

集計モデルの実証比較：組成データ解析の観点から

輸送ネットワークと都市の空間構造の研究プロジェクト

2025 年 6 月

公益社団法人 日本交通政策研究会

1. “日交研シリーズ”は、公益社団法人 日本交通政策研究会の実施するプロジェクトの研究
成果、本研究会の行う講演、座談会の記録、交通問題に関する内外文献の紹介、等々を印
刷に付して順次刊行するものである。
2. シリーズはAよりEに至る5つの系列に分かれる。
シリーズAは、本研究会のプロジェクトの成果である書き下ろし論文を収める。
シリーズBは、シリーズAに対比して、より時論的、啓蒙的な視点に立つものであり、
折にふれ、重要な問題を積極的にとりあげ、講演、座談会、討論会、その他の方法によっ
てとりまとめたものを収める。
シリーズCは、交通問題に関する内外の資料、文献の翻訳、紹介を内容とする。
シリーズDは、本研究会会員が他の雑誌等に公けにした論文にして、本研究会の研究調査
活動との関連において復刻の価値ありと認められるもののリプリントシリーズである。
シリーズEは、本研究会が発表する政策上の諸提言を内容とする。
3. 論文等の内容についての責任はそれぞれの著者に存し、本研究会は責任を負わない。
4. 令和2年度以前のシリーズは印刷及び送料実費をもって希望の向きに頒布するものとする。

公益社団法人日本交通政策研究会

代表理事 山 内 弘 隆
同 原 田 昇

令和2年度以前のシリーズの入手をご希望の向きは系列番号
を明記の上、下記へお申し込み下さい。

〒102-0073 東京都千代田区九段北 1-12-6

守住ビル 4階

公益社団法人日本交通政策研究会

電 話 (03) 3263-1945 (代表)

F a x (03) 3234-4593

E-Mail:office@nikkoken.or.jp

日交研シリーズ A-920

令和 6 年度自主研究プロジェクト

輸送ネットワークと都市の空間構造の研究

刊行：2025 年 6 月

集計モデルの実証比較：組成データ解析の観点から

Empirical comparison of aggregate-models: From the perspective of compositional data analysis

主査：高橋 孝明（東京大学空間情報科学研究センター・教授）

Takaaki Takahashi

要 旨

本稿では、組成データ解析（compositional data analysis: CoDA）の知見を援用して、土木計画学分野で用いられてきた「集計モデル」を再考する。土木計画学分野において、交通需要および交通行動のモデリングは重要な研究課題の一つであり、それらのモデルは非集計モデルと集計モデルに大別される。近年、非集計モデルは洗練と拡張が進み、ランダム効用最大化理論による裏付けのもとで分野の主要テーマの一つと扱われているのに対し、集計モデルは四段階推定法の古典的な手法と位置付けられ、さらに計算機性能の向上により簡便性の利点を失ったことで、いくつかの利用場面は残されているものの主要テーマとは扱われていない。一方、他分野をみると、地質学分野を中心に発展した CoDA が発展し、集計モデル、特に構成比や比率からなるデータを扱うモデルが体系化され、様々な分野で応用が行われるようになってきている。そこで本研究では、CoDA を含む他分野で提案されたモデルを用いて、それぞれの精度指標を比較し、「集計モデル」に関する知見の更新を図ることを目的とする。実証分析の例として、パーソントリップ調査から得られた東京都市圏における交通機関分担率を用いて、比較分析を行う。その結果、モデルの当てはまりの良さの比較から、データにゼロ値がない場合は CoDA で提案されたモデルが最良であること、ゼロ値の割合が増えるほどその精度が低下し、ゼロ値の割合が 1 割を超えると条件付き集計ロジットモデルやディリクレ回帰モデルの精度に良くなることを確認した。本稿の結果は、土木計画学分野に対して CoDA の応用可能性を示すだけでなく、CoDA に対してもゼロ値の影響の度合いを把握する知見となり得ることから、両分野への貢献が期待できるものである。

キーワード：集計モデル、集計ロジットモデル、組成データ解析、対数比変換、ゼロ値

Keywords : aggregate-model, aggregate-logit model, compositional data analysis, log-ratio transformation, zero value

目 次

1 章	はじめに	1
2 章	組成データ解析	3
2.1	組成データ	3
2.2	対数比解析	5
3 章	実証分析の設定	7
3.1	モデル	7
3.2	データ	10
3.3	精度指標	13
3.4	ゼロ値率	14
4 章	実証分析の結果	14
4.1	パラメータ推定結果	14
4.2	精度評価結果	19
4.3	ゼロ値率と精度評価の関係	20
5 章	おわりに	22
	参考文献	22

研究メンバーおよび執筆者（敬称略・順不同）

研究メンバー

高橋孝明（東京大学空間情報科学研究センター教授）
佐藤泰裕（東京大学大学院経済学研究科教授）
金本良嗣（政策研究大学院大学客員教授）
田渕隆俊（中央大学国際経営学部教授）
浅田義久（日本大学経済学部教授）
小川 光（東京大学大学院経済学研究科教授）
吾郷貴紀（専修大学商学部教授）
河端瑞貴（慶応義塾大学大学院経済学研究科教授）
太田 充（筑波大学大学院システム情報工学研究科准教授）
藤嶋翔太（一橋大学大学院経済学研究科准教授）
中川万理子（一橋大学経済研究所講師）
大津優貴（東京大学空間情報科学研究センター講師）
相場郁人（東京大学大学院経済学研究科博士課程）
久住竜也（東京大学大学院経済学研究科博士課程）

執筆者

吉田崇紘（東京大学空間情報科学研究センター講師）

集計モデルの実証比較：組成データ解析の観点から

1. はじめに

土木計画学分野において、交通需要および交通行動のモデリングは重要な研究課題の一つである。同分野においてそれらのモデルは、「非集計モデル」と「集計モデル」に大別される（例えば、土木学会土木計画学研究委員会，1995）。この2つの用語は、分析対象とする交通行動データが個人や世帯単位の個票のデータであるか、市区町村やゾーンなどの空間的な範囲で集計されたデータであるかを区別する独自の概念として慣用されている。「集計モデル」は、1950年代から交通需要予測に用いられてきた4段階推定法として利用・開発されてきた（Ben-Akiva and Lerman, 1985）。

交通機関の分担交通量を予測する集計ロジットモデルは我が国の同分野においてよく利用されてきた「集計モデル」の一つである。ここで、個人や世帯単位の個票のデータを y_i ($i = 1, \dots, n$)、集計されたデータを $S_I = \sum_{i \in I} y_i$ として、対応する集計範囲におけるデータ数 N_I を用いて $P_I = S_I/N_I$ と比率に変換することでロジットモデルにおける選択確率とみなすことができ、その推定に利用することが可能である。ここで、 y_i を二項変数としたときの条件付き確率の対数オッズ比が説明変数 X_i と回帰係数 β で表されると仮定すると、そのモデルは

$$\ln \frac{\Pr(y_i = 1|X_i)}{\Pr(y_i = 0|X_i)} = \ln \frac{p_i}{1 - p_i} = \text{logit}(p_i) = X_i \beta$$

となる。このモデルの推定法には、誤差項が平均0、分散 $\sigma_I^2 = 1/[N_I P_I (1 - P_I)]$ であるとして重み付き最小二乗法（最小カイ二乗法）を用いる方法がある。この方法は、Berkson (1955) によって提案され、Theil (1969) によって多項ロジットモデルに拡張された Berkson-Theil 法として知られ、計算が非常に簡便であることに特徴がある（Ben-Akiva and Lerman, 1985）。このモデルは交通需要予測の実務において利用されてきたが、ランダム効用最大化理論（MacFadden, 1974）による理論的背景の整備と計算機性能の向上により非集計モデルの活用と洗練が進んだことから、同分野でこのモデルを利用する機会は近年極めて限定的となっている（Cameron and Trivedi, 2005）。

一方、他分野を見ると「集計モデル」の改良や応用に関する研究が進み、再び脚光が当たるようになってきている。特に岩石や土層の化学組成を扱う地質学分野において「組成デー

タ解析 (Compositional Data Analysis: CoDA)」と呼ばれる手法が発展し、各要素の値が非負でそれら値の和が一定となるようなデータの分析手法が幅広い分野で利用されるようになってきている (e.g, Aitchison, 1982; Filzmoser et al., 2018)。CoDA における近年の「集計モデル」の発展を土木計画学分野に応用した研究はないが、Morais et al. (2017) はマーケティング分野に応用し、集計ロジットモデルを含め関連モデルの比較を行っている。彼女らは、モデルの当てはまりの良さの観点から CoDA で提案されたモデルが従来のモデルより優れた性能を発揮しうる実証分析の結果を報告している。

土木計画学分野においても、CoDA の知見を援用して、「集計モデル」の再整理・更新する余地は大きいと考えられる。その際には、同分野で用いるデータの特徴を踏まえながら知見を取り入れることが望ましい。特に、CoDA の源流である地質学分野においてはあまり想定されない要素 (交通機関分担率) の値がゼロとなる場合の対処とその影響の把握は重要な課題となりえる。地質学分野において、分類器の検出限界で発生したゼロ (rounded zero) は検出限界の半分の値で置換するなどの操作が行われており (e.g., Martín-Fernández et al., 2012)、意味のあるゼロ (essential zero) が多い社会科学のデータの場合にも同様の操作を行う必要があるものの、そのモデルの当てはまりの良さに与える影響の程度を把握しておくことは重要である。

本稿では、交通機関分担率データを対象に、土木計画学分野でこれまでに用いられてきたモデルや、CoDA を含む他分野で提案されたモデルを比較することで「集計モデル」に関する知見の更新を図ることを目標とする。具体的には、Morais et al. (2017) の比較分析を下敷きにして、土木計画学分野で用いられる交通機関分担率データに応用し、各モデルの当てはまりの良さを比較する。ここで、交通機関分担率データは分担率に 0 が含まれることが多いことから、ゼロの割合の程度に対する各モデルの当てはまりの良さの頑健性を検証することとする。なお、モデルを推定する際、ゼロを微小値に変換することは分野を問わず慣例であることを踏まえ本研究でもこれを採用し、実際にはゼロではなく微小値に変換した割合の程度を比較検証する。本稿の分析は、土木計画学分野に対して CoDA の応用可能性を示すだけでなく、社会経済データを対象とする CoDA においてはゼロの対処は重要な課題であり、その影響の度合いを把握することは重要な知見となり得ることから、CoDA に対する貢献も期待できる。

本研究の構成は以下の通りである。まず第 2 節では、組成データ解析のアプローチを紹介し、これにより各分野で用いられてきた「集計モデル」を整理する。第 3 節では、比較分析

の設定や取り上げるモデル、実証分析で用いるパーソントリップ調査のデータを説明する。第4節では実証分析の結果を報告、考察し、最後に第5節で本研究のまとめを述べる。

2. 組成データ解析

2.1 組成データ

比率のように値が非負で和が一定となるような制約のついたデータは「組成データ」と呼ばれている（図2.1）。この制約により、組成データの標本空間 $\mathcal{S}^J$ は、 J を次元数とすると、実数空間 $\mathcal{R}^J$ とは異なり、式（1）に表す単体空間 $\mathcal{S}^J \subset \mathcal{R}^J$ に限定されることになる。

$$\mathcal{S}^J = \left\{ \mathbf{p} = (p_1, p_2, \dots, p_J) \left| p_j > 0, j = 1, \dots, J, \sum_{j=1}^J p_j = \kappa \right. \right\} \quad (1)$$

ここで、 κ は定数であり、たとえば成分を割合で表す場合は $\kappa = 1$ 、百分率で表す場合は $\kappa = 100$ となる。組成データは、その名称自体は一般的には知られていないものの、至る所で目にするデータの種類である。普段あまり意識されないが、統計学的には、疑似相関の問題から組成データの分析の際には値の総和が一定であるという「定数和制約（Constant-sum constraint）」を考慮する必要がある、地質学を中核にこれを考慮した「組成データ解析（Compositional Data Analysis: CoDA）」が発展している。組成データの定義域は、定数和制約により変数の自由度が一つ下がることから、単体空間上に限定される（図2.2：3次元の組成データの定義域は、定数和制約により2次元の平面に制限される）。このために通常の距離指標（ユークリッド距離）に基づいた統計手法を組成データに直接適用すると、データ間の距離関係を過少に評価し誤った結果の解釈を導く可能性が古くから指摘されている（Aitchison, 1986）。しかし、この分野の嚆矢となった書籍 Aitchison (1986) から既に35年以上が経過しているにも関わらず、こうした警告は、組成データがありふれたデータであること、そして

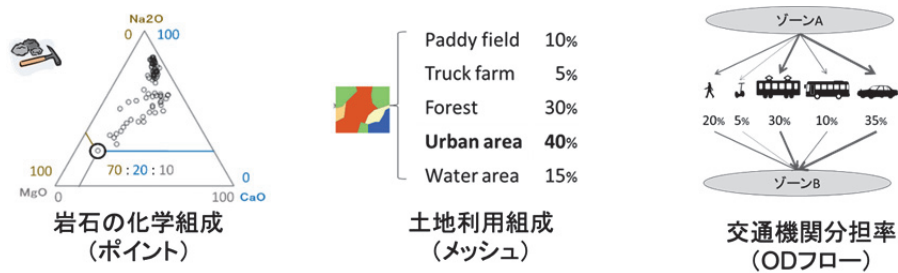


図 2.1 組成データの例

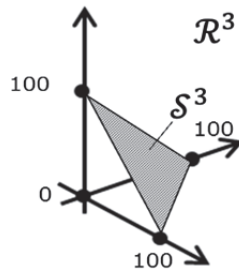


図 2.2 組成データの標本空間 ($J = 3$; $\kappa = 100$ の場合)

統計解析ソフトの上では通常のデータと並列に分析が実行できてしまうことから、多くの分野で見過ごされてきた (Aitchison and Egozcue, 2005)。

こうした経緯があるものの現在は多様な分野で CoDA の応用が進んでいる。関連する研究動向の整理を図 2.3 に示す。前述のように、岩石の化学組成を扱う地質学においてまず CoDA の高度化が進められてきた。一方で、統計学では、定数和制約に起因する疑似相関の問題などが古くから認識・議論されてきた。しかし、組成データの扱いに関する問題は、定数和制約を解く操作として対数比変換を核に据え組成データの解析手法をはじめて体系化した統計学者 J. Aitchison による著書 Aitchison (1986) が出版されるまで不幸にも長らく放置されてきたという指摘がある (Aitchison and Egozcue, 2005)。ごく僅かな研究者らの尽力によって手法の精緻化と検証が進められていたが、2010 年代には生物学、脳科学にも応用が広がり、2015 年の CoDA の国際ワークショップ (CoDAWork) では翌年の学会設立が提言されるなど、いま CoDA は開花期を迎えようとしている。現在は社会経済データへの適用上の課題、例えば時系列相関の考慮、ゼロ過剰・欠損データへの処理などへの対処法に関する研究が進んでいる。

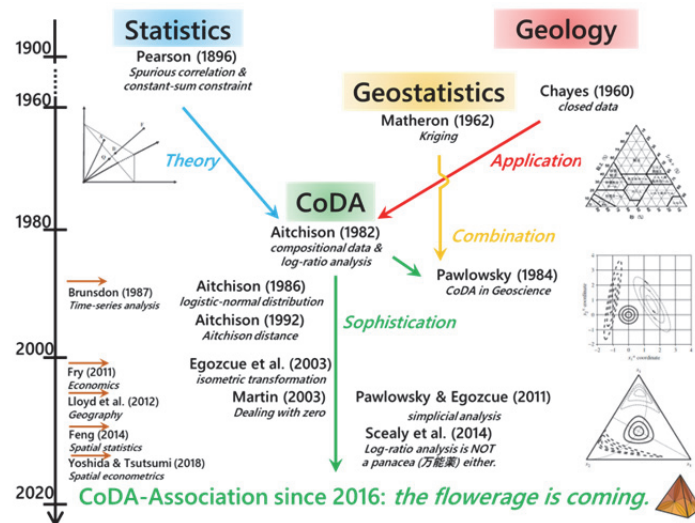


図 2.3 組成データ解析の研究動向

2.2 対数比解析

組成データを分析するためのアプローチの一つに、写像変換を利用するものがある。これは、組成データに従来の統計解析手法を適用したい場合には、式 (1) で表した組成データの標本空間から実数空間へ写像変換をまず行えばよい、とするものである。その変換は、操作の特徴から対数比変換 (log-ratio transformation) と呼ばれている (Aitchison, 1986)。この変換操作として 3 つの代表的な変換法が知られており、以下に整理しよう。

2.2.1 加法対数比変換 (additive log-ratio (alr) transformation)

組成データ $\mathbf{p} \in \mathcal{S}^J$ の alr 変換は、例えば p_J 成分で規格する場合は式 (2) で表すことができる。

$$\text{alr}(\mathbf{x}) = \left(\ln \frac{p_1}{p_J}, \ln \frac{p_2}{p_J}, \dots, \ln \frac{p_{J-1}}{p_J} \right) \quad (2)$$

この alr 変換によって、組成データは実空間 $\mathcal{R}^{J-1}$ 上に写像される。ここで alr 変換において注意を要するのは、元々 J 成分であった組成データが $J-1$ 成分の実数データに変換されることと、分析者による規格化する成分の選択によって変換したサンプル間の距離、角度の配置関係が異なるものになることである。後述で紹介する 2 つの変換法と比べて変換した変数の意味解釈は容易であることには利点を有する。また、回帰モデルの推定においては、定義域の制限 (比率であった組成データの各成分のとりうる値の範囲は $(0, 100)$) から解放されるため、前述の Berkson-Theil 法のような簡便法が利用可能となる点も利点として挙げることができる。

2.2.2 有心対数比変換 (centered log-ratio (clr) transformation)

組成データ $\mathbf{p} \in \mathcal{S}^J$ の clr 変換は、 $\mathbf{p}$ の幾何平均 $g(\mathbf{p})$ で各成分を規格する変換法であり、式 (3) で表すことができる。

$$\text{clr}(\mathbf{p}) = \left(\ln \frac{p_1}{g(\mathbf{p})}, \ln \frac{p_2}{g(\mathbf{p})}, \dots, \ln \frac{p_J}{g(\mathbf{p})} \right) \quad (3)$$

この clr 変換によって組成データは、元々の組成データの次元数と同じ空間に写像されることである。幾何平均で規格化することで、alr 変換において注意を要した規格化成分の選択による変換したサンプル間の配置関係の問題は解消されている。一方で、変換したデータの和がゼロとなり一定なため定数和制約を緩和できていないことで実空間への写像となっていないことに注意を要する。

2.2.3 等長対数比変換 (isometric log-ratio (ilr) transformation)

組成データ $\mathbf{p} \in \mathcal{S}^J$ の ilr 変換は、変換したデータが $\mathcal{R}^{J-1}$ で正規直交系となるように写像する方法である。その変換式は、式 (4) で表すように、alr 変換や clr 変換とは異なり、直感的には解釈が難しいものとなっている。

$$\text{ilr}(\mathbf{p})_j = \sqrt{\frac{J-j}{J-j+1}} \ln \frac{p_j}{\left(\prod_{j'=j+1}^J p_{j'}\right)^{\frac{1}{J-j}}}, \quad j = 1, \dots, J-1 \quad (4)$$

わかりにくいため、3 成分からなる組成データ $\mathbf{p} = (p_1, p_2, p_3)$ の場合は

$$\begin{aligned} \text{ilr}(\mathbf{p})_1 &= \sqrt{\frac{2}{3}} \ln \frac{p_1}{\sqrt{p_2 p_3}} = \sqrt{\frac{2}{3}} \ln p_1 - \frac{1}{\sqrt{6}} (\ln p_2 + \ln p_3), \\ \text{ilr}(\mathbf{p})_2 &= \frac{1}{\sqrt{2}} \ln \frac{p_2}{p_3} = \frac{1}{\sqrt{2}} (\ln p_2 - \ln p_3) \end{aligned}$$

で表すことができ、この場合の $\mathcal{S}^3 \rightarrow \mathcal{R}^2$ への変換行列は

$$\begin{bmatrix} \sqrt{\frac{2}{3}} & 0 \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

となる。変換したデータの解釈は困難であるが、この ilr 変換によって、組成データは実空間 $\mathcal{R}^{J-1}$ 上に写像され、alr 変換や clr 変換で問題となった点は解消されている。このため、ilr 変換は CoDA において推奨される変換法となっている。近年は、正規直交系であるように変換されることが慣例となっていることと、clr 変換も等長な変換法であり混同を招く恐れがあることから、ilr 変換を正規直交対数比変換 (orthonormal log-ratio (olr) transformation) と呼称することが提唱されている (Martín-Fernández, 2019)。

2.2.4 対数比変換と集計モデルの対応

ここで、alr 変換は交通工学分野における「集計ロジットモデル」、clr 変換はマーケティング分野において「Multiplicative Competitive Interaction (MCI) モデル」(Nakanishi and Cooper, 1974) と呼ばれるモデルの変換法にそれぞれ相当する。第 1 節で説明した集計ロジットモデルのオッズ比の形をとる被説明変数

$$\ln \frac{p_i}{1 - p_i} = \text{logit}(p_i)$$

は、式 (2) で示した alr 変換の 2 次元の場合そのものである。詳細は割愛するが、MCI モデルの被説明変数は幾何平均で除する方法をとることから、同様に clr 変換そのものである。このように、CoDA における対数比変換法の整理に依拠することで、既存のモデルを対数比変換法の観点から俯瞰的に整理し、位置付けることができると言える。

3 実証分析の設定

本節では、Morais et al. (2017) の比較分析を下敷きにして、土木計画学分野で用いられる代表的な組成データであるゾーン間交通機関分担率データに応用し、各モデルの当てはまりの良さを比較する。被説明変数はゾーン間交通機関別分担率とし、説明変数は、選択肢固有変数 x_{ij} として発地ゾーンー着地ゾーン (Origination-Destination: OD) 間の移動時間、OD ペア固有変数 w_i として OD 間の直線距離の 2 つとする。ここで、ゾーン間交通機関分担率データは分担率にゼロが含まれることが多いことから、ゼロの割合の程度に対する各モデルの当てはまりの良さの頑健性を検証することとする。なお、モデルを推定するにはゼロは微小値に変換することが分野を問わず慣例であり、実際には微小値に変換した割合の程度を比較検証する。本節の分析は、土木計画学分野に対して CoDA の応用可能性を示すだけでなく、社会経済データを対象とする CoDA においてはゼロの対処は重要な課題であり、その影響の度合いを把握することは重要な知見となり得ることから CoDA に対する貢献も期待できる。

3.1 モデル

3.1.1 集計ロジットモデル

本稿では、土木計画学分野でよく用いられてきた集計モデルとして、集計ロジット (Aggregated Logit: AL) モデルを用いる。AL モデルは、 x_{ij} と選択肢間共通のパラメータ $\beta^{(x)}$ を仮定し、OD ペア i で交通機関 j を選択したときの効用 U_{ij} を

$$U_{ij} = a_j + \beta^{(x)} x_{ij}$$

とする。この効用に誤差項 ε_{ij} を加え、 ε_{ij} にガンベル分布に仮定すると、OD ペア i で交通機関 j を選択する確率 P_{ij} は

$$P_{ij} = \frac{\exp(a_j + \beta^{(x)} x_{ij})}{\sum_j \exp(a_j + \beta^{(x)} x_{ij})}$$

と表される。AL モデルの推定は、これに選択肢間で対数比をとった

$$\ln\left(\frac{P_{ij'}}{P_{ij}}\right) = (a_{j'} - a_j) + \beta_j^{(x)}(x_{ij'} - x_{ij}) + (\varepsilon_{ij'} - \varepsilon_{ij})$$

に対して重み付き最小二乗法（最小カイ二乗法）により実行可能である。この方法は、Berkson（1955）によって二項の場合が提案され、Theil（1969）によって多項の場合に拡張された Berkson-Theil 法として知られ、計算が非常に簡便であることに特徴がある（Ben-Akiva and Lerman, 1985）。

3.1.2 多項条件付(集計)ロジットモデル

多項条件付(集計)ロジット (Multinomial Conditional (Aggregated) Logit: MNCL) モデルは、AL モデルで仮定した変数とパラメータに加えて、 w_i と選択肢固有の回帰パラメータ $\beta_j^{(w)}$ を仮定し、効用 U_{ij} は

$$U_{ij} = a_j + \beta_j^{(x)}x_{ij} + \beta_j^{(w)}w_i$$

で表される。ここで、パラメータの識別のため、 $a_1 = 0, \beta_1 = 0$ とする。AL モデルと同様にすると、確率 P_{ij} は

$$P_{ij} = \frac{\exp(a_j + \beta_j^{(x)}x_{ij} + \beta_j^{(w)}w_i)}{\sum_j \exp(a_j + \beta_j^{(x)}x_{ij} + \beta_j^{(w)}w_i)}$$

と表される。MNCL モデルの推定には最尤法を用いる。

3.1.3 ディリクレ回帰モデル

ディリクレ回帰 (Dirichlet Regression: DIR) モデルは、被説明変数にディリクレ分布を仮定した一般化線形モデルの枠組みのモデルである。AL モデルと MNCL モデルとの説明変数とパラメータの違いは、 x_{ij} のパラメータに選択肢固有のパラメータ $\beta_j^{(x)}$ を仮定することである。ディリクレ分布の集中度パラメータ $\alpha = (\alpha_1, \dots, \alpha_J)$ は対数リンク関数を用いて

$$\ln(\alpha_j | x_i) = a_j + \beta_j^{(x)}x_{ij} + \beta_j^{(w)}w_i$$

となり、確率 P_{ij} は

$$P_{ij} = \frac{\alpha_j | x_i}{\sum_j \alpha_j | x_i} = \frac{\exp(a_j + \beta_j^{(x)}x_{ij} + \beta_j^{(w)}w_i)}{\sum_j \exp(a_j + \beta_j^{(x)}x_{ij} + \beta_j^{(w)}w_i)}$$

と表すことができる。なお、本稿では、説明変数のみに依存して α をモデル化することができる共通パラメータ化 (common parameterization) を採用する。DIR モデルの推定には最

尤法を用いる。

3.1.4 組成回帰モデル

組成回帰（CoDA）モデルは、組成データ解析の枠組みの中で提案されたモデルである。CoDA モデルは、ilr 変換（式 4）を用いて最小二乗法により簡便に集計モデルを推定するものとする。 $\mathcal{S}^J$ 上の CoDA モデルは、

$$\mathbf{P}_i = \mathbf{a} \oplus \mathbf{B}^{(x)} \boxdot \mathbf{z}_i \oplus \mathbf{w}_i \odot \boldsymbol{\beta}^{(w)} \oplus \boldsymbol{\varepsilon}_i \quad (5)$$

と組成データ解析で提案・整理された演算子で表される。ここで、 $\mathbf{P}_i, \mathbf{a}, \mathbf{z}_{ik}, \boldsymbol{\varepsilon}_i \in \mathcal{S}^J, \mathbf{B} \in \mathcal{R}_{J \times J}$ である。この式で使われている演算子を整理すると、摂動演算（perturbation operation）は

$$\mathbf{p} \oplus \mathbf{q} = \mathcal{C}(p_1 \cdot q_1, p_2 \cdot q_2, \dots, p_J \cdot q_J) \in \mathcal{S}^J$$

であり、冪演算（powering operation）は

$$\lambda \odot \mathbf{p} = \mathcal{C}(p_1^\lambda, p_2^\lambda, \dots, p_J^\lambda) \in \mathcal{S}^J$$

である。なお、 $\mathbf{p}, \mathbf{q} \in \mathcal{S}^J, \lambda \in \mathcal{R}$ とする。摂動演算、冪演算はそれぞれユークリッド空間上の足し算、掛け算に相当する。行列積は

$$\mathbf{B} \boxdot \mathbf{p} = \mathcal{C}\left(\prod_{j=1}^J p_j^{b_{1j}}, \prod_{j=1}^J p_j^{b_{2j}}, \dots, \prod_{j=1}^J p_j^{b_{Jj}}\right) \in \mathcal{S}^J$$

である。ここで、 $\mathbf{B} \in \mathcal{R}_{J \times J}, \mathbf{p} \in \mathcal{S}^J$ としている。式（5）を演算子を使って展開すると、

$$\mathbf{P}_i = \mathcal{C}\left(a_1 \prod_{m=1}^J x_{im}^{\beta_{1m}^{(x)}} w_i^{\beta_1^{(w)}} \varepsilon_i, a_2 \prod_{m=1}^J x_{im}^{\beta_{2m}^{(x)}} w_i^{\beta_2^{(w)}} \varepsilon_i, \dots, a_J \prod_{m=1}^J x_{im}^{\beta_{Jm}^{(x)}} w_i^{\beta_J^{(w)}} \varepsilon_i\right)$$

となる。ここで、他 3 つのモデルとの整合から誤差項 ε_i を除くと、効用 U_{ij} が

$$U_{ij} = a_j \cdot \prod_{m=1}^J x_{im}^{\beta_{jm}^{(x)}} \cdot w_i^{\beta_j^{(w)}}$$

と書けることから、確率 P_{ij} は

$$P_{ij} = \frac{\exp\left(\ln(a_j) + \sum_{m=1}^J \beta_{jm}^{(x)} \ln(x_{im}) + \beta_j^{(w)} \ln(w_i)\right)}{\sum_j \exp\left(\ln(a_j) + \sum_{m=1}^J \beta_{jm}^{(x)} \ln(x_{im}) + \beta_j^{(w)} \ln(w_i)\right)}$$

となる。前述の 3 つのモデルとの違いは、定数項と説明変数が対数となることに加え、選

択肢 j の効用に選択肢 j 以外の選択肢の変数とパラメータが考慮されることである。これを Morais et al. (2017) は相互効果 (cross effects) と呼んでいる。CoDA モデルの推定は、ilr 変換して出力される $l (= 1, \dots, (J - 1))$ 本のモデル (式 6) をそれぞれ最小二乗法で推定し、逆対数比変換によって各回帰パラメータを得る。

$$p_{ij}^{(l)} = \ln(a_j^{(l)}) + \sum_{m=1}^J \beta_{jm}^{(x)(l)} \ln(x_{im}^{(l)}) + \beta_j^{(w)(l)} \ln(w_i) \quad (6)$$

3.2 データ

東京都市圏 (東京 MA) の 2008 年パーソントリップ調査データを用いる。東京都市圏交通計画協議会のウェブページから分担率データを国土数値情報から「OD 量データ」をそれぞれ入手した。対象地域のマップを図 3.1 に示す。「OD 量データ」は、パーソントリップ調査の結果に基づき、国土交通省が位置情報を追加したデータである。両データには、出発地 ID と目的地 ID があり、これに基づいてデータを連結・結合した。

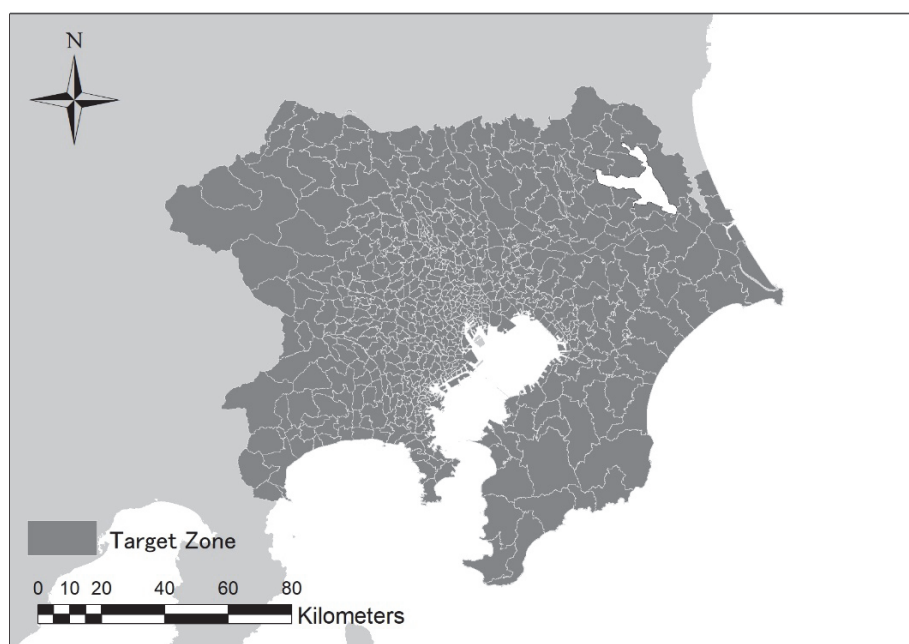


図 3.1 対象ゾーン

本研究ではゾーン間交通機関分担率に焦点を当てるが、データからはトリップ目的に応じたゾーン間交通機関分担率を得ることができる。トリップ目的別にデータを分けることで複数の実証的な検証を行うことが可能なため、図 3.2 に示す 5 つのトリップ目的 (To work place,

To school, Back, Private, Bussiness) 別に、5 つの交通機関分担 (Train, Bus, Car, Bike, Others) を対象として各モデルを適用する。

		着目的						
		c.自宅へ	a.勤務先へ	f.業務先へ	b.通学先へ	d.買物へ	e.その他私事	不明
発目的	c.自宅へ	Back	To work place		To school	Private		
	a.勤務先へ							
	f.業務先へ							
	b.通学先へ							
	d.買物へ							
	e.その他私事							
	不明							
	8.不明							

Transportation Planning council for Tokyo Metropolitan Area (2012)

図 3.2 トリップ目的

表 3.1 はデータの全体の概況、表 3.2 はトリップ目的別の交通機関分担率とゼロ比率である。ゼロ値率は、交通機関分担にどれだけゼロが含まれているかを示し、その算出例を図 3.3 に示す。3 つの OD ペア (N) と 5 つの交通機関 (J) があり、全 15 の OD ペアと交通機関のセルがある内で、交通機関を使った OD がなかったセルが 3 ケースあった場合に、そのゼロ値率は $100 * 3/15 = 20 (\%)$ とする。

東京 MA における全体およびトリップ目的別にみると、others の場合を除き Car の比率が大きいことが表 3.1、表 3.2 からわかる。表 3.2 の To school をみると、その others の比率は、他のトリップ目的の others の比率よりも著しく大きいことも確認できる。これは、others に walking と biking が含まれており、通学する人々はそれらを利用する傾向があるためと考えられる。

表 3.1 データの概況

東京 MA	
面積	約 15,000 km ²
対象年	2008
交通機関分担	
Train	7.4%
Bus	3.6%
Car	27.0%
Bike	2.4%
Others	60.0%
トリップ目的	
To work place	11.8%
To school	7.0%
Back	40.6%
Private	36.4%
Business	4.3%

表 3.2 トリップ目的別の交通機関分担率、ゼロ比率

トリップ目的	交通機関	Tokyo MA	
		分担率 (%)	ゼロ値率 (%)
To work place N _{Tokyo MA} =3357	Train	14.4	27.28
	Bus	4.5	
	Car	29.5	
	Bike	5.3	
	Others	46.3	
To school N _{Tokyo MA} =2631	Train	4.5	59.96
	Bus	1.4	
	Car	6.2	
	Bike	0.4	
	Others	70.9	
Private N _{Tokyo MA} =3419	Train	5.5	24.74
	Bus	3.7	
	Car	28.4	
	Bike	1.4	
	Others	60.9	
Business N _{Tokyo MA} =3159	Train	11.3	56.9
	Bus	1.5	
	Car	54.5	
	Bike	4.7	
	Others	27.9	
Back N _{Tokyo MA} =3391	Train	7.1	17.44
	Bus	3.9	
	Car	25.5	
	Bike	2.4	
	Others	61.1	

	Share of Travel Mode				
	Train	Bus	Car	Bike	Others
1	10	10	20	20	40
2	15	<u>0</u>	<u>0</u>	30	55
3	20	30	10	<u>0</u>	40

$$\text{Rate of 0s} = 3 / 15 = 20.0\%$$

図 3.3 ゼロ値率の算出例

本研究で用いる説明変数の記述統計量を表 3.3 に示す。本研究では、前述のように選択肢固有変数 x_{ij} として OD 間の移動時間、OD ペア固有変数 w_i として OD 間の直線距離を用いる。表 3.3 より、Train の平均所要時間が最も長く、Bus が 2 番目、Car が 3 番目に長いことがわかる。東京 MA に住む人々は、長時間・長距離の移動には公共交通機関を利用する傾向を反映したものと言えよう。

表 3.3 説明変数（OD 間の移動時間、OD 間の直線距離）の記述統計量

	Tokyo MA	
	Mean	S.D.
OD 間の移動時間 (min)		
Train	40.11	14.72
Bus	36.79	18.08
Car	22.31	8.95
Bike	19.85	11.85
Others	15.08	8.28
OD 間の直線距離 (m)	3820.732	2576.03

3.3 精度指標

各モデルの当てはまりの良さを比較するため、Morais et al. (2017) に倣い以下の 3 つの指標を採用する。 R_T^2 と R_A^2 は値が大きいほど精度が良いことを示し、 $KL\ div.$ は逆に値が小さいほど精度が良いことを示す。

(a) R^2 based on the total variability

$$R_T^2 = \frac{\text{totvar}(\hat{\mathbf{y}})}{\text{totvar}(\mathbf{y})}, \quad \text{totvar}(\mathbf{y}) = \frac{1}{2J} \sum_{j,j^*} \text{var} \left(\ln \frac{y_j}{y_{j^*}} \right)$$

(b) R^2 based on Aichison's distance

$$R_A^2 = 1 - \frac{\sum_{i=1} \{d_A(\mathbf{y}_i, \hat{\mathbf{y}}_i)\}^2}{\sum_{i=1} \{d_A(\mathbf{y}_i, \bar{\mathbf{y}}_i)\}^2}, \quad d_A(\mathbf{y}_i, \hat{\mathbf{y}}_i) = \sqrt{\sum_{j=1} \left\{ \log \frac{y_{ij}}{g(\mathbf{y}_i)} - \log \frac{\hat{y}_{ij}}{g(\hat{\mathbf{y}}_i)} \right\}^2}$$

(c) The Kullback-Leibler divergence

$$KL \text{ div.} = \sum_{i=1} \sum_{j=1} \left\{ y_{ij} \log \left(\frac{y_{ij}}{\hat{y}_{ij}} \right) \right\}$$

3.4 ゼロ値率

ゼロ値率の大小と当てはまりの良さの関係を調べるため、以下のシミュレーションを行う。
そのステップは次の通りとする：

1. 被説明変数がゼロ値のグループとそうでないグループに分割する
2. ランダムにゼロ値削減率を決定し、被説明変数がゼロ値のグループの中から削減する対象をランダムに選定する
3. 被説明変数がゼロ値を含まないグループのデータと2の過程で残ったデータを結合する
4. 各モデルのパラメータを推定し、精度指標を算出する
5. ステップ1から4の手順を300回繰り返す

4 実証分析の結果

4.1 パラメータ推定結果

表 4.1 から 4.4 に ALM、MNCL、DIR、CoDA の各モデルのパラメータ推定結果を示す。
MNCL の結果に注目すると、Train's intercept と length のパラメータは 0 であり、これは 2.3 で示した識別可能性の理由によるものである。表に示すように、ALM、MNCL、DIR のほとんどのパラメータは有意であり、直感的に整合する結果が得られている。なお、CoDA の結

果は対数比変換をしたものであり、有意性について直接的に解釈することは難しい (Morais et al., 2018; Morais & Thomas-Agnan, 2021)。

表 4.1 パラメータ推定結果：ALM

Purpose		Coef.	S.E.
To work place	Intercept	-3.82	6.54×10^{-2}
	Time	-4.06×10^{-3}	1.72×10^{-3}
	Train D	1.57	8.17×10^{-2}
	Bus D	-2.71×10^{-1}	7.93×10^{-2}
	Car D	1.73	7.38×10^{-2}
	Length	5.19×10^{-4}	1.07×10^{-5}
To school	Intercept	-6.56	8.93×10^{-2}
	Time	-5.56×10^{-3}	2.38×10^{-3}
	Train D	2.88	1.14×10^{-1}
	Bus D	9.25×10^{-1}	1.10×10^{-1}
	Car D	1.91	1.03×10^{-1}
	Length	5.98×10^{-4}	1.51×10^{-5}
Private	Intercept	-4.92	5.83×10^{-2}
	Time	-3.16×10^{-3}	1.48×10^{-3}
	Train D	1.95	7.17×10^{-2}
	Bus D	1.37	7.04×10^{-2}
	Car D	3.57	6.60×10^{-2}
	Length	4.53×10^{-4}	9.40×10^{-6}
Business	Intercept	-3.14	6.94×10^{-2}
	Time	-3.07×10^{-3}	1.85×10^{-3}
	Train D	7.55×10^{-1}	8.73×10^{-2}
	Bus D	-5.16×10^{-1}	8.49×10^{-2}
	Car D	4.88	7.89×10^{-2}
	Length	4.75×10^{-4}	1.15×10^{-5}
Back	Intercept	-4.31	5.31×10^{-2}
	Time	-4.03×10^{-3}	1.40×10^{-3}
	Train D	1.64	6.63×10^{-2}
	Bus D	4.44×10^{-1}	6.44×10^{-2}
	Car D	2.40	5.99×10^{-2}
	Length	4.62×10^{-4}	8.68×10^{-6}

表 4.2 パラメータ推定結果：MNCL

Purpose	Generic Variable		Train		Bus		Car		Bike		Others	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
To work place	-	-	0.00	-	-6.38×10 ⁻¹	1.46×10 ⁻¹	-2.94×10 ⁻²	1.07×10 ⁻¹	-1.11	1.52×10 ⁻¹	1.65	1.06×10 ⁻¹
	-2.29×10 ⁻²	2.56×10 ⁻³	-	-	-	-	-	-	-	-	-	-
	-	-	0.00	-	-1.51×10 ⁻⁴	1.88×10 ⁻⁵	-6.68×10 ⁻⁵	1.88×10 ⁻⁵	-1.16×10 ⁻⁴	2.93×10 ⁻⁵	-5.47×10 ⁻⁴	2.52×10 ⁻⁵
To school	-	-	0.00	-	-8.43×10 ⁻¹	1.88×10 ⁻¹	6.97×10 ⁻²	1.54×10 ⁻¹	-2.77	3.67×10 ⁻¹	2.41	1.22×10 ⁻¹
	-1.53×10 ⁻²	2.95×10 ⁻³	-	-	-	-	-	-	-	-	-	-
	-	-	0.00	-	-1.54×10 ⁻⁴	3.75×10 ⁻⁵	-2.61×10 ⁻⁴	3.12×10 ⁻⁵	-1.07×10 ⁻⁴	6.94×10 ⁻⁵	-5.55×10 ⁻⁴	2.70×10 ⁻⁵
Private	-	-	0.00	-	-1.26×10 ⁻¹	1.48×10 ⁻¹	9.02×10 ⁻¹	1.09×10 ⁻¹	-1.73	2.39×10 ⁻¹	2.28	1.15×10 ⁻¹
	-1.74×10 ⁻²	2.21×10 ⁻³	-	-	-	-	-	-	-	-	-	-
	-	-	0.00	-	-1.41×10 ⁻⁴	2.94×10 ⁻⁵	-9.75×10 ⁻⁵	1.89×10 ⁻⁵	-1.12×10 ⁻⁴	4.62×10 ⁻⁵	-5.49×10 ⁻⁴	2.55×10 ⁻⁵
Business	-	-	0.00	-	-1.33	2.45×10 ⁻¹	1.56	1.23×10 ⁻¹	-4.18×10 ⁻¹	1.88×10 ⁻¹	1.34	1.41×10 ⁻¹
	-2.64×10 ⁻²	3.63×10 ⁻³	-	-	-	-	-	-	-	-	-	-
	-	-	0.00	-	-1.04×10 ⁻⁴	5.38×10 ⁻⁵	-4.55×10 ⁻⁵	2.18×10 ⁻⁵	-2.59×10 ⁻⁴	4.40×10 ⁻⁵	-5.15×10 ⁻⁴	3.43×10 ⁻⁵
Back	-	-	0.00	-	-2.59×10 ⁻¹	1.51×10 ⁻¹	5.32×10 ⁻¹	1.12×10 ⁻¹	-1.40	1.88×10 ⁻¹	2.21	1.13×10 ⁻¹
	-1.99×10 ⁻²	2.56×10 ⁻³	-	-	-	-	-	-	-	-	-	-
	-	-	0.00	-	-1.77×10 ⁻⁴	3.07×10 ⁻⁵	-1.15×10 ⁻⁴	1.94×10 ⁻⁵	-1.02×10 ⁻⁴	3.45×10 ⁻⁵	-5.70×10 ⁻⁴	2.55×10 ⁻⁵

表 4.3 パラメータ推定結果 : DIR

Purpose	Mode	Train			Bus			Car			Bike			Others		
		Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	S.E.
To work place	Intercept	-6.44×10^{-1}	4.93×10^{-2}	-1.39	3.91×10^{-2}	-3.85×10^{-1}	5.09×10^{-2}	-1.30	3.76×10^{-2}	4.97×10^{-1}	4.78×10^{-2}					
	Time	-1.32×10^{-2}	1.24×10^{-3}	-9.31×10^{-4}	9.05×10^{-4}	-3.23×10^{-2}	3.07×10^{-3}	-1.76×10^{-3}	1.49×10^{-3}	5.60×10^{-3}	1.72×10^{-3}					
	Length	4.44×10^{-5}	7.06×10^{-6}	-5.34×10^{-5}	7.35×10^{-6}	5.01×10^{-5}	1.11×10^{-5}	-4.88×10^{-5}	7.72×10^{-6}	-3.37×10^{-4}	8.86×10^{-6}					
To school	Intercept	-1.79	5.89×10^{-2}	-1.98	4.51×10^{-2}	-1.32	5.95×10^{-2}	-2.14	4.14×10^{-2}	9.96×10^{-2}	6.22×10^{-2}					
	Time	-4.53×10^{-3}	1.40×10^{-3}	1.97×10^{-4}	1.09×10^{-3}	-1.43×10^{-2}	3.58×10^{-3}	1.56×10^{-6}	1.71×10^{-3}	2.35×10^{-3}	2.44×10^{-3}					
	Length	9.06×10^{-5}	8.94×10^{-6}	-2.67×10^{-5}	8.57×10^{-6}	-5.64×10^{-5}	1.20×10^{-5}	-1.96×10^{-5}	8.58×10^{-6}	-3.22×10^{-4}	1.21×10^{-5}					
Private	Intercept	-6.17×10^{-1}	4.55×10^{-2}	-8.21×10^{-1}	4.18×10^{-2}	7.28×10^{-1}	4.61×10^{-2}	-1.33	3.85×10^{-2}	1.38	4.43×10^{-2}					
	Time	-1.31×10^{-2}	9.57×10^{-4}	-2.49×10^{-3}	1.08×10^{-3}	-3.83×10^{-2}	2.36×10^{-3}	-1.48×10^{-3}	1.66×10^{-3}	3.30×10^{-3}	1.39×10^{-3}					
	Length	-2.14×10^{-6}	6.44×10^{-6}	-1.02×10^{-4}	7.30×10^{-6}	-3.49×10^{-5}	8.12×10^{-6}	-9.17×10^{-5}	7.96×10^{-6}	-4.61×10^{-4}	8.60×10^{-6}					
Business	Intercept	-1.38	5.71×10^{-2}	-1.99	4.33×10^{-2}	3.13×10^{-1}	5.82×10^{-2}	-1.76	3.82×10^{-2}	-8.58×10^{-1}	5.06×10^{-2}					
	Time	-1.20×10^{-2}	1.41×10^{-3}	-2.19×10^{-4}	1.11×10^{-3}	-2.75×10^{-2}	3.56×10^{-3}	9.82×10^{-4}	1.49×10^{-3}	4.46×10^{-3}	2.10×10^{-3}					
	Length	-1.39×10^{-7}	7.59×10^{-6}	-3.03×10^{-5}	7.76×10^{-6}	7.03×10^{-6}	1.17×10^{-5}	-6.70×10^{-5}	7.98×10^{-6}	-2.21×10^{-4}	9.30×10^{-6}					
Back	Intercept	-4.56×10^{-1}	4.95×10^{-2}	-9.35×10^{-1}	3.83×10^{-2}	6.24×10^{-1}	5.05×10^{-2}	-1.11	3.80×10^{-2}	1.48	4.62×10^{-2}					
	Time	-1.40×10^{-2}	1.25×10^{-3}	-5.38×10^{-4}	8.45×10^{-4}	-4.79×10^{-2}	2.97×10^{-3}	-3.24×10^{-3}	1.58×10^{-3}	6.85×10^{-4}	1.61×10^{-3}					
	Length	6.30×10^{-5}	6.85×10^{-6}	-7.89×10^{-5}	7.14×10^{-6}	3.35×10^{-5}	1.00×10^{-5}	-5.40×10^{-5}	7.66×10^{-6}	-4.17×10^{-4}	8.67×10^{-6}					

表 4.4 パラメータ推定結果 : CoDA

Trip Purpose	Mode	Train Coef.	Bus Coef.	Car Coef.	Bike Coef.	Others Coef.
To work place	Intercept	0.184	0.100	0.112	0.095	0.508
	ilr (Xtrain)	0.320	0.196	0.129	0.210	0.145
	ilr (Xbus)	0.384	0.210	0.068	0.193	0.144
	ilr (Xcar)	0.211	0.193	0.209	0.193	0.194
	ilr (Xbike)	0.203	0.182	0.232	0.178	0.205
	Length	0.200	0.200	0.200	0.200	0.200
To school	Intercept	0.140	0.147	0.137	0.161	0.415
	ilr (Xtrain)	0.221	0.217	0.221	0.204	0.137
	ilr (Xbus)	0.232	0.229	0.212	0.211	0.116
	ilr (Xcar)	0.205	0.199	0.201	0.190	0.206
	ilr (Xbike)	0.202	0.197	0.200	0.194	0.207
	Length	0.200	0.200	0.200	0.200	0.200
Private	Intercept	0.073	0.054	0.176	0.038	0.659
	ilr (Xtrain)	0.228	0.214	0.173	0.221	0.164
	ilr (Xbus)	0.299	0.222	0.086	0.240	0.153
	ilr (Xcar)	0.199	0.185	0.233	0.171	0.212
	ilr (Xbike)	0.167	0.172	0.282	0.148	0.231
	Length	0.200	0.200	0.200	0.200	0.200
Business	Intercept	0.170	0.136	0.329	0.132	0.233
	ilr (Xtrain)	0.214	0.193	0.186	0.200	0.207
	ilr (Xbus)	0.231	0.182	0.156	0.196	0.235
	ilr (Xcar)	0.195	0.200	0.219	0.193	0.193
	ilr (Xbike)	0.191	0.193	0.241	0.191	0.184
	Length	0.200	0.200	0.200	0.200	0.200
Back	Intercept	0.125	0.072	0.108	0.050	0.645
	ilr (Xtrain)	0.374	0.206	0.104	0.198	0.117
	ilr (Xbus)	0.453	0.215	0.044	0.192	0.096
	ilr (Xcar)	0.214	0.194	0.210	0.184	0.198
	ilr (Xbike)	0.191	0.175	0.244	0.161	0.229
	Length	0.200	0.200	0.200	0.200	0.200

4.2 精度評価結果

第 3.3 節で述べた 3 つの尺度 (R_T^2 , R_A^2 , $KL\ div.$) をそれぞれのモデルについて評価し比較する。その結果を表 4.5 に示す。各モデルで最も良い値を示す結果を太字で表す。表 4.5 より、DIR が一部で最良であるものの、多くの場合で MNCL が最良の結果にあることがわかる。ゼロ値率も合わせて結果を見てみると、CoDA はゼロ値率が大きくなるほど結果が悪化していることも確認できる。

表 4.5 精度評価結果

		Tokyo MA			Rate of 0s
Purpose		R_T^2	R_A^2	$KL\ div.$	
To work place	AML	0.003	0.321	1571.282	27.28
	MNCL	0.379	0.377	1409.675	
	DIR	0.172	0.369	1429.145	
	CoDA	0.174	0.299	1500.553	
To school	AML	0.002	-0.881	2050.681	59.96
	MNCL	0.270	-0.262	1700.765	
	DIR	0.136	0.434	2042.531	
	CoDA	0.079	0.308	2450.485	
Private	AML	0.003	0.473	1548.335	24.74
	MNCL	0.291	0.547	1265.112	
	DIR	0.207	0.491	1347.265	
	CoDA	0.222	0.463	1439.185	
Business	AML	0.001	-0.065	2542.396	56.9
	MNCL	0.417	0.183	1463.168	
	DIR	0.069	0.596	1720.234	
	CoDA	0.024	0.280	2823.069	
Back	AML	0.005	0.470	1257.041	17.44
	MNCL	0.304	0.522	1079.314	
	DIR	0.208	0.470	1124.695	
	CoDA	0.262	0.493	1084.849	

4.3 ゼロ値率と精度評価の関係

前節の結果から、ゼロ値率の大小と当てはまりの良さの間に関係があることが示唆された。本節では、この関係をさらに調べるため、第 3.4 節で述べたゼロ値率を変動させるシミュレーションの結果を説明する。各精度指標の推移結果を図 4.1 (R_T^2)、図 4.2 (R_A^2)、図 4.3 ($KL\ div$) に示す。ここで、分析に用いたトリップ目的は **Private** とした。

図 4.1 から 4.3 から、ゼロ値率がほとんどない場合は、CoDA が最良のモデルであることが読み取れる。また、ゼロ値率が高くなるにつれて、各指標の値は悪化し、ゼロ値率が約 10% になると最良のモデルが CoDA から MNCL に変わる傾向があることが読み取れる。この分析から、比説明変数のゼロ値率が各モデルの精度指標に影響を与える可能性が示唆され、比較対象とした 4 つのモデルの中では、最良のモデルに置き換わるボーダーラインがゼロ値率約 10% のあたりにあることが確認できた。

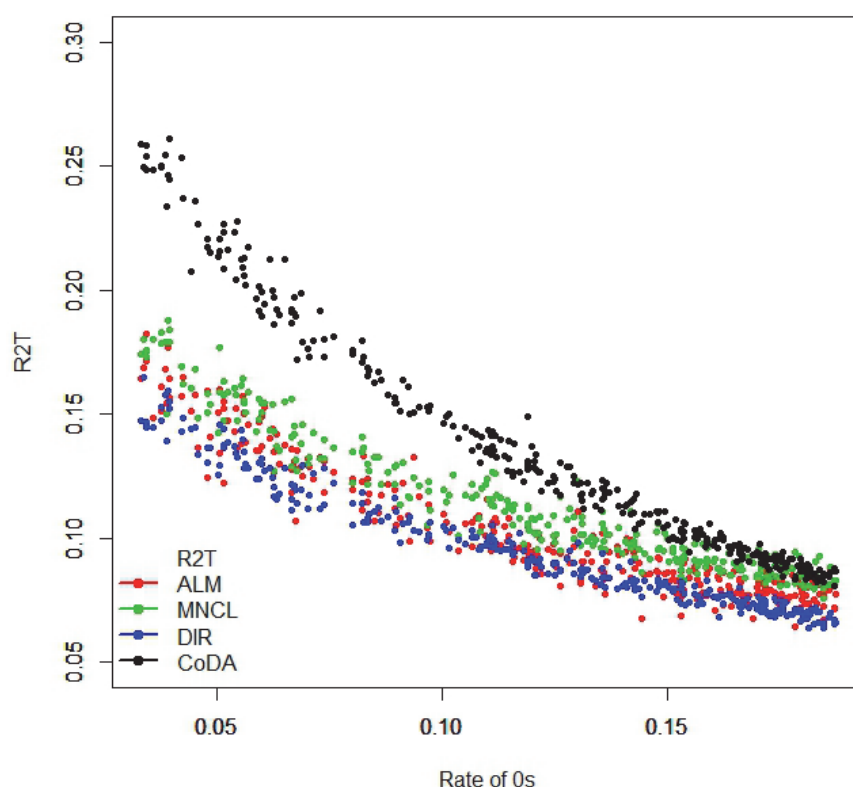


図 4.1 ゼロ値率と R_T^2 の関係

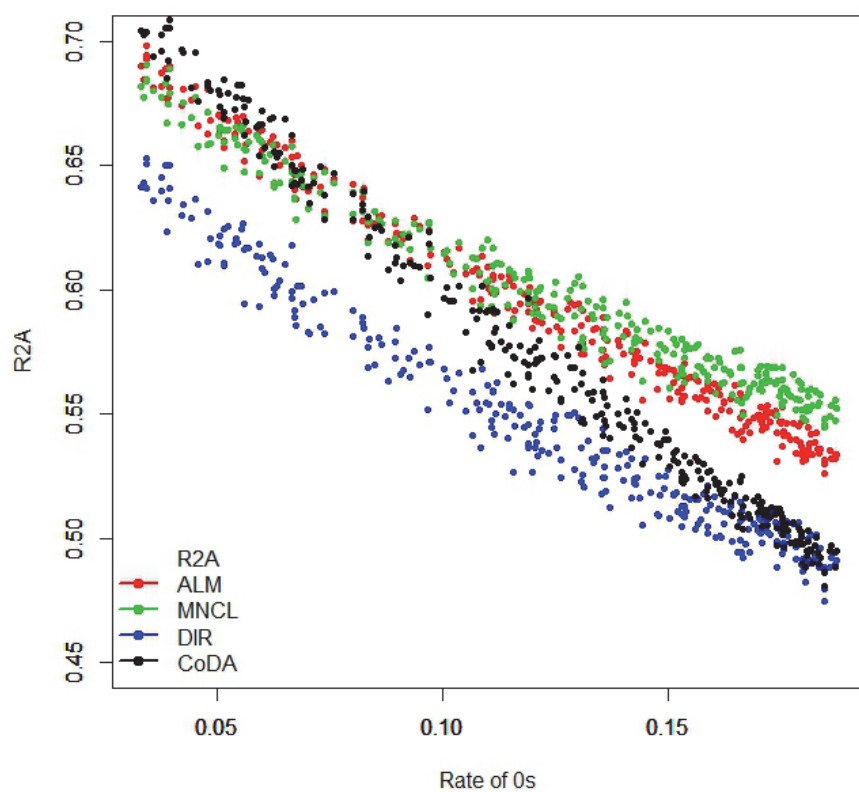


図 4.2 ゼロ値率と R^2_A の関係

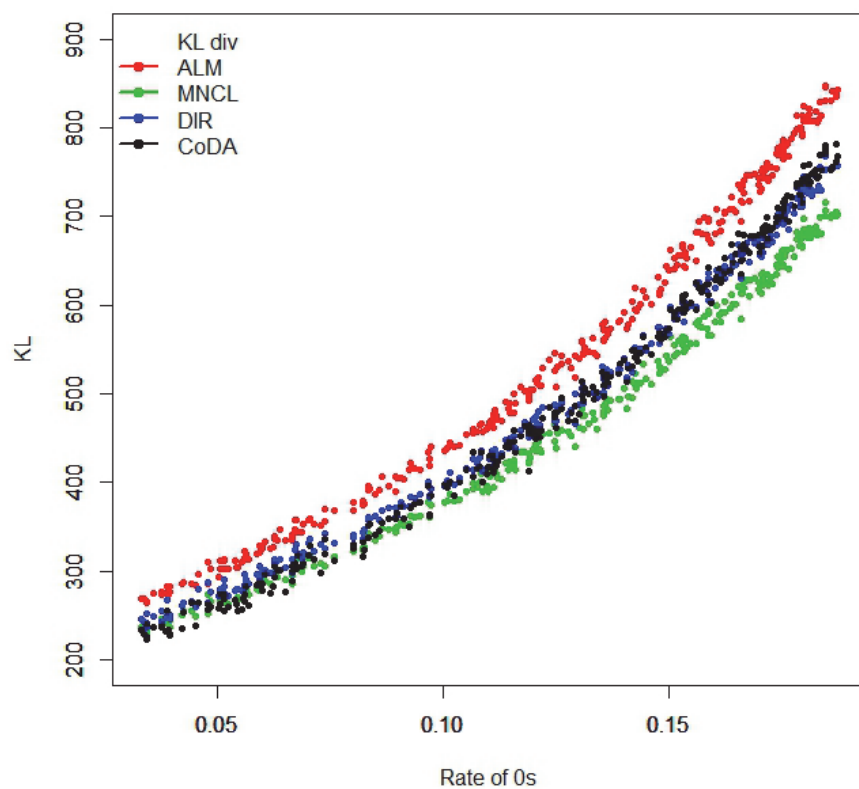


図 4.3 ゼロ値率と KL div.の関係

5 おわりに

本稿では、CoDA の知見を援用して、土木計画学分野で用いられてきた「集計モデル」の再考を行った。まず、同分野における「集計モデル」の現在の位置付けの認識を述べ、そして CoDA という「集計モデル」に焦点を当てた方法論の発展が地質学分野で興っていることを説明した。続いて、CoDA で検討されてきた様々な対数比変換の観点からみると、同分野で用いられてきた集計ロジットモデルやマーケティング分野で用いられてきた MCI モデルが俯瞰的に整理できることを説明した。実証分析の例として、通常ゼロ値が含まないデータを対象としている CoDA の方法論を、土木計画学分野で用いられる代表的な組成データであり、かつゼロ値を含みうる交通機関分担率データに適用し、既往モデルとの比較結果を報告した。興味深い結果として、ゼロ値率に応じて当てはまりの良さが最良のモデルが変わることを示した。この理由の解析的な分析と、そもそも比率を扱う集計モデルにおけるゼロ値の取り扱いについては今後の主要な課題である。

参考文献

- 1) Aitchison, J. (1982) “The statistical analysis of compositional data”, *Journal of the Royal Statistical Society Series B (Methodological)*, Vol. 44, No. 2, pp. 139–177.
- 2) Aitchison, J (1986) *The Statistical Analysis of Compositional Data*. Chapman & Hall.
- 3) Aitchison, J. and Egozcue, JJ (2005) “Compositional Data Analysis: Where Are We and Where Should We Be Heading?”, *Mathematical Geology*, Vol. 27, pp. 829–850.
- 4) Ben-Akiva, M and Lerman, SR (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, The MIT Press.
- 5) Berkson, J (1955) “Maximum likelihood and minimum chi-square estimations of the logistic function”, *Journal of the American Statistical Association*, Vol. 50, pp. 130–161.
- 6) Cameron AC and Trivedi PK (2005) *Microeconometrics Methods and Applications*, Cambridge University Press.
- 7) Filzmoser, P., Hron, K, and Temp, M (2018) *Applied Compositional Data Analysis with Worked Examples in R*. Springer.
- 8) McFadden, D (1974) “The measurement of urban travel demand.” *Journal of Public Economics*, Vol. 3, No. 4, pp. 303–328.
- 9) Martín-Fernández, J (2019) “Comments on: Compositional data: the sample space and its structure”, *TEST*, Vol. 28, No. 3, pp. 653–657
- 10) Martín-Fernández, J, Hron, K, Templ, M, Filzmoser, P, and Palarea-Albaladejo, J (2012) “Model-based replacement of rounded zeros in compositional data: classical and robust approaches”. *Computational Statistics & Data Analysis*, Vol. 56, No. 9, pp. 2688–2704.
- 11) Morais, J, Thomas-Agnan, C, and Simioni, M (2017) “Using compositional and Dirichlet models for market share regression.” *Journal of Applied Statistics*, Vol. 45, No. 9, pp. 1670–1689.
- 12) Morais, J, Thomas-Agnan, C, and Simioni, M (2018) “Interpretation of explanatory variables impacts in compositional regression models.” *Austrian Journal of Statistics*, Vol. 47, No. 5, pp.

1–25.

- 13) Morais, J, and Thomas-Agnan, C (2021) “Impact of covariates in compositional models and simplicial derivatives.” *Austrian Journal of Statistics*, Vol. 50, No. 2, pp. 1–15.
- 14) Nakanishi, M. and Cooper LG (1974) “Parameter estimation for a multiplicative competitive interaction model—least squares approach”. *Journal of Marketing Research*, Vol. 11, No. 3, pp. 303–311.
- 15) Theil, H (1969) “A multinomial extension of the linear logit model”, *International Economic Review*, Vol. 10, pp. 251–259.
- 16) 土木学会土木計画学研究委員会（1995）*非集計モデルの理論と実際*、土木学会.

日交研シリーズ目録は、日交研ホームページ
http://www.nikkoken.or.jp/publication_A.html を参照してください

A-920 「集計モデルの実証比較：組成データ解析の観点から」

輸送ネットワークと都市の空間構造の研究プロジェクト

2025 年 6 月 発行

公益社団法人日本交通政策研究会