スーパーの店舗分布の二点相関関数

Inuverse

2026 年 11 月 14 日

要旨

店舗の並び方を宇宙論で銀河の分布を測るのに使う二点相関関数で調べた。オーストラリアの Coles（865 店）と Woolworths（1,116 店）、米国の Walmart（4,564 店）は OpenStreetMap から、日本のトライアル（375 店）、西友（243 店）、ドラッグストアのコスモス（1,735 店）は公式サイトから座標を取った。店どうしに相関がない場合の基準には、人口に比例して点を撒いたランダムカタログを使った。

結果は三つの問いに分けて示す。

人口で説明できるか。Coles と Woolworths の並び方は人口の分布でほぼ説明でき、この基準に対する Landy–Szalay 推定量は 3 km 以上で 0 と合う。トライアル、コスモス、Walmart は人の密集地に店が少なく、西友は逆に多い。この違いの大半は出店している地方の偏りによるが、西友は地方の中でも密集地に寄り、Walmart は避けている。

地方を取り除いても固まるか。出店している地方を選択関数としてランダムカタログに入れ、店を選択関数に従って独立に撒いた模擬データ 100 個と比べた。トライアルの店は 3–25 km の距離で互いに固まり、8–20 km の値は模擬データのどれよりも大きい。西友の店は 8 km より短い距離で、密集地への寄り方もそろえた模擬データのどれよりも強く固まっている。コスモスは全国では強く固まるが、九州の中だけで測ると固まりの大半が消え、トライアルも同じである。Walmart は 8–50 km では弱く固まり、8 km より短い距離ではトライアルとは逆に店どうしが離れている。

二社の関係。トライアルと西友の相互相関は、出店している地方を取り除いても 8–30 km で -0.2 前後と負になる。ただし、トライアルの店を実際の位置のまま残して西友だけを撒き直すと、100 個のうち 10 個が実際の値以下になり、二社が互いを避けて店を置いていると考える必要はない。トライアルとコスモスの相互相関は、全国では正、九州の中だけでは負になる。

店が数百しかないときは jackknife の共分散から求めた χ^2 の上側確率が信用できないので、有意かどうかは模擬データと比べて判断した。付録では店の数を固定したときのデータ同士の組の数の分散を導いて Lean で検証し、人口の濃淡が分散を大きくするが、Landy–Szalay 推定量ではその分がほぼ打ち消されることを示した。

目次

1 序論 4

2 枠組み 5
2.1 二点相関関数 5
2.2 推定量 7
2.3 ランダムカタログ 9
2.4 選択関数 9
2.5 誤差の見積もり 11

3 データ 16
3.1 オーストラリア 16
3.2 日本 18
3.3 米国 22

4 結果 24
4.1 人口で説明できるか 24
4.2 地方を取り除いても固まるか 27
4.3 二社の関係 34

5 結論と議論 35

付録 A toy model 40

付録 B jackknife の領域 41

付録 C 選択関数を 10 km 格子の表で求める 41

付録 D DD の分散 43
D.1 設定 43
D.2 店の数を固定した場合 44
D.3 店の数もポアソン分布で揺らした場合 44
D.4 Lean での検証 45
D.5 toy model での検算 46
D.6 オーストラリアでの値 46

付録 E 推定量の恒等式 47

付録 F　結果の確かめ 48
F.1　設定を変えた計算 48
F.2　模擬データの分布の形 49

付録 G　再現の手順 49

表 1　比べるチェーンと、トライアルから見た位置。店の数は本稿で使った数。

役割	チェーン	店の数	トライアルから見た位置
基準	トライアル	375	九州が地盤のスーパーセンター
内側の横	西友	243	2025 年に子会社にした会社
外側の横	コスモス	1,735	九州で地盤が重なる競合（ドラッグストア）
世界の大手	Walmart	4,564	世界規模のスーパーセンター
市場の構造	Coles・Woolworths	865・1,116	二社寡占が進んだ市場 [1]

1　序論

オーストラリアの食品スーパーは、Coles と Woolworths の二社が大きな割合を占める。オーストラリア競争・消費者委員会（ACCC）の推計では、スーパーでの食料品の売上のうち Woolworths が 38%、Coles が 29% で、二社で約三分の二になる。ACCC はこれを寡占の市場構造と呼んでいる [1]。現地で聞いた話では、商品の入れ替わりは日本よりずっと遅く、物流拠点から店舗までの距離は日本より長い。これを調べるうちに、日本とオーストラリアで店舗の並び方が統計的に同じかどうかが気になった。

店舗の並び方を数で表すため、店舗を銀河に見立てる。宇宙論では、銀河がどれだけ集まっているかを二点相関関数で測る。宇宙には大規模な構造があり、銀河が集中している領域もあれば、銀河がほとんどない空っぽの領域もある。二点相関関数はこうした集まり方を距離ごとの数で表す。本稿では二点相関関数の推定量を、オーストラリアの Coles と Woolworths、日本のトライアルと西友の店舗の座標に当てる。日本では比べる相手として、トライアルと同じく九州が地盤のドラッグストア、コスモスも測る。さらに、スーパーセンターを中心に店を持つ大手として、米国の Walmart も測る。

本稿の問いは、店舗の並び方が国やチェーンによって違うかどうかである。比べる相手は、トライアルを真ん中に置き、役割で選んだ（表 1）。トライアルは 2025 年に西友を子会社にしたので、日本では二社の店が近くに並ぶか、離れて並ぶかも相互相関関数で調べる。

日本の二社は、地盤とする地方が違う。トライアルは 1974 年に福岡市で開業したリサイクルショップを前身とし、1992 年に福岡県でディスカウントストアを、1996 年に北九州市でスーパーセンターを開いた [2]。九州を中心に店を広げ、2001 年からは全国に出店を進めてきた [2, 3]。西友は関東を中心に、駅に近い店を多く持つ [3]。トライアルは西友を子会社にすることで、関東、中部、関西の人口の多い地方での足場を早く固めることを見込んでいる [3]。したがって、二社の店が地方ごとに違う所に多いことは、初めから見込まれる。本稿で調べるのは、そうした地方の違いを取り除いても残る並び方の違いである。

店舗の座標をそのまま数えると、店の置き方とは関係のない理由でも、店が集まっているように見える。海の上や人の住まない山奥には、どのチェーンも店を出さない。そのため店は陸地の中でも人の多い所に固まって見える。これはチェーンの店の置き方の性質ではなく、地図と人口の性質

である。

銀河の観測にも同じ問題がある。天の川に隠れた方向の銀河は見えにくく、暗い銀河まで見えるかどうかも方向によって違う。そのため、数えた銀河の分布には観測の都合が混ざる。宇宙論では、観測と同じ条件で相関のない点（ランダムカタログ）を撒き、銀河と比べることで、この観測の都合を取り除いてきた。店舗にも同じ方法が使える。陸地だけにランダムな点を撒いて比べれば、海岸線の形で説明できる分が取り除ける。人口に比例して撒いて比べれば、人の分布で説明できる分も取り除ける。

同じ道具は、空間統計学の空間点過程の分野では別の名前で知られている。二点相関関数 ξ は対相関関数 $g = 1 + \xi$ と同じもので、その積分が Ripley の K 関数である [4]。人口のように場所によって濃さの違う基準と比べることは非一様な K 関数 [5] に、二社の相互相関は種類つきの点の cross-K 関数に、帰無仮説の模擬データとの比較はモンテカルロの包絡線検定に当たる。経済地理学では、Duranton と Overman [6] が事業所の間の距離の分布を、全事業所から無作為に選んだ模擬データと比べて業種の集積を判定しており、考え方は本稿とほぼ同じである。こうした距離に基づく指標は Marcon と Puech [7] が整理している。本稿は宇宙論の推定量と選択関数の扱いをそのまま持ち込み、チェーンどうしを比べる。

本稿の構成は次のとおりである。2 節では二点相関関数と推定量、ランダムカタログ、選択関数、誤差の見積もりをまとめる。3 節ではオーストラリア、日本、米国の店舗と人口のデータを述べる。4 節では結果を三つの問い（人口で説明できるか、出店している地方を取り除いても固まるか、二社の関係）の順に示す。5 節では結論と限界を述べる。付録 F では設定を変えた計算と模擬データの分布の形を確かめる。付録 A では相関のない点に推定量を当てた toy model を示す。付録 D ではデータ同士の組の数 DD の分散を、店の数を固定した場合とポアソン分布で揺らした場合とで導き、Lean で検証する。付録 E では四つの推定量の関係を式で示す。

2　枠組み

2.1　二点相関関数

二点相関関数はある点から距離 r の所にほかの点が、でたらめに置いたときより多いか少ないかを表す。図 1 の点で考える。どれか一つの点を選び、そのまわりに半径 r の円を描いて、円の中のほかの点を数える。点がでたらめに置かれていれば（図 1(b)）、円の中の点の数は平均すると点の密度と円の面積の積になる。点が固まっていれば（図 1(a)）、選んだ点のまわりにはほかの点も多いので、円の中の点はこれより多い。点が間隔をあけて並んでいれば（図 1(c)）、すぐ近くにはほかの点がないので、円の中の点は少ない。二点相関関数はこの多い・少ないを、でたらめに置いたときに比べて何割多いかで表す。距離ごとに見るため、実際には円の中ではなく、半径 r と $r + \delta r$ のあいだの細い輪の中を数える。店舗でいえば、ある店から距離 r の所に同じチェーンのほかの店が、でたらめに置いたときより多いかどうかである。

でたらめに置くことを、式ではポアソン過程で表す [8]。面密度 n のポアソン過程は次の二つの

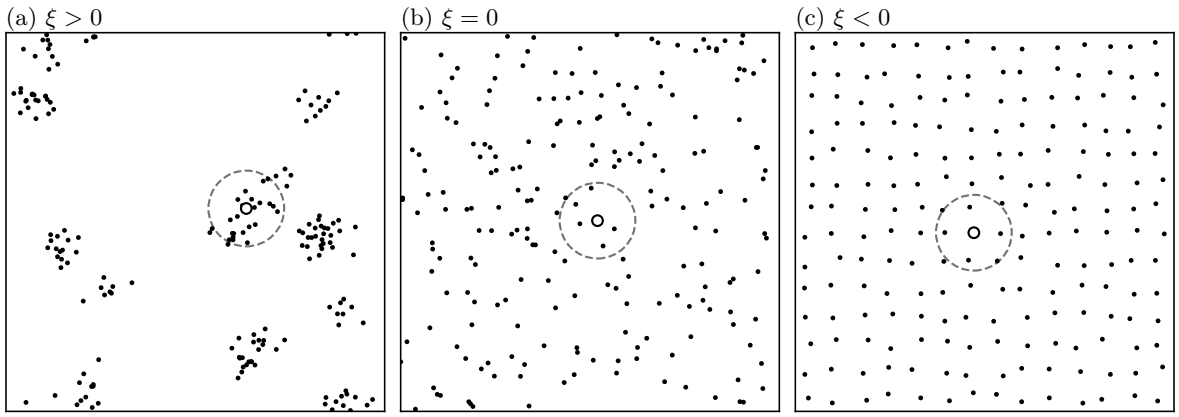


図 1　二点相関関数の符号と点の並び方。(a) 固まった点。(b) 互いに独立に、一様に置いた点（ポアソン過程）。(c) 間隔をあけて並んだ点。白丸の点のまわりの破線の円は、半径 r（正方形の辺の 0.1 倍）の円である。円の中のほかの点は、(b) に比べて (a) では多く、(c) では少ない。ξ の符号は、円の大きさ程度の距離での値を指す。説明のための絵で、ξ は測っていない。

性質をもつ点の置き方である。一つ目に、面積 S の領域に入る点の数は、平均 nS のポアソン分布に従う。二つ目に、重ならない領域に入る点の数は、互いに独立である。二つ目の性質が、点がほかの点の位置に左右されないことを表す。店でいえば、ほかの店がどこにあるかを気にせず、どこにでも同じ確率で店を出すことに当たる。

点の置き方がポアソン過程とどれだけ違うかを、次のように表す。面密度 n の点の置き方で、距離 r だけ離れた面積要素 δS_1, δS_2 に一点ずつ見つかる確率を

$$\delta P = n^2\big[1 + \xi(r)\big]\,\delta S_1\,\delta S_2 \tag{1}$$

と書き、$\xi(r)$ を二点相関関数と呼ぶ [8]。ここでは、点の置き方が場所にも向きにもよらず、ξ は距離 r だけで決まるとしている。

ポアソン過程では $\xi = 0$ になる。小さな面積 δS に点が一つだけ入る確率は、一つ目の性質から $n\,\delta S\,e^{-n\,\delta S} = n\,\delta S + O(\delta S^2)$ である。δS_1 と δS_2 は重ならないので、二つ目の性質から、二つの確率は掛け算になる。

$$\delta P = (n\,\delta S_1)(n\,\delta S_2) = n^2\,\delta S_1\,\delta S_2. \tag{2}$$

式 (1) と比べると、すべての r で $\xi(r) = 0$ である。逆は成り立たない。$\xi = 0$ でも、三つ以上の点の組に相関があれば、ポアソン過程とは限らない。二点相関関数が見るのは、二つの点の組の数だけである。

式 (1) は、最初に述べた円の中の点の数と次のようにつながる。δS_1 に点が一つあるという条件のもとで、δS_2 にもう一つの点がある確率は、δP を $n\,\delta S_1$ で割って $n[1 + \xi(r)]\,\delta S_2$ である。δS_2 を半径 r と $r + \delta r$ のあいだの輪（面積 $2\pi r\,\delta r$）にとると、輪の中のほかの点の数の平均は

$$\langle N(r)\rangle = n\big[1 + \xi(r)\big]\,2\pi r\,\delta r \tag{3}$$

になる。ポアソン過程なら $n\,2\pi r\,\delta r$ なので、$\xi(r)$ は、輪の中の点の数がポアソン過程より何割多いかを表す。

$\xi(r) > 0$ なら、距離 r だけ離れた店舗の組が、独立に撒いたときより多い。店舗がその距離で集まっているということである。$\xi(r) < 0$ なら、その距離の組は少なく、店舗は互いに離れて並ぶ。

2.2 推定量

式 (1) は点の置き方を決める確率である。実際に手にあるのは、限られた領域に置かれた店の座標の一組だけである。そこで、店の組の数を数え、その期待値を式 (1) で書いて、ξ を取り出す。

店は面積 S の領域 W（陸地）の中に n_d 店あるとし、平均の面密度を n_d/S とする。距離のビン b（たとえば 10–13 km）に入る店の組の数を DD と書く。式 (1) を W の中の二点について積分すると、その期待値は

$$\langle DD\rangle = \frac{n_d(n_d-1)}{2}\,\frac{1}{S^2}\int_W\!\!\int_W \Theta_b(r_{12})\,\bigl[1+\xi(r_{12})\bigr]\,dS_1\,dS_2 \tag{4}$$

になる。r_{12} は二点の距離で、Θ_b は r_{12} がビン b に入れば 1、入らなければ 0 をとる。$\xi = 0$ でも、この積分は Θ_b の積分

$$A_b = \int_W\!\!\int_W \Theta_b(r_{12})\,dS_1\,dS_2 \tag{5}$$

を残す。A_b は領域 W の形で決まる。海岸の近くの店では、距離 r の相手の候補の一部が海に落ちるので、A_b は同じ面積の正方形より小さくなる。そのため、店の組の数を数えるだけでは、組が多いのが店の置き方のせいか、領域の形のせいかを分けられない。

そこで、同じ領域 W に相関のない点を n_r 個撒き、その組の数と比べる。この点の集まりをランダムカタログと呼ぶ。距離のビンごとに、三種類の組の数を数える。データ同士の組の数が DD、データとランダムの組の数が DR、ランダム同士の組の数が RR である[*1]。組の総数はそれぞれ $n_d(n_d-1)/2$、n_dn_r、$n_r(n_r-1)/2$ と違うので、総数で割った dd、dr、rr を使う。式 (4) と同じ計算で、期待値は

$$\langle dd\rangle = \frac{A_b}{S^2}\bigl(1+\bar{\xi}_b\bigr), \qquad \langle dr\rangle = \langle rr\rangle = \frac{A_b}{S^2} \tag{6}$$

になる。ここで $\bar{\xi}_b$ はビン b に入る組について平均した ξ である。

$$\bar{\xi}_b = \frac{1}{A_b}\int_W\!\!\int_W \Theta_b(r_{12})\,\xi(r_{12})\,dS_1\,dS_2. \tag{7}$$

dr の期待値に ξ が出ないのは、ランダムの点が店と相関をもたないからである。式 (6) の比をとると、領域の形 A_b が消えて

$$\frac{\langle dd\rangle}{\langle rr\rangle} - 1 = \bar{\xi}_b \tag{8}$$

[*1] DR は、実際の店のまわりにランダムの点がいくつあるかを数える。海岸の近くにある店ほど、距離 r の所のランダムの点は少ない。DR で割ったり DR を引いたりすると、実際の店が置かれている場所から見た領域の形の効果を、その店の位置のまま取り除ける。RR はランダムの点どうしなので、領域全体で平均した形の効果しか表さない。

が残る。ランダムの点を増やすと rr と dr はそれぞれの期待値に近づき、ばらつきは $n_r^{-1/2}$ で小さくなる。したがって、ランダムの点が十分多い極限で、次の推定量の期待値はビンで平均した ξ になる。ここでは、店の一点ずつの密度が W の中で一様だとしている。人口に比例してランダムカタログを撒くときは、この一様を人口に比例する密度に置き換える（2.3 節）。

比べるため、推定量は次の 4 つを実装した。

$$\hat{\xi}_{\mathrm{N}} = \frac{dd}{rr} - 1, \qquad \hat{\xi}_{\mathrm{DP}} = \frac{dd}{dr} - 1, \tag{9}$$

$$\hat{\xi}_{\mathrm{H}} = \frac{dd\,rr}{dr^2} - 1, \qquad \hat{\xi}_{\mathrm{LS}} = \frac{dd - 2\,dr + rr}{rr}. \tag{10}$$

$\hat{\xi}_{\mathrm{DP}}$ は Davis と Peebles [9]、$\hat{\xi}_{\mathrm{H}}$ は Hamilton [10]、$\hat{\xi}_{\mathrm{LS}}$ は Landy と Szalay [11] による。実装は `crates/retail-cosmology/src/lib.rs` にある。本文の数値と図表を作るコマンドは付録 G にまとめた。toy model で、相関のない点に当てると 0 のまわりに散ることを確かめた（付録 A）。

四つの推定量に式 (6) の期待値を入れると、どれも $\bar{\xi}_b$ になる。ただし二つの近似がある。一つ目に、$\hat{\xi}_{\mathrm{DP}}$ と $\hat{\xi}_{\mathrm{H}}$ はデータから数えた dd と dr の比なので、比の期待値を期待値の比で置き換えた差が、データの揺れの 2 乗の大きさで残る。$\hat{\xi}_{\mathrm{N}}$ と $\hat{\xi}_{\mathrm{LS}}$ は、rr を期待値に置き換えれば dd と dr の 1 次式なので、この差はない。二つ目に、平均の面密度をデータ自身の n_d/S で決めているので、領域全体で積分した ξ が 0 に近づくように偏る（積分の制約）。本稿ではこの偏りを補正していない。

極限では同じ量を測っているが、店の数が限られているときの揺れ方と、領域の端の効果の入り方が違う。natural の推定量 $\hat{\xi}_{\mathrm{N}}$ は上の比をそのまま使う、いちばん素直な形である。$\hat{\xi}_{\mathrm{DP}}$ は RR の代わりに DR で割り、実際の店の位置から見た領域の形を使う。$\hat{\xi}_{\mathrm{H}}$ と $\hat{\xi}_{\mathrm{LS}}$ は DD、DR、RR の三つを組み合わせる。Hamilton は、平均の密度の見積もりの誤差が $\hat{\xi}_{\mathrm{H}}$ には 2 次でしか入らないことを示した [10]。Landy と Szalay は、相関が弱いとき $\hat{\xi}_{\mathrm{LS}}$ の分散がポアソン過程の値まで小さくなり、端の効果による偏りも小さいことを示した [11]。本稿では主に $\hat{\xi}_{\mathrm{LS}}$ を使い、ほかの三つは結果が推定量の選び方に左右されないかを見るために使う。四つの推定量の違いは付録 E で $\hat{\xi}_{\mathrm{N}}$ と $dr/rr - 1$ の二つの量で正確に書いた。

二つのチェーンの店がどれだけ近くに並ぶかは、相互相関関数 $\xi_\times(r)$ で表す。式 (1) で、一点目を一方のチェーンの店、二点目をもう一方のチェーンの店に限ったものである。$\xi_\times > 0$ なら二社の店は近くに並び、$\xi_\times < 0$ なら互いに離れて並ぶ。推定量には Landy–Szalay の形を二種類の点に広げたものを使った。

$$\hat{\xi}_\times = \frac{d_1 d_2 - d_1 r_2 - d_2 r_1 + r_1 r_2}{r_1 r_2}. \tag{11}$$

添字の 1 と 2 は二つのチェーンを、r_1 と r_2 はそれぞれのランダムカタログに関する量を表す。たとえば $d_1 r_2$ はチェーン 1 の店とチェーン 2 のランダムの点の組の数を、組の総数 $n_{d_1} n_{r_2}$ で割ったものである。ランダムカタログをチェーンごとに分けるのは、出店している地方が二社で違うときに、それぞれの選択関数（2.4 節）を入れるためである。二社が同じランダムカタログを使うなら、式 (11) は $(d_1 d_2 - d_1 r - d_2 r + rr)/rr$ になる。

2.3　ランダムカタログ

ランダムカタログは「店どうしに相関がなかったら、店はどこにあるはずか」という基準である。$\hat{\xi}$ はこの基準からのずれを測るので、どんな基準を選ぶかで、$\hat{\xi}$ に何が出るかが決まる。ここでは二つの基準を使った。

一つ目は陸地に一様に撒いたランダムカタログである（図 2(a)）。宇宙論のランダムカタログが観測の範囲だけを表すのと同じ考え方で、この基準が知っているのは陸地の形だけである。人がどこに住んでいるかは知らない。そのため、店が人の多い所にあることも、店どうしが固まることも、区別されずにまとめて $\hat{\xi}$ に出る。4 節で見るように、オーストラリアの店の $\hat{\xi}$ はこの基準では人口そのものの $\hat{\xi}$ とほぼ重なる。これは、店の固まり方の大部分が人口の固まり方と同じ形をしているということで、店が人の多い所に出ているのか、人が店の多い所へ通っているのかは、この比較では分けられない。

二つ目は人口に比例して撒いたランダムカタログである（図 2(b)）。この基準は「店が人口に比例して、ほかの店と無関係に置かれていたら」という置き方を表す。人の分布で説明できる分は $\hat{\xi}$ から消え、人がいる所に対して店がさらに固まるか、間隔をあけて並ぶかが残る。日本と比べるときは、この二つ目の基準の方が適している。海岸線の形や人の住む場所の違いが、基準の方に吸収されるからである。

ランダムカタログは人口の格子（1 辺 1 km）を一つ選び、その格子の中に一様に点を一つ置く、という操作を繰り返して作る。格子 k を選ぶ確率は格子の重み w_k に比例させる。

$$P(k) = \frac{w_k}{\sum_j w_j}. \tag{12}$$

和は陸地の格子すべてについてとる。格子 k の人口を N_k とすると、重みは基準ごとに次のとおりである。陸地に一様な基準ではすべての格子で $w_k = 1$ である。人が住む格子に一様な基準では $N_k > 0$ の格子で $w_k = 1$、それ以外で $w_k = 0$ である。人口に比例する基準では $w_k = N_k$ である。人が住む格子に一様な基準は、陸地に一様な基準が作れない日本で、その代わりに使う（3.2 節）。どの基準でも、ランダムカタログは 20,000 点である。この点の数は格子の数ではなく、撒いた点の数である。

人口がどれだけ固まっているかも、店と同じ物差しで測った。人口に比例して 2,000 点を撒き（図 2(c)）、これを店の代わりのデータとして、陸地に一様なランダムカタログと比べた。あるチェーンがほかの店のことを気にせず、人の数に比例して店を出したとすると、店の並びはこの点のようになる。この点の $\hat{\xi}$ が人口だけで生じる固まり方である。距離のビンは 2–2000 km を対数で 15 等分した。

2.4　選択関数

人口に比例するランダムカタログでも、あるチェーンが一部の地方にしか出店していなければ、$\hat{\xi}_{\mathrm{LS}}$ にはその地方の偏りが出る（トライアルの例は 4.1 節）。店どうしの間隔の取り方だけを見る

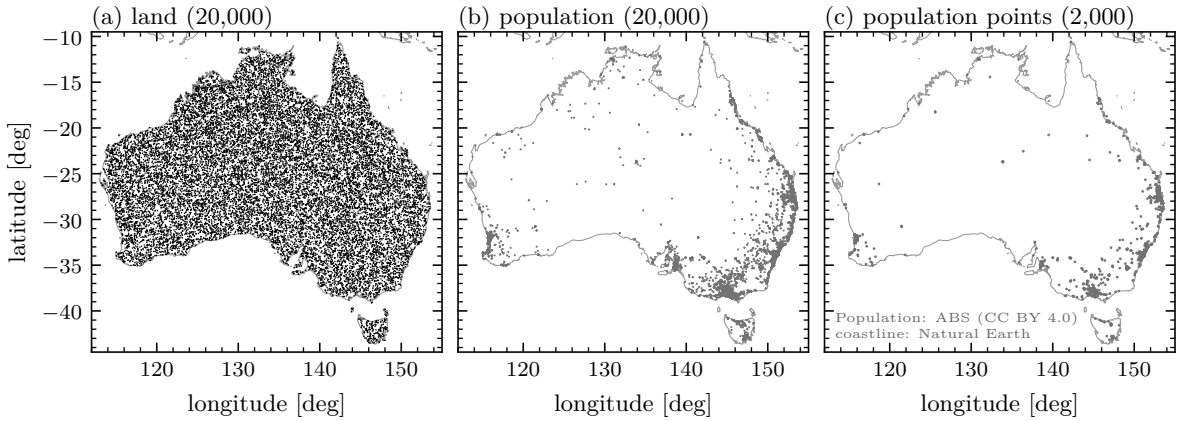


図 2　オーストラリアのランダムカタログと人口の点。(a) 陸地に一様に撒いた 20,000 点。(b) 人口に比例して撒いた 20,000 点。(c) 人口に比例して撒いた 2,000 点で、人口そのものの $\hat{\xi}$ を測るときに店の代わりのデータとして使う。どれも式 (12) の重みで格子を選び、格子の中に一様に置いた。種は二点相関関数の計算と同じで、同じ点である。

には、出店している地方もランダムカタログに入れる必要がある。

宇宙論にも同じ扱いがある。銀河のカタログにはある明るさより明るく見える銀河しか載らない。遠い銀河ほど暗く見えるので、カタログの銀河の数は遠くほど減る。これは銀河の並び方ではなく観測の性質である。銀河がカタログに載る確率を場所の関数として表したものを選択関数（selection function）と呼び、ランダムカタログも同じ割合で減らしてから比べる [8, 9]。本稿では、チェーンがその地方に人口あたりどれだけ店を出しているかを、この選択関数に見立てる。

選択関数は店の密度と人口の密度を幅 R でならした比として作った。

$$S_R(\boldsymbol{x}) = \frac{N_\mathrm{s}^{-1}\sum_i K_R(\boldsymbol{x},\boldsymbol{x}_i)}{N_\mathrm{p}^{-1}\sum_j K_R(\boldsymbol{x},\boldsymbol{y}_j)}, \qquad K_R(\boldsymbol{x},\boldsymbol{y}) = \exp\!\left(-\frac{c(\boldsymbol{x},\boldsymbol{y})^2}{2\sigma_R^2}\right). \tag{13}$$

分子の和は N_s 個の店、分母の和は人口に比例して撒いた $N_\mathrm{p} = 50{,}000$ 個の参照の点についてとる。分子は点 $\boldsymbol{x}$ のまわり（幅 R 程度）の店の数を、分母は同じ範囲の人の数を、それぞれ全体に対する割合で表す。したがって $S_R(\boldsymbol{x})$ は、点 $\boldsymbol{x}$ のまわりで人口あたりの店の数が全国の平均の何倍かを表す。$S_R = 1$ なら全国の平均と同じ割合で店を出していて、$S_R = 2$ ならその倍、S_R が 0 に近ければほとんど店を出していない。参照の点（人口に比例して撒いた点）で平均すると、トライアルの S_R は $R = 50,\ 100,\ 200\,\mathrm{km}$ で 0.97, 0.99, 0.93 で、ほぼ 1 である。図 3(b)–(d) はトライアルの S_R である。九州（北緯 34° より南、東経 132.2° より西）の参照の点で平均すると 2.6–3.5、関東（東経 138.9–140.9°、北緯 35.0–36.6°）では 0.30–0.43 である。ランダムカタログは人口に S_R を掛けた重みで撒く。このランダムカタログは「地方ごとの人口あたりの店の濃さは実際のとおりで、その中では人に比例して、ほかの店とは無関係に店を置いたら」という置き方を表す。

ならし方の関数 K_R は二点の距離で決まるガウス関数である。c は二点を結ぶ単位球の弦の長さで、大円距離を d、地球の半径を $R_\oplus$ とすると $c = 2\sin(d/2R_\oplus)$ である（図 3(a)）。σ_R は大円距

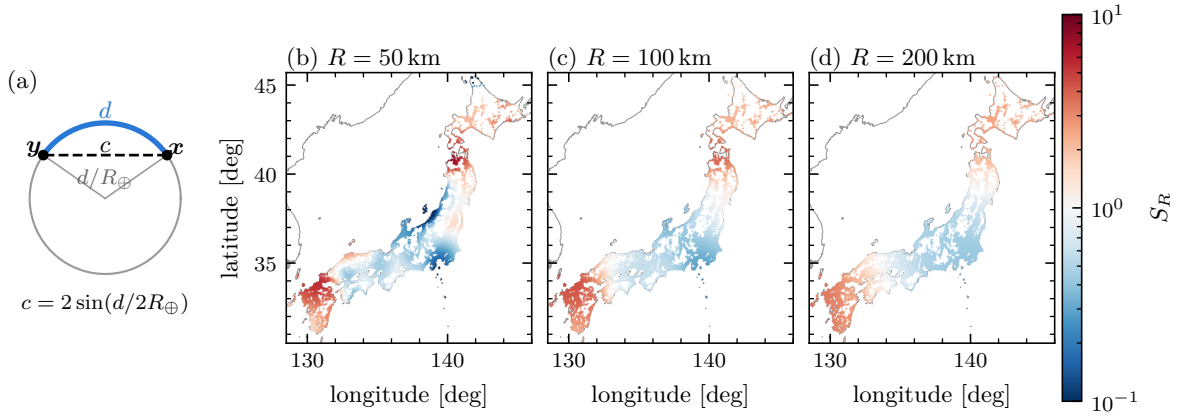


図 3　選択関数。(a) 地球の断面。表面の二点 $\boldsymbol{x}, \boldsymbol{y}$ の大円距離（太線）が d、中心から見た角度が $d/R_\oplus$ である。破線の弦の長さ c は、地球を半径 1 の球に縮めて測る。$R_\oplus$ は地球の半径である。説明のための絵。(b)–(d) トライアルの選択関数 S_R を、人口に比例して撒いた 50,000 個の参照の点ごとに色で示した。$R = 50, 100, 200\,\mathrm{km}$。色は対数で、0.1 未満と 10 より大きい値はそれぞれの端の色にした。参照の点は二点相関関数の計算と同じ点である。

離 R に当たる弦の長さ $2\sin(R/2R_\oplus)$ である。弦の長さは二点の単位ベクトルの差の長さなので、三角関数を使わずに求められる。本稿の距離（数百 km まで）では $c \approx d/R_\oplus$ なので、K_R は距離 d について標準偏差 R のガウス関数と同じである。ガウス関数は $3\sigma_R$ で打ち切った。ガウス関数を選んだのは、幅が一つの数 R で決まり、なめらかで、どの向きにも同じだからである。一定の半径の円の中を数えるような、ほかの窓関数でも同じ役割を果たす。窓関数の形による違いは確かめていないが、幅 R による違いは $R = 50, 100, 200\,\mathrm{km}$ で比べた（4.2 節）。

ランダムカタログは棄却法で撒いた [12]。まず人口に比例して候補の点を撒き、それぞれの候補を確率 $S_R/S_{\max}$ で残す。残った点の密度は人口と S_R の積に比例する。$S_{\max}$ は最初に撒いた 40,000 個の候補での最大値で、これを超えた候補は多くても 5 個（候補の 0.01% 未満）だった。

この選択関数は幅 R より長い距離の構造を基準に取り込む。そのため、$R/2$ より短い距離だけを読む。また、選択関数を店そのものから作るので、店が固まっている分の一部も基準に取り込まれ、$\hat{\xi}_{\mathrm{LS}}$ は小さめに出る。

2.5　誤差の見積もり

店の配置は一つの国に一通りしかないので、同じ国を何度も観測して $\hat{\xi}$ の散らばりを見ることはできない。そこで、手元のデータを部分に分け、部分を抜いたときに値がどれだけ動くかから誤差を見積もる。本稿ではその一つの jackknife を使った [13, 14]。銀河の二点相関関数でもよく使われ、ほかの方法との比較は Norberg ら [15] にある。

国を K 個の領域に分け、領域 k を除いたデータとランダムカタログで計算した $\hat{\xi}$ のビン i の

値を $\hat{\xi}_{(k),i}$ と書く。jackknife の共分散は

$$C_{ij} = \frac{K-1}{K} \sum_{k=1}^{K} \bigl(\hat{\xi}_{(k),i} - \bar{\xi}_i\bigr)\bigl(\hat{\xi}_{(k),j} - \bar{\xi}_j\bigr), \qquad \bar{\xi}_i = \frac{1}{K}\sum_{k=1}^{K} \hat{\xi}_{(k),i} \tag{14}$$

で、ビン i の標準偏差は $\sqrt{C_{ii}}$ である。ふつうの標本分散と違って、前に $1/(K-1)$ ではなく $(K-1)/K$ がつく。K 個の $\hat{\xi}_{(k)}$ はどれもデータの $(K-1)/K$ を共有しているので、互いにほとんど同じ値になり、散らばりは小さい。$(K-1)/K$ はその分を戻す係数で、$\hat{\xi}$ が標本の平均なら、式 (14) は標本の平均の分散 s^2/K（s^2 は標本分散）にちょうど一致する。

本稿では $K = 64$ とした。人口が大都市に偏っているので（オーストラリアならシドニーとメルボルン）、領域は面積ではなく人口でそろえた。人口に比例するランダムの点 20,000 個を、点の数が等しくなるように二等分することを 6 回繰り返した（$2^6 = 64$）。分ける向きは、そのときの点の広がりが東西と南北で大きい方にした。こうしてできた領域を付録 B の図 14 に示す。大都市の領域は小さく、人口の少ない地方の領域は大きい。領域 k を除いた組の数は、領域の組み合わせごとに数えておいた組の数から、領域 k を含むものを引いて求めた（数え直した値と一致することをテストで確かめた）。jackknife の標準偏差が実際のばらつきとどれだけ合うかは、toy model で確かめた（付録 A）。

$\hat{\xi}_{\rm LS}$ が全体として 0 と合うかは、3 km 以上の $p = 14$ 個のビンをまとめて調べた。ビン i の値を並べたベクトルを $\boldsymbol{x} = (\hat{\xi}_1, \ldots, \hat{\xi}_p)^{\mathsf{T}}$ とし、

$$\chi^2 = \boldsymbol{x}^{\mathsf{T}} C^{-1} \boldsymbol{x} \tag{15}$$

を求める。ビンの間に相関がなく C が対角なら、$\chi^2 = \sum_i (\hat{\xi}_i/\sigma_i)^2$ で、各ビンが 0 から誤差の何倍離れているかの二乗の和である。隣り合うビンは同じ店の組を一部共有するので、実際には相関がある。C^{-1} を使うと、相関のあるビンで同じずれを二重に数えないように重みが変わる。$\boldsymbol{x}$ が平均 0、共分散 C の多変量正規分布に従うなら、χ^2 は自由度 p の χ^2 分布

$$f_p(t) = \frac{t^{p/2-1}\, e^{-t/2}}{2^{p/2}\,\Gamma(p/2)} \qquad (t \ge 0) \tag{16}$$

に従う。これは独立な標準正規分布の値 p 個の二乗の和が従う分布で、平均は p である。この分布を考えるのは、得られた χ^2 が $\hat{\xi} = 0$ のもとでどれほど起こりにくいかを、上側確率 $\int_{\chi^2}^{\infty} f_p(t)\,dt$ として数えるためである。上側確率が小さいほど、$\hat{\xi} = 0$ とは合いにくい。

式 (15) の C は真の共分散ではなく、式 (14) で見積もった値である。見積もった C が平均として正しくても、その逆行列 C^{-1} は平均として真の逆行列より大きくなり、χ^2 も大きく出る。Hartlap ら [16] は、K 個の標本が互いに独立で正規分布に従うなら、C^{-1} の期待値が真の逆行列の $(K-1)/(K-p-2)$ 倍になることを示した。そこで C^{-1} に Hartlap の因子 $(K-p-2)/(K-1)$ をかけて偏りを除いた。$K = 64$、$p = 14$ なら因子は $48/63 = 0.762$ で、χ^2 は 2 割あまり小さくなる。ビンの数 p が領域の数 K に近づくほど偏りは大きく、$p \ge K - 2$ では直せない（表 14 の 16 領域）。この形は、標本が独立で正規分布に従うという仮定のもとで C^{-1} の期待値を合わせる

定数として一つに決まる。ただし、jackknife の $\hat{\xi}_{(k)}$ は互いに独立ではないので、本稿ではこの仮定は近似でしかない。また、因子は C^{-1} の平均を合わせるだけで、χ^2 が式 (16) に従うことまでは保証しない。因子をかける代わりに尤度の形を変える方法もある [17]。本稿では、こうした近似のずれを、後で述べる模擬データとの比較で校正する。

いくつかのビンを一つの値にまとめるときは、ビンの単純な平均をとった。誤差は jackknife の標本ごとに同じ平均をとって求めた標準偏差である。標本ごとに平均をとるので、ビンの間の相関はこの誤差に含まれる。共分散で重みをつけた平均、つまり一般化最小二乗法（generalized least squares; GLS）で定数を当てはめた値 $m = \mathbf{1}^{\mathsf{T}} C^{-1} \boldsymbol{x} / \mathbf{1}^{\mathsf{T}} C^{-1} \mathbf{1}$ は使わなかった。店が数百しかないと共分散が粗く、GLS の重みがその粗さに左右される。ビンの間の相関が強いと、GLS はどのビンよりも外の値を返すことがある [18]。トライアルの $R = 50\,\mathrm{km}$（4.2 節）では、ビンの値 0.056 と 0.103 から GLS は 0.118 を返した。単純な平均なら、日本のどの量でも、乱数の種を変えて jackknife の領域の分け方を変えても、まとめた値は誤差の 0.15 倍以下しか動かない。[*2]

店の数が数百しかないときは、jackknife の共分散が粗くなり、χ^2 の上側確率は信用できなくなる（下の表 2 と、それに続く基準）。そこで、有意かどうかは帰無仮説の模擬データと比べて判断する。実際の店から作った選択関数 S_R に従って、店と同じ数の点を独立に撒き、店の代わりに使う。この模擬の店には、選択関数より短い距離の相関がない。これに実際と同じ手順（模擬の店から選択関数を作り直す）を当て、乱数の種を変えて繰り返す。人口だけを基準にした χ^2 には、店と同じ数の点を人口だけに比例して撒いた模擬の店を使う。模擬データの店の数は実際と同じに固定するので、比べるばらつきは店の数を固定したときのばらつきである。店の数もポアソン分布で揺らしたときとは DD の分散の式が違い、その導出は付録 D にまとめた。

本稿で使う χ^2 をこうした模擬データ 100 個と比べて校正した（表 2）。模擬データと比べた上側確率は、模擬データのうち χ^2 が実際の値以上になったものの割合である。100 個のどれも届かなければ 0.01 未満としか言えない。校正のずれは一方向ではない。人口が基準の模擬データでは、14 ビンの χ^2 の中央値が 7–9 で、自由度 14 の χ^2 分布の中央値 13.3 より小さい。jackknife が誤差を大きめに見積もるためで、χ^2 分布から求めた上側確率は大きく出すぎる。一方、トライアルの $R = 50\,\mathrm{km}$ の模擬データでは、22 個が χ^2 分布で上側確率 0.05 未満になり、上側確率は小さく出すぎる。そこで本稿では、模擬データの値の分布と実際の店の値を比べる。

上側確率が「信用できる」かどうかは、次の基準で判断した。上側確率 P が正しく求まっていれば、帰無仮説が正しいとき P は 0 と 1 の間に一様に分布する。つまり、どの α についても $\Pr(P < \alpha) = \alpha$ である。帰無仮説の模擬データ M 個のうち $P < \alpha$ になる個数は、二項分布 $\mathrm{Bin}(M, \alpha)$ に従い、平均は $M\alpha$、標準偏差は $\sqrt{M\alpha(1-\alpha)}$ である。$M = 100$、$\alpha = 0.05$ なら 5 ± 2.2 個になる。実際の個数がこの二項分布でまず起こらない数なら、その上側確率は信用できないとみなす。トライアルの $R = 50\,\mathrm{km}$ の 22 個は、$\mathrm{Bin}(100, 0.05)$ で 22 個以上になる確率が 4×10^{-9} で、この基準で信用できない。逆に、オーストラリアの人口が基準の模擬データでは、Coles と Woolworths の 200 個のうち $P < 0.05$ は 0 個で、見込みの 10 ± 3.1 個より少なすぎ

[*2] 設計の記録は `docs/adr/002-combine-bins-with-plain-mean.md`。

表 2　χ^2 の上側確率を、帰無仮説の模擬データ 100 個と比べて校正した値。χ^2 は Hartlap の因子をかけた値で、「χ^2 分布」はそれを自由度がビンの数の χ^2 分布（式 (16)）と比べた上側確率。模擬データは、人口が基準の行では店と同じ数の点を人口だけに比例して撒き、R の行では幅 R の選択関数に従って撒いた。日本の $R = 50$、100 km の模擬データでは選択関数を点ごとに求め、ほかは 10 km 格子の表で求めた（付録 C）。3–25 km の 4 ビンには 8 km より短い 2 ビンも入る。ビンごとの比較は表 8。

国	標本	基準	範囲 [km]	ビン	χ^2	上側確率	
						χ^2 分布	模擬データ
オーストラリア	Coles	人口	≥ 3	14	10.4	0.733	0.28
	Woolworths	人口	≥ 3	14	10.6	0.719	0.27
	Coles	$R = 50$ km	3–25	4	0.8	0.939	0.95
	Woolworths	$R = 50$ km	3–25	4	2.1	0.713	0.81
	Coles	$R = 100$ km	3–50	5	3.4	0.636	0.71
	Woolworths	$R = 100$ km	3–50	5	2.4	0.787	0.79
	Coles	$R = 200$ km	3–100	7	4.5	0.722	0.66
	Woolworths	$R = 200$ km	3–100	7	3.4	0.844	0.85
日本	TRIAL	人口	≥ 3	14	72.9	< 0.001	< 0.01
	SEIYU	人口	≥ 3	14	40.0	< 0.001	< 0.01
	TRIAL×SEIYU	人口	3–2000	14	46.8	< 0.001	< 0.01
	TRIAL	$R = 50$ km	3–25	4	18.3	0.001	0.05
	SEIYU	$R = 50$ km	3–25	4	10.2	0.037	0.01
	TRIAL×SEIYU	$R = 50$ km	3–20	4	5.0	0.283	0.21
	TRIAL	$R = 100$ km	3–50	5	38.3	< 0.001	< 0.01
	SEIYU	$R = 100$ km	3–50	5	7.3	0.196	0.09
	TRIAL×SEIYU	$R = 100$ km	3–32	5	5.3	0.378	0.12
	TRIAL	$R = 200$ km	3–100	7	45.1	< 0.001	< 0.01
	SEIYU	$R = 200$ km	3–100	7	7.7	0.361	0.10
	TRIAL×SEIYU	$R = 200$ km	3–80	7	6.1	0.526	0.25

る。この上側確率も正しくないが、大きめに出る向きなので、相関のないものを有意と読む誤りにはつながらない。

上の模擬データでは、点を撒く重みは人口 $\times S_R$ である。S_R は幅 R（50 km 以上）でならした値なので、九州に多い、関東に少ないといった地方ごとの店の多さは模擬データに入る。一方、同じ地方の中では、点は人口に比例して撒かれる。たとえば $R = 50$ km の円の中に大きな市とまわりの町があれば、模擬の店は人口の割合どおりに市と町に分かれる。実際の店がこの割合より市の中心に集まっていれば、店どうしの距離は模擬データより短くなり、店が互いに固まっていなくても $\hat{\xi} > 0$ になる。逆に人の多い所を避けていれば $\hat{\xi}$ は小さく出る。店が人口の割合より人口密度の高い所に寄るか低い所に寄るかを、本稿では「密集地への寄り方」と呼ぶ。この寄り方は地方の濃淡よりずっと短い距離の話なので、選択関数では取り除けない。

そこで、撒く重みを人口 $\times S_R(\boldsymbol{x})\,\rho(\boldsymbol{x})^\gamma$ にした模擬データも作った。$\rho(\boldsymbol{x})$ は点 $\boldsymbol{x}$ を含む 1 km

格子の人口（格子はほぼ 1 km 四方なので、人口密度とみなせる）である。$\gamma = 0$ なら上の模擬データに戻り、$\gamma > 0$ なら人口の割合より密度の高い格子に、$\gamma < 0$ なら低い格子に多く撒く。γ はチェーンごとに、店のある格子の $\ln \rho$ の平均が、模擬データと実際の店とで等しくなるように決めた。

$$\frac{\sum_j S_R(\boldsymbol{y}_j)\, \rho(\boldsymbol{y}_j)^\gamma \ln \rho(\boldsymbol{y}_j)}{\sum_j S_R(\boldsymbol{y}_j)\, \rho(\boldsymbol{y}_j)^\gamma} = \frac{1}{N_\mathrm{s}} \sum_i \ln \rho(\boldsymbol{x}_i). \tag{17}$$

左辺の和は人口に比例して撒いた参照の点 $\boldsymbol{y}_j$ について、右辺の和は実際の店 $\boldsymbol{x}_i$（N_s 店）についてとる。左辺は、重みを ρ^γ 倍した模擬の店が立つ格子の $\ln \rho$ の平均である。

この決め方には二つの理由がある。一つは、ρ が日本の 1 km 格子で 1 人から 3 万人あまりまで 4 桁を超えて広がるので、ρ そのものの平均は少数の都心の格子で決まってしまうことである。$\ln \rho$ の平均（人口密度の幾何平均の対数）なら、店のある格子の典型的な密度を表す。もう一つは、式 (17) が、店を重み人口 $\times\, S_R \rho^\gamma$ に従って独立に置いたとみなしたときの γ の最尤推定の式と同じになることである。その対数尤度は、規格化を $Z(\gamma) = \sum_j S_R(\boldsymbol{y}_j) \rho(\boldsymbol{y}_j)^\gamma$ として

$$\ell(\gamma) = \sum_i \big[\gamma \ln \rho(\boldsymbol{x}_i) + \ln S_R(\boldsymbol{x}_i)\big] - N_\mathrm{s} \ln Z(\gamma) + \mathrm{const} \tag{18}$$

で、$d\ell/d\gamma = 0$ が式 (17) になる。$d^2\ell/d\gamma^2 = -N_\mathrm{s} \times$（模擬の店の $\ln \rho$ の分散）< 0 なので、左辺は γ について単調に増え、解は一つに決まる。これを二分法で解いた。

以下では、二つの模擬データを次のように呼ぶ。

- 地方だけをそろえた模擬データ：重みは人口 $\times\, S_R$。チェーンがどの地方に多いかは実際の店と同じだが、地方の中では人口の割合どおりに店を置く。
- 密集地への寄り方もそろえた模擬データ：重みは人口 $\times\, S_R\, \rho^\gamma$。地方ごとの多さに加えて、店のある格子の $\ln \rho$ の平均も実際の店と同じにする。

どちらも店どうしは互いに独立に置くので、店が互いに固まることは含まない。実際の店の $\hat{\xi}$ がこれらの模擬データより大きければ、地方の多さや密集地への寄り方では説明できない固まりがあることになる。

二社の相互相関では二社を互いに独立に撒くので、どちらも「二社は互いに無関係に店を置く」という帰無仮説になる。ただし、これらの模擬データの店は互いに固まっていない。そこで相互相関では一方のチェーンを実際の位置のまま残し、もう一方だけを撒き直した模擬データとも比べた。この模擬データには残したチェーンの店の固まりと置き方がそのまま入るので、二社が無関係でも生じる相互相関を、店が固まっていない模擬データより実際に近く表す。

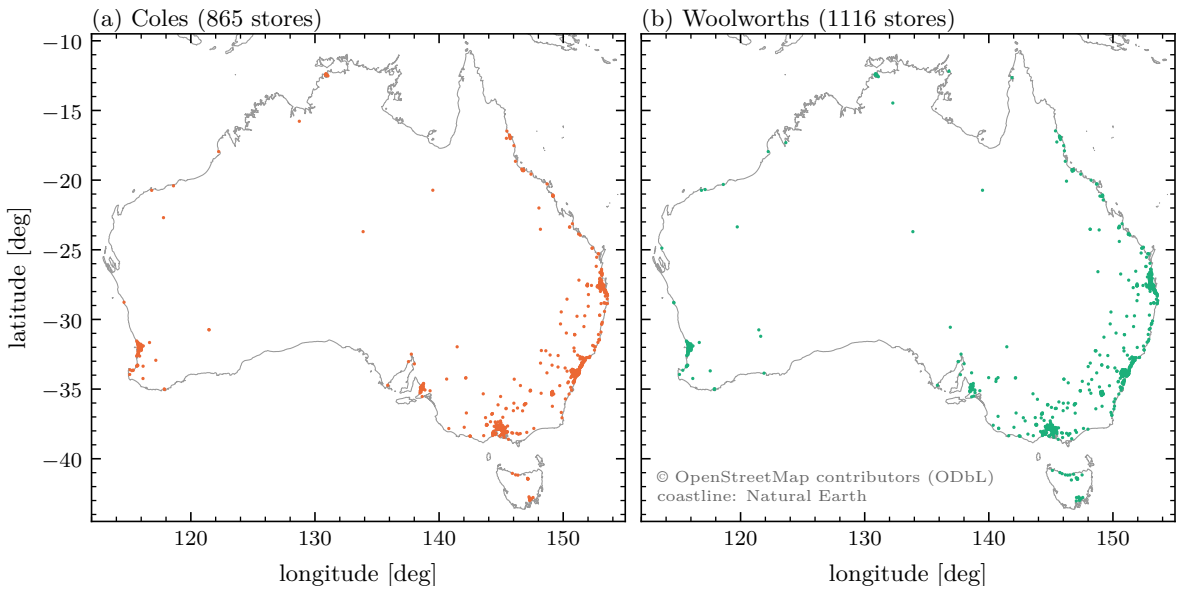


図 4　OSM から取り出した Coles と Woolworths の店舗（2026 年 11 月 1 日時点）。店舗は東海岸と南東部の都市に集まり、内陸にはほとんどない。灰色の線は Natural Earth の海岸線 [20]。

3　データ

3.1　オーストラリア

3.1.1　OpenStreetMap から取り出す

店舗の座標は OpenStreetMap（OSM）[19] から取り出した。OSM のデータは ODbL で公開されていて、出典を示せば研究に使える。取り出しには Overpass API を使い、2026 年 11 月 1 日 0 時（UTC）時点のデータを指定した。日付を固定したので、同じ問い合わせはいつ送っても同じ結果を返す。問い合わせは `scripts/osm/stores.overpassql` にある。店舗を州・特別地域ごとに取り出すので、各店舗がどの州にあるかも同時にわかる。

店舗として数えたのは `shop=supermarket` のタグがついたもののうち、`brand` が次のどれかのものである。Coles では Coles・Coles Local・Coles Central、Woolworths では Woolworths・Woolworths Metro を数えた。各社の決算資料では、Coles はスーパー全体、Woolworths はスーパーと Metro の合計を数えているので、それに合わせた（3.1.2 節）。

OSM では、同じ店が点（node）と建物の輪郭（way）の両方で描かれていることがある。そこで、同じチェーンの二点が 100 m より近ければ、同じ店とみなして一方を落とした。点と輪郭の組では輪郭を残し、同じ種類どうしの組では先に出た方を残した。こうして落とした店は 7 つで、そのうち 5 組は二点の距離が 25 m 以内だった。残った店舗を図 4 に示す。

表 3　地域ごとの店舗数。OSM は 2026 年 11 月 1 日、公式は 2026 年 6 月 28 日時点。

チェーン	地域	OSM	公式	OSM/公式
Coles	NSW & ACT	265	268	0.99
Coles	QLD	187	189	0.99
Coles	VIC & TAS	249	248	1.00
Coles	SA & NT	60	61	0.98
Coles	WA	104	104	1.00
Woolworths	NSW & ACT	359	360	1.00
Woolworths	QLD	259	265	0.98
Woolworths	VIC	276	279	0.99
Woolworths	SA & NT	80	80	1.00
Woolworths	WA	110	115	0.96
Woolworths	TAS	32	31	1.03
Coles	計	865	870	0.99
Woolworths	計	1116	1130	0.99

3.1.2　公式の店舗数との比較

OSM がどれだけ店舗を拾えているかを確かめるため、各社の決算資料にある 2026 年 6 月 28 日時点の店舗数と比べた（表 3）。Coles は Full Year Results Release の Appendix 1 [21]、Woolworths は F26 決算発表の Appendix Six [22] の値である。地域の分け方は各社の表に合わせた。全国の合計では、Coles は OSM が 865 店、公式が 870 店で、OSM は公式の 99.4% だった。Woolworths は OSM が 1116 店、公式が 1130 店で、98.8% だった。地域ごとに見ても、OSM は公式の 96–103% に収まる。

3.1.3　OpenStreetMap の限界

数がほぼ合っても、OSM の一つ一つの店が実在し、正しい位置にあるとは限らない。OSM を使うときの限界を次にまとめる。

1. 公式の数と比べられるのは地域ごとの合計までである。公式の資料には店舗ごとの一覧がない。そのため、閉店した店が OSM に残り、新しい店が OSM から抜けていても、両者が打ち消し合えば合計は合ってしまう。
2. OSM と公式の時点が 4 か月ずれている。6 月 28 日より後に開いた店が OSM に描かれていれば、OSM の数は公式より多くなる。その後に閉じた店が OSM から消されていれば、少なくなる。Coles は 2027 年度（2026 年 6 月 29 日から）に、約 25 店の開店と約 3 店の閉店を予定している [21]。同じ割合で開くなら、6 月 28 日から 11 月 1 日までの 4 か月で

約 8 店が開く。

3. 選び方によって数が変わる。`shop` のタグがない要素（Coles 5 件、Woolworths 5 件）は数えていない。`shop=convenience` の Woolworths MetroGo（6 件）も数えていない。Woolworths の公式の数に MetroGo が入っているかは、資料からはわからない。
4. 同じ店かどうかを二点の距離だけで決めている。60 m と 97 m の組は、移転の前後の店が両方とも OSM に残っているだけかもしれない。地図で一つずつ確かめる必要がある。
5. 店の座標は描き方によって数十 m ずれる。点で描かれた店では入口やロゴの位置、輪郭で描かれた店では建物の中心が座標になる。ショッピングセンターの中の店では、センターの建物全体の中心が座標になっていることもありうる。二点相関関数を 1 km より短い距離で測るときは、このずれで組が隣のビンに移り、結果が変わる。

3.1.4 人口

人口にはオーストラリア統計局（ABS）の 2025 年の格子人口 [23] を使った。1 辺 1 km の格子ごとの推計人口で、座標系は EPSG:3577（Albers 正積円錐図法）である。値が 0 以上の格子を陸地とみなすと、陸地の格子は 7,693,195 個、人口の合計は 27,613,318 人になる。格子の中の点は楕円体の Albers 図法の逆変換で緯度・経度に戻した [24]。楕円体は GRS80（長半径 $a = 6{,}378{,}137\,\mathrm{m}$、扁平率 $1/298.257222101$、離心率 e）、中央子午線は $\lambda_0 = 132°\mathrm{E}$、標準緯線は $\phi_1 = 18°\mathrm{S}$ と $\phi_2 = 36°\mathrm{S}$、原点の緯度は $\phi_0 = 0$ である。

$$
\begin{aligned}
q(\phi) &= (1-e^2)\left[\frac{\sin\phi}{1-e^2\sin^2\phi} - \frac{1}{2e}\ln\frac{1-e\sin\phi}{1+e\sin\phi}\right], \qquad m(\phi) = \frac{\cos\phi}{\sqrt{1-e^2\sin^2\phi}}, \\
n &= \frac{m(\phi_1)^2 - m(\phi_2)^2}{q(\phi_2) - q(\phi_1)}, \qquad C = m(\phi_1)^2 + n\,q(\phi_1), \qquad \rho_0 = \frac{a\sqrt{C - n\,q(\phi_0)}}{n}
\end{aligned}
\tag{19}
$$

を定数として、平面座標 (x, y) から

$$
\rho = \sqrt{x^2 + (\rho_0 - y)^2}, \qquad \theta = \arctan\frac{x}{\rho_0 - y}, \qquad \lambda = \lambda_0 + \frac{\theta}{n}, \qquad q(\phi) = \frac{C - \rho^2 n^2/a^2}{n} \tag{20}
$$

で経度 λ を求め、最後の式を満たす緯度 ϕ を反復で解く（$n < 0$ なので、ρ と θ の符号は n に合わせる）。実装は `crates/retail-cosmology/src/util/albers.rs` にあり、PROJ で変換した 4 都市の座標との差は 1 mm 未満である（テストで確かめた）。OSM の店舗はすべて陸地の格子に入り、人口が 0 の格子に入ったのは Coles の 2 店と Woolworths の 1 店だった。

3.2 日本

3.2.1 トライアル

日本の題材にはスーパーのトライアル（TRIAL）を選んだ。まず OSM から、オーストラリアと同じ方法で店を取り出した（2026 年 11 月 1 日時点、47 都道府県ごと）。ところが、見つかったのは 232 店で、決算説明資料にある 378 店（2026 年 6 月末）[25] の 61% だった。オーストラリ

表 4　トライアルの地方ごとの店舗数。サイトは 2026 年 11 月 3 日、公式は 2026 年 6 月末時点。OSM で一致は、サイトの店から 300 m 以内に OSM の店があった数。

地方	サイト	公式	サイト/公式	OSM で一致	OSM の割合
北海道	36	34	1.06	29	0.81
東北	42	40	1.05	33	0.79
関東	52	62	0.84	34	0.65
中部	21	20	1.05	10	0.48
近畿	43	40	1.07	25	0.58
中国・四国	36	35	1.03	23	0.64
九州	145	147	0.99	70	0.48
計	375	378	0.99	224	0.60

アの Coles と Woolworths が 99% だったのとは、事情がまったく異なる。

そこで、店の座標はトライアルの公式サイトの店舗検索 [26] から取った。サイトの地図は店ごとの業態・緯度・経度・都道府県を返すので、住所から座標を推定する必要はない。サイトに載っていた 401 店のうち、トライアルの名前の業態を数えた。西友の店（トライアル西友 8 店）、ドラッグストア（トライウェルと TRIDRUG GO の 9 店）、別のブランド（スーパーセンタータイヨー 3 店）、非表示の 2 店は除いた。残りは 375 店で、決算説明資料の 378 店の 99.2% になる。取ったデータの著作権はトライアルグループにあるので、リポジトリには載せていない。

地方ごとの数を表 4 に示す。決算説明資料の地方別の数は棒グラフから読んだ。資料に地方の定義が書かれていないので、8 地方区分（中国と四国はまとめる）を仮定した。関東だけサイトの数が公式より 1 割以上少ないのは、県の割り振りが資料と違うためかもしれない。

OSM の店とサイトの店を突き合わせると、OSM が拾えていたのは 60% で、九州と中部では 48% だった。一方、拾えていた店の OSM の座標は、サイトの座標から中央値 11 m、90% 点 52 m しかずれていなかった。日本の OSM の問題は描かれている店の位置ではなく描かれていない店が多いことにある。

3.2.2　西友

トライアルは 2025 年 7 月 1 日に西友を完全子会社にした [25]。そこで、二社の店の並び方を比べるため、西友の店も調べた。

店の座標は西友の公式サイトの店舗検索 [27] から取った（2026 年 11 月 4 日）。都道府県ごとのページに、店ごとの名前、住所、緯度、経度が載っている。ただし、トライアル西友（西友の店を業態転換した店）の一覧には東京都の 3 店しか載っていない。そこで、サイトの店のページの一覧（サイトマップ）と突き合わせ、一覧にない店のページも取った。これで浦安、二俣川、武蔵新城のトライアル西友 3 店が加わった。店のページの地図の座標は一覧の座標と同じ値である（西友青井店で確かめた）。店名に【閉店】とある 1 店（山科店）は除き、【改装中】の 2 店は数えた。決算説明資料はトライアル西友を西友の数に入れているので [25]、ここでも西友として数えた。取った

表 5　西友の地方ごとの店舗数。サイトは 2026 年 11 月 4 日、公式は 2026 年 6 月末時点。

地方	サイト	公式	サイト/公式
東北	19	19	1.00
関東	135	134	1.01
中部	71	71	1.00
近畿	18	19	0.95
計	243	243	1.00

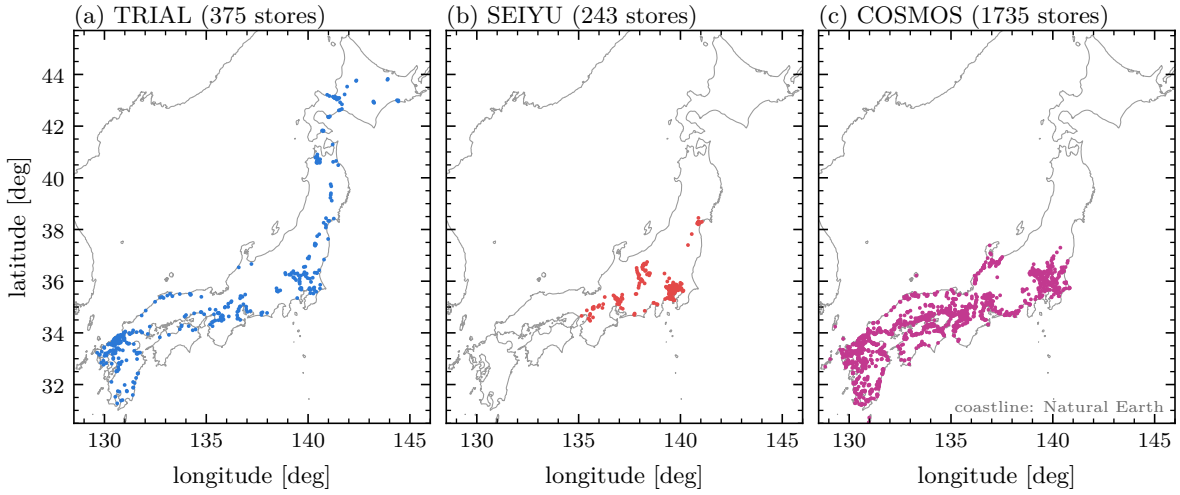


図 5　公式サイトから取った (a) トライアル（2026 年 11 月 3 日）、(b) 西友（2026 年 11 月 4 日）、(c) コスモス（2026 年 11 月 12 日）の店舗。コスモスの座標は住所から推定した（3.2.3 節）。灰色の線は Natural Earth の海岸線 [20]。店のない沖縄は範囲の外にある。

データの著作権は西友にあるので、リポジトリには載せていない。

数えた店は 243 店で、決算説明資料の 243 店（2026 年 6 月末）と一致した（表 5）。地方別の数は資料の棒グラフから業態ごとに読み取った。関東が 1 店多いのは、2026 年 8 月に南林間店が開いたためである [25]。近畿が 1 店少ないのは、閉店した山科店を除いたためである。

トライアルの公式サイトにも、業態がトライアル西友の店が 8 店載っている。8 店とも、西友のサイトのいずれかの店から 60 m 以内にあった。トライアルの店（3.2 節）からはトライアル西友を除いてあるので、同じ店を二社に数えてはいない。

二社の店を図 5 に示す。トライアルは九州の北部に最も多く、そこから山陽、近畿、関東、東北、北海道へ散らばっている。西友は関東と中部に集まり、九州と北海道には店がない。二社の店が同じ地方に重なるのは、主に関東と中部である。

表 6　コスモスの地区ごとの店舗数。サイトは 2026 年 11 月 12 日、公式は 2026 年 5 月末時点。東北はその後に初めて出店した福島県の 2 店。

地区	サイト	公式	サイト/公式
関東	211	202	1.04
中部	208	201	1.03
関西	249	244	1.02
中国	229	229	1.00
四国	160	159	1.01
九州	676	673	1.00
東北	2	—	—
計	1735	1708	1.02

3.2.3　コスモス

トライアルと地盤の重なるチェーンとして、ドラッグストアのコスモス（コスモス薬品）も調べた。コスモスは、2026 年 5 月末の 1,708 店のうち 673 店が九州にある [28]。OSM でトライアルと同じように店を取り出すと（2026 年 11 月 1 日時点）、449 店しか見つからず、決算短信の 26% だった。そこで、店は公式サイトの店舗検索 [29] から取った。サイトはスクリプトからの取得を受け付けないので、全店舗の一覧のページをブラウザで保存した（2026 年 11 月 12 日）。一覧には店ごとの名前、住所、地区が載っているが、緯度と経度はない。そこで、住所から国土地理院の住所検索 [30] で座標を引いた。「オープン予定」の 11 店は除いた。取ったデータの著作権はコスモス薬品にあるので、リポジトリには載せていない。店を図 5(c) に示す。

数えた店は 1,735 店で、地区ごとに決算短信の数と比べると 0–5% 多い（表 6）。多い分は 6 月以降の開店で、決算短信が「新商勢圏」として出店を広げるとしている関東と中部で多い [28]。地区の分け方は資料に書かれていないので、サイトの地区に従った（三重県は中部に入る）。

1,735 店のうち 1,300 店は、住所検索が街区か地番まで当たった。残りの 435 店は、町か字（大字）までしか当たらなかった。座標の精度は、OSM にある 449 店と突き合わせて確かめた。OSM の店からいちばん近いサイトの店までの距離は、街区か地番まで当たった店で中央値 49 m、90% 点 175 m だった。町か字までの店では中央値 478 m、90% 点 1.4 km だった。本稿で見る 3 km 以上の距離には十分な精度である。

3.2.4　人口

人口には令和 2 年国勢調査の 1 km メッシュ（3 次メッシュ）[31] を使った。メッシュは 176,964 個で、人口の合計は 126,146,099 人になり、国勢調査の総人口 [32] と一致した。ただし、このファイルに載っているのは人が住むメッシュだけで、人口が 0 の陸地は載っていない。そのため日本では、陸地に一様なランダムカタログは作れない。代わりに、人が住む格子に一様なランダムカタログを作り、オーストラリアでも同じ基準で計算し直した。人口に比例するランダムカタログは、

オーストラリアと同じ方法で作った。

3.3 米国

3.3.1 Walmart

Walmart は、2026 年 1 月期の売上高が 7,132 億ドルで、米国と海外の 18 か国に計 10,955 店（会員制の Sam's Club を含む）を持つ [33]。米国では Supercenter を中心に 4,611 店を持ち、トライアルと同じスーパーセンターの業態の店が多いので、比べる相手として店の並び方を測った。なお、トライアルのスーパーセンターも、当時米国で盛況だった業態に着想を得て 1996 年に始まった [2]。範囲は米国本土の 48 州とワシントン D.C. とし、アラスカとハワイは除いた。

店はオーストラリアと同じく OSM から取り出した。ただし、日付を固定した問い合わせは使えなかった。州ごとに問い合わせると 1 州に約 2 分かかって Overpass API の時間切れ（30 分）に届かず、米国本土を囲む矩形で一度に問い合わせても時間切れになった。そこで日付を固定せずに取り出し、データの時刻（2026 年 11 月 12 日 23 時 56 分（UTC））を記録した。問い合わせは `scripts/osm/walmart.overpassql` にある。矩形（北緯 24–50°、西経 60–130°）にはカナダとメキシコの一部も入るので、店がどの州にあるかを Natural Earth の州境 [34] で決め、どの州にも入らない 558 個の要素を落とした。

店舗として数えたのは、`shop=supermarket`（Supercenter と Neighborhood Market）か `shop=department_store`（Discount Store）のタグがあり、`brand:wikidata` が Walmart か Walmart Neighborhood Market の項目であるか、`brand` か `name` が Walmart で始まるものである。店の中の売り場（薬局、ガーデンセンター、オートケア、眼鏡）、ガソリンスタンドに併設の店（`shop=convenience`）、注文した品を受け取るだけの施設は数えなかった。会員制の Sam's Club は別のブランドなので入らない。オーストラリアと同じく 100 m より近い組を同じ店とみなし、8 店を落とした。残った店は 4,564 店である（図 6）。

OSM がどれだけ店を拾えているかを、10-K [33] にある 2026 年 1 月 31 日時点の店数と比べた（表 7）。10-K にあるのは全米（50 州、ワシントン D.C.、プエルトリコ）の業態ごとの合計で、州ごとの数はない。そこで、米国本土の外（アラスカ、ハワイ、プエルトリコ）の店も同じ規則で OSM から取り出して足した（`scripts/osm/walmart_outside.overpassql`）。業態は名前で分け、名前に Supercenter とあれば Supercenter、Neighborhood Market（か Market で終わる名前）なら Neighborhood Market、それ以外（名前が Walmart だけの店など）は Discount Store とみなした。全米の合計では、OSM が 4,603 店、10-K が 4,611 店で、OSM は 10-K の 99.8% だった。業態ごとにも 97–102% に収まる。ただし、名前の Walmart だけの店がすべて Discount Store とは限らず、業態ごとの数は名前の付け方に左右される。

3.3.2 人口

人口には JRC の GHS-POP [35] の 2020 年の推計を使った。1 辺 1 km の格子ごとの推計人口で、座標系は ESRI:54009（Mollweide 正積図法）である。世界全体の一枚から、米国本土の 49

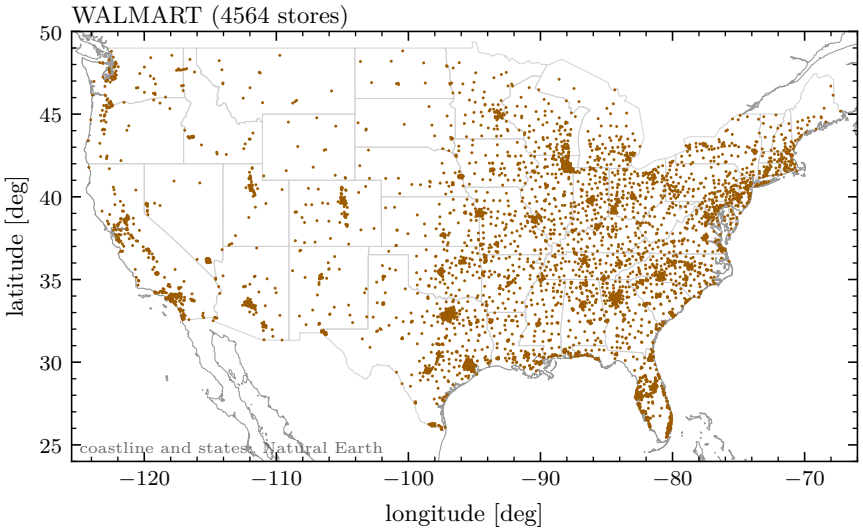


図 6　OSM から取り出した米国本土の Walmart の店舗（2026 年 11 月 12 日 23 時 56 分（UTC）時点）。店は東部と南部に多く、西部の山地と砂漠では都市の周りにしかない。灰色の線は Natural Earth の海岸線 [20]、薄い線は州境 [34]。

表 7　Walmart の業態ごとの店数。OSM は 2026 年 11 月 12–13 日、10-K は 2026 年 1 月 31 日時点。10-K の Neighborhood Market はほかの小型店を含む。

業態	本土	本土の外	OSM 計	10-K	OSM/10-K
Supercenter	3547	22	3569	3566	1.00
Discount Store	341	17	358	351	1.02
Neighborhood Market	676	0	676	694	0.97
計	4564	39	4603	4611	1.00

の州（D.C. を含む）の州境 [34] を囲む範囲を切り出し、格子の中心がどの州にも入らない格子を除いた。海の格子は値がないので、値が 0 以上の格子を陸地とみなした。陸地の格子は 7,953,624 個、人口の合計は 331,589,927 人になる。2020 年国勢調査の人口 [36] から米国本土の分を求めると 329,260,619 人（全米 331,449,281 人からアラスカとハワイを引いた値）で、GHS-POP の合計はこれより 0.7% 多い。

格子の中の点は、球の Mollweide 図法の逆変換で緯度・経度に戻した [24]。球の半径は $R = 6{,}378{,}137\,\mathrm{m}$ である。平面座標 (x, y) から

$$\theta = \arcsin\frac{y}{\sqrt{2}R}, \qquad \phi = \arcsin\frac{2\theta + \sin 2\theta}{\pi}, \qquad \lambda = \frac{\pi x}{2\sqrt{2}R\cos\theta} \tag{21}$$

で緯度 ϕ と経度 λ が求まる。実装は `crates/retail-cosmology/src/util/mollweide.rs` にある。順変換と逆変換を往復させた差と、正積であることをテストで確かめた。陸地の格子の外に出た店は 4 店、人口が 0 の格子に入った店は 29 店だった。

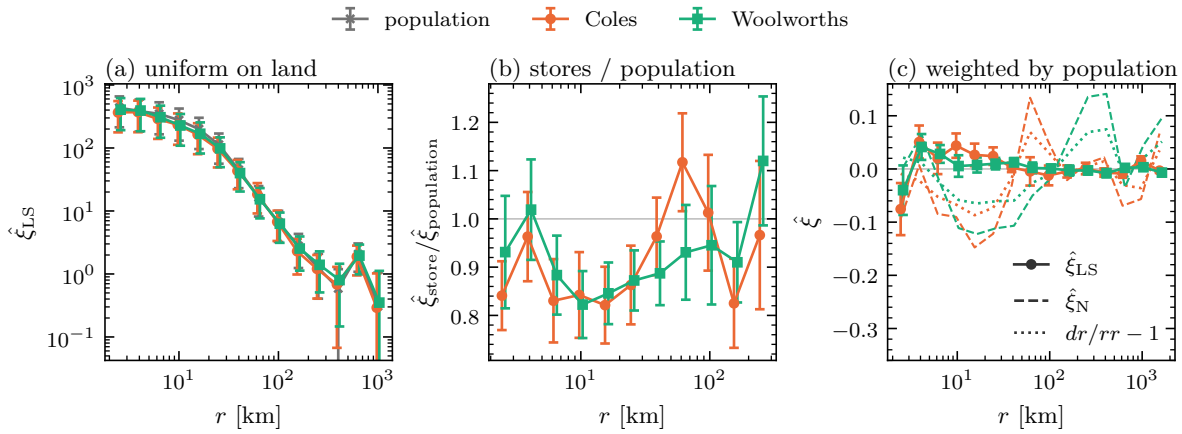


図 7　店舗の二点相関関数。誤差棒は jackknife（64 領域）の標準偏差。(a) 陸地に一様なランダムカタログを使った $\hat{\xi}_{\rm LS}$。灰色は人口そのもの。(b) (a) の店の値を人口の値で割った比。(c) 人口に比例するランダムカタログを使った $\hat{\xi}_{\rm LS}$（点と実線）、natural の推定量 $\hat{\xi}_{\rm N}$（破線）、$dr/rr-1$（点線）。破線と実線の差は点線の 2 倍（式 (22)）。

4　結果

結果を三つの問いの順に示す。4.1 節では、店の並び方が人口の分布で説明できるかを見る。4.2 節では、出店している地方を選択関数で取り除いても、店が互いに固まっているかを見る。4.3 節では、二社の店が近くに並ぶか、離れて並ぶかを相互相関で見る。どの節でも、人口でほぼ説明できるオーストラリアの二社を基準に置き、トライアル、西友、コスモス、Walmart をそれと比べる。ビンをまとめた値は単純な平均で（2.5 節）、人口が基準のときは中点が 10–40 km の 4 ビン（8.0–50.2 km）、選択関数を入れたときは 8 km から $R/2$ までのビンをまとめる。

4.1　人口で説明できるか

4.1.1　オーストラリア：人口でほぼ説明できる

陸地に一様な基準では（図 7(a)）、Coles と Woolworths の $\hat{\xi}_{\rm LS}$ は人口の $\hat{\xi}_{\rm LS}$ とほぼ重なる。2.5 km では約 400 で、距離が延びると下がり、100 km で約 6 になる。500–800 km のビンが盛り上がるのは、シドニーとメルボルン（713 km）のような大都市どうしの組が入るからである。(a) の誤差棒は 3–100 km のどのビンでも値の約 5 割あるが、店と人口の比をとると小さくなる（図 7(b)）。領域を一つ除いたとき、店と人口の $\hat{\xi}$ が同じ向きに揺れるからである。

比が 1 なら、その距離で店は人口と同じだけ固まっている。1 より小さければ、店は人口ほどには固まっていない。比は 10–40 km で 1 を下回り、4 ビンをまとめると Woolworths が 0.857 ± 0.052（1 より 2.8σ 小さい）、Coles が 0.873 ± 0.069（1.8σ）だった。

人口に比例する基準では（図 7(c) の実線）、$\hat{\xi}_{\rm LS}$ は 0 と合う。3 km 以上の 14 ビンで、Coles は

$\chi^2 = 10.4$、Woolworths は $\chi^2 = 10.6$ だった。人口に比例して撒いた模擬データと比べた上側確率は、Coles が 0.28、Woolworths が 0.27 である（表 2）。人口の分布から見込まれる以上に店どうしが固まる様子は見えない。

人口に比例する基準では、$\hat{\xi}_{\rm LS}$ とは別に $dr/rr - 1$ も見る（図 7(c) の点線）。$\hat{\xi}_{\rm LS}$ が「店どうしが固まっているか」を問うのに対し、$dr/rr - 1$ は「店が人口に比例して置かれているか」を問う量である。dr は店とランダムの点の組、rr はランダムの点どうしの組を数えたもので、ランダムの点は人口に比例して撒いてある。店が人口に比例して置かれていれば、店はランダムの点と同じ所に同じ割合で立つので、$dr/rr = 1$ になる。店が人の密集地に少なければ、ランダムの点が多い都市の中で店とランダムの点の組が減り、$dr/rr - 1 < 0$ になる。密集地に多ければ $dr/rr - 1 > 0$ である。以下では、この量をチェーンどうしの比較の主な物差しにする。

この量は推定量の違いにも現れる。式 (10) で定義した natural の推定量 $\hat{\xi}_{\rm N} = dd/rr - 1$ と $\hat{\xi}_{\rm LS}$ の差は

$$\hat{\xi}_{\rm N} - \hat{\xi}_{\rm LS} = 2\left(\frac{dr}{rr} - 1\right) \tag{22}$$

で、式 (10) から直接出る（ほかの推定量との関係は付録 E）。$\hat{\xi}_{\rm N}$ は、店が人口に比例していないことによるずれも含む。$\hat{\xi}_{\rm LS}$ は dr を使ってそのずれを差し引き、店どうしの固まりだけを残す。図 7(c) では、破線の $\hat{\xi}_{\rm N}$ が 8–40 km で -0.1 前後になるのに対し、実線の $\hat{\xi}_{\rm LS}$ は 0 の近くにとどまる。この差は点線の $dr/rr - 1$ の 2 倍で、natural の推定量の負の値は店どうしが離れているためではなく、店が人口に比例していないために生じたことがわかる。

4 ビンをまとめると、$dr/rr - 1$ は Woolworths が -0.061 ± 0.025 (2.4σ)、Coles が -0.061 ± 0.040 (1.5σ）だった。どちらも負で、人が密集している所ほど人口あたりの店が少ない。上の比が 1 を下回ることと同じ向きである。

まとめると、二社とも店の並び方は人口の分布でほぼ説明できる。ずれがあるとすれば、店が人口に比例するより少しだけ、人の少ない所に寄っていることである。大きさは店とランダムの点の組で 6% ほど、誤差の 1.5–3 倍なので、本稿ではこのずれを示唆にとどめる。都市の中では一店あたりの客が多い（店が大きい）といった説明が考えられるが、確かめていない。ランダムカタログの点の数、jackknife の領域の数、乱数の種を変えても、この結論は変わらない（F.1）。

4.1.2　日本と米国：人の密集地への寄り方がチェーンで違う

日本と米国のチェーンは、オーストラリアの二社よりずっと人口から外れている。$dr/rr - 1$ (8.0–50.2 km）は次のとおりだった。

- トライアル：-0.634 ± 0.075。3–40 km で -0.5 から -0.7 になり（図 8(a))、店とランダムの点の組は、人口に比例して置いた場合の 4 割に満たない。Coles と Woolworths の約 10 倍の偏りである。
- コスモス：-0.657 ± 0.075 で、トライアルとほぼ同じ。
- Walmart：-0.475 ± 0.104 で、Coles と Woolworths の約 8 倍。

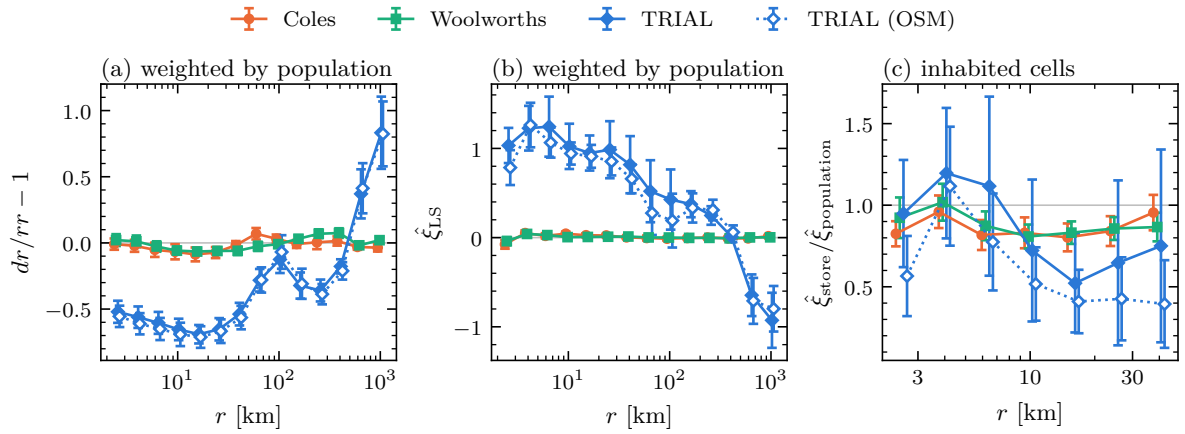


図 8　トライアル（日本）と Coles・Woolworths（オーストラリア）の比較。誤差棒は jackknife（64 領域）。(a) 人口に比例するランダムカタログに対する $dr/rr - 1$。(b) 人口に比例するランダムカタログに対する $\hat{\xi}_{\rm LS}$。(c) 人が住む格子に一様なランダムカタログでの、店と人口の $\hat{\xi}_{\rm LS}$ の比。白抜きの点はトライアルの店を OSM から取ったもの。

- 西友：$+0.620 \pm 0.283$ で、ほかのチェーンと符号が逆になり、人の密集地に寄っている。

人が住む格子を基準にした比も、8.0–50.2 km でトライアルが 0.660±0.443、Coles が 0.857±0.078、Woolworths が 0.841 ± 0.057 だった（図 8(c)）。

人口に比例する基準の $\hat{\xi}_{\rm LS}$ も、オーストラリアと違って 0 と合わない。トライアルは 3–25 km で約 +1 になり、距離が延びると下がって 400 km でほぼ 0、600–1000 km では負になる（図 8(b)）。3 km 以上の 14 ビンで $\chi^2 = 72.9$ で、店と同じ数の点を人口だけに比例して撒いた模擬データ 100 個はどれもこの値に届かなかった（最大 25.6、表 2）。西友は 2.0–3.2 km のビンで 3.9、12.6–20.0 km のビンで 1.5 で、50–126 km のビンではほぼ 0 になる（図 10(a) の灰色）。Walmart は 14 ビンで $\chi^2 = 106.8$ だった（模擬データでは校正していない）。

これらの値は、店どうしの間隔の取り方よりも、どの地方に出店しているかを主に表している。トライアルは 375 店のうち 145 店が九州に（表 4）、西友は 243 店のうち 135 店が関東にある（表 5）。短い距離の正の値は自社の強い地方に店が固まっていることを、トライアルの長い距離の負の値は九州と関東のような遠い地方どうしの組が人口から見込まれるより少ないことを表す。西友は関東を地盤とし、駅に近い店を多く持つので [3]、人の密集地に寄ることは事業の成り立ちからも自然である。トライアルの店を OSM から取った 232 店で計算しても、$dr/rr - 1$ は -0.659 ± 0.081、χ^2 は 69.1 で、結論は変わらない（図 8 の白抜きの点）。

4.1.3　地方の中での寄り方

地方の偏りを除いても密集地への寄り方が残るかを、式 (17) の γ で見る。γ は、選択関数で地方の多さをそろえたうえで、店のある 1 km 格子の人口密度 ρ の対数の平均を実際の店に合わせる指数で、正なら密集地に寄り、負なら避ける（2.5 節）。種 1 の模擬データで解くと、γ は

西友で 0.417、トライアルで -0.044 だった。西友の店のある格子の $\ln\rho$ の平均は 8.979（幾何平均で格子あたり約 7,900 人）で、人口と選択関数で撒いた基準の平均 8.419（約 4,500 人）より大きい。トライアルでは 7.731 と 7.825 で、ほとんど差がない。Walmart は $R = 50\,\mathrm{km}$ で $\gamma = -0.180$、$R = 100\,\mathrm{km}$ で $\gamma = -0.203$ だった。西友は出店している地方の中でも人の多い所に寄り、Walmart は避けているが、トライアルは地方の中では人の多い所を特に避けてはいない。

九州の中だけで測っても、同じことが確かめられる。人口の格子を福岡、佐賀、長崎、熊本、大分、宮崎、鹿児島の 7 県に入る 1 km 格子に限り、店もその格子に入るものだけを残した。格子がどの県に入るかは、総務省統計局の市区町村別メッシュ・コード一覧 [37] で決めた。残った店はトライアル 145 店、コスモス 674 店である（西友は九州に店がない）。$dr/rr - 1$ は、トライアルが $+0.196 \pm 0.085$ と正になり、コスモスは -0.123 ± 0.037 に縮む。人口が基準の $\hat{\xi}_{\mathrm{LS}}$ は、3 km 以上の 11 ビンで χ^2 がトライアル 9.9、コスモス 12.4 で、χ^2 分布で見ればどちらも 0 と合う（模擬データでは校正していない）。九州の中では、二社の店の並び方は人口の分布でほぼ説明できる。

まとめると、全国で見えたトライアルとコスモスの「密集地を避ける」傾向は、大都市の多い関東などに店が少ないことから来ている。地方の中まで密集地への寄り方が残るのは、寄る向きの西友と、避ける向きの Walmart である。

4.2 地方を取り除いても固まるか

選択関数を入れたランダムカタログ（2.4 節）で $\hat{\xi}_{\mathrm{LS}}$ を測り、帰無仮説の模擬データ（地方だけをそろえたものと、密集地への寄り方もそろえたもの、2.5 節）100 個と比べる。比べる距離は、8 km から $R/2$ までのビンをまとめた値と、8 km より短い距離（3.2–8.0 km の 2 ビンの単純な平均と、2.0–3.2 km のビン）である。2.0–3.2 km のビンは、ランダムカタログが 1 km 格子の中で一様なので格子の大きさの影響を受ける。ただし模擬データの店も同じように格子の中で一様に撒くので、模擬データとの比較は同じ条件で行える。トライアルと西友の 8 km より短い距離の比較（表 8）では、模擬データをビンごとの値を残すように作り直した（まとめた値と χ^2 は前と同じ値が出ることを確かめた）。

4.2.1 オーストラリア：固まりはない

出店している地方を選択関数として入れても、Coles と Woolworths の $\hat{\xi}_{\mathrm{LS}}$ は 0 と合う（図 9(b)）。$R = 100\,\mathrm{km}$ で 8.0–31.7 km の 3 ビンをまとめると、Coles は 0.008 ± 0.008、Woolworths は -0.007 ± 0.009 だった。人口だけでほぼ説明できていたので、予想どおりである。

4.2.2 トライアル：3–25 km で固まる

トライアルは出店している地方を取り除いても、短い距離に正の相関が残る（図 9(a)）。相関は 6 km のビンが最も大きく、25 km あたりまで正で、40 km より長い距離ではほぼ 0 になる。8 km から $R/2$ までのビンをまとめると、$R = 50\,\mathrm{km}$ で 0.080 ± 0.087、$R = 100\,\mathrm{km}$ で 0.196 ± 0.120、$R = 200\,\mathrm{km}$ で 0.377 ± 0.121 だった。R が小さいと店が固まっている分まで基準に取り込まれ、

表 8　8 km より短い距離の、選択関数つきの $\hat{\xi}_{\rm LS}$ と相互相関 $\hat{\xi}_{\times}$ を、帰無仮説の模擬データ 100 個と比べた値。「模擬データ」の列は平均 ± 標準偏差、「以上」「以下」は実際の値以上・以下になった模擬データの数。模擬データの組は、地方だけをそろえたもの、密集地への寄り方もそろえたもの、一方のチェーンを実際の位置のまま残してもう一方を撒き直したもの（2.5 節）。一方を残す組では、残したチェーンの $\hat{\xi}_{\rm LS}$ は実際と同じなので相互相関だけを示す。

R	模擬データ	標本	範囲 [km]	実際	模擬データ	以上	以下
50	地方だけ	トライアル	3.2–8.0	+0.150	−0.044 ± 0.047	0	100
50	地方だけ	トライアル	2.0–3.2	−0.063	−0.059 ± 0.114	52	48
50	地方だけ	西友	3.2–8.0	+0.648	−0.015 ± 0.060	0	100
50	地方だけ	西友	2.0–3.2	+1.100	−0.008 ± 0.160	0	100
50	地方だけ	相互相関	3.2–8.0	+0.041	−0.003 ± 0.084	31	69
50	地方だけ	相互相関	2.0–3.2	−0.145	−0.021 ± 0.196	70	30
50	密集地も	トライアル	3.2–8.0	+0.150	−0.039 ± 0.045	0	100
50	密集地も	トライアル	2.0–3.2	−0.063	−0.039 ± 0.116	59	41
50	密集地も	西友	3.2–8.0	+0.648	+0.081 ± 0.077	0	100
50	密集地も	西友	2.0–3.2	+1.100	+0.065 ± 0.183	0	100
50	密集地も	相互相関	3.2–8.0	+0.041	−0.021 ± 0.087	25	75
50	密集地も	相互相関	2.0–3.2	−0.145	−0.047 ± 0.216	66	34
50	トライアルを残す	相互相関	3.2–8.0	+0.041	−0.140 ± 0.097	3	97
50	トライアルを残す	相互相関	2.0–3.2	−0.145	−0.159 ± 0.182	49	51
50	西友を残す	相互相関	3.2–8.0	+0.041	−0.006 ± 0.118	28	72
50	西友を残す	相互相関	2.0–3.2	−0.145	−0.076 ± 0.219	61	39
100	地方だけ	トライアル	3.2–8.0	+0.290	−0.020 ± 0.057	0	100
100	地方だけ	トライアル	2.0–3.2	+0.007	−0.043 ± 0.130	34	66
100	地方だけ	西友	3.2–8.0	+0.870	−0.014 ± 0.068	0	100
100	地方だけ	西友	2.0–3.2	+1.447	−0.030 ± 0.153	0	100
100	地方だけ	相互相関	3.2–8.0	−0.068	−0.004 ± 0.069	81	19
100	地方だけ	相互相関	2.0–3.2	−0.237	−0.007 ± 0.177	92	8
100	密集地も	トライアル	3.2–8.0	+0.290	−0.022 ± 0.051	0	100
100	密集地も	トライアル	2.0–3.2	+0.007	−0.026 ± 0.120	39	61
100	密集地も	西友	3.2–8.0	+0.870	+0.141 ± 0.098	0	100
100	密集地も	西友	2.0–3.2	+1.447	+0.151 ± 0.231	0	100
100	密集地も	相互相関	3.2–8.0	−0.068	−0.025 ± 0.075	76	24
100	密集地も	相互相関	2.0–3.2	−0.237	−0.036 ± 0.190	86	14
100	トライアルを残す	相互相関	3.2–8.0	−0.068	−0.204 ± 0.078	5	95
100	トライアルを残す	相互相関	2.0–3.2	−0.237	−0.237 ± 0.147	48	52
100	西友を残す	相互相関	3.2–8.0	−0.068	−0.010 ± 0.119	68	32
100	西友を残す	相互相関	2.0–3.2	−0.237	−0.073 ± 0.211	76	24

R が大きいと地方の濃淡が取り除ききれずに残るので、$R = 50$ km の値が最も控えめな見積もりである。

$R = 50$ km の模擬データ 100 個では、8.0–20.0 km の 2 ビンをまとめた値は平均 −0.027、標準偏差 0.024、最大 0.049 だった（図 9(c)）。平均が負なのは、選択関数が相関の一部を取り込むためである。トライアルの 0.080 以上になった模擬データは一つもなく（$p < 0.01$）、模擬データの平均から 4.5σ 離れている（σ は模擬データの標準偏差。分布の形は F.2）。トライアル自身の jackknife の誤差 0.087 で数えると 0.080 は 0 から 0.9σ しか離れていないが、これは誤差の大きい 8.0–12.6 km のビン（0.056 ± 0.118）に単純な平均で半分の重みがかかるためで、jackknife は誤差を大きめに見積もってもいる。χ^2 の上側確率はトライアルについては信用できないので（2.5 節）、有意かどうかはまとめた値を模擬データと比べて判断する。

8 km より短い距離でも、トライアルは固まっている（表 8）。3.2–8.0 km の値は $R = 50$ km で +0.150、$R = 100$ km で +0.290 で、どちらの模擬データの 100 個も届かなかった（模擬データの最大は +0.107）。一方、2.0–3.2 km のビンは −0.063 と +0.007 で、模擬データの分布の真ん

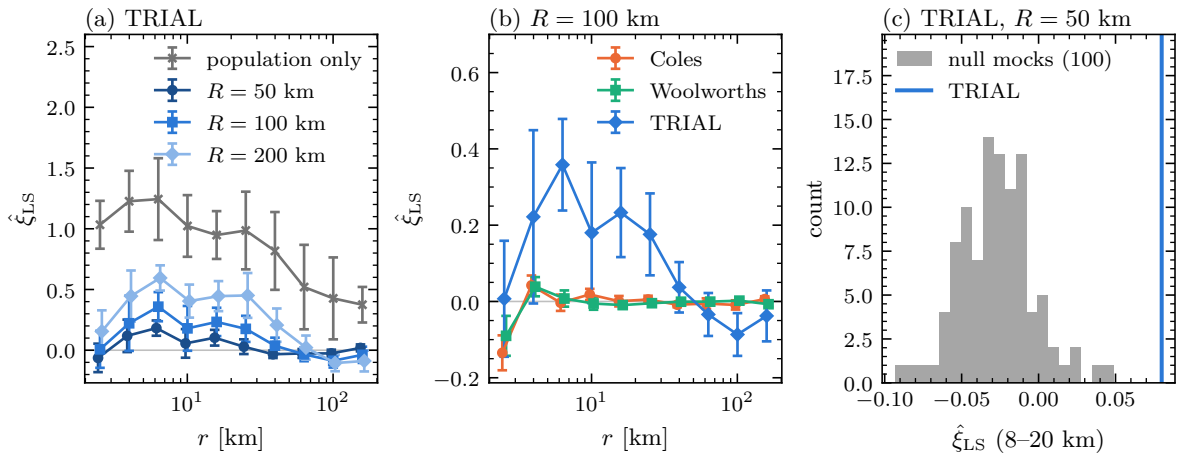


図 9　選択関数を入れたランダムカタログに対する $\hat{\xi}_{\rm LS}$。誤差棒は jackknife（64 領域）。(a) トライアル。灰色は人口だけを基準にしたもの、青は選択関数を入れたもので、濃いほど R が小さい。(b) $R = 100\,$km でのトライアル、Coles、Woolworths。(c) 帰無仮説の模擬データ 100 個での、8.0–20.0 km の 2 ビンをまとめた値の分布（$R = 50\,$km）と、トライアルの値。

中にある。トライアルの固まりは 3 km ほどより長い距離に現れ、それより近くには店を重ねていない。

同じ地域に店を重ねて出す出店の仕方（ドミナント出店）は、この結果と合う説明の一つだが、この解析では確かめていない。店を近くに集めると、配送や人の融通の費用が下がる。米国の Walmart について、店が密集することの利点を、近くの店どうしが客を奪い合う損と比べて見積もった研究がある [38]。トライアルには、説明の候補がもう一つある。小型店の TRIAL GO は、近くの大型店を拠点に、そこで作った惣菜などの商品を高頻度で届ける「サテライト型店舗」として運営されている [39, 40]。このような店は、拠点の大型店の近くにしか置けない。本稿のトライアルの店（375 店）のうち 33 店が TRIAL GO で、最寄りのスーパーセンターかメガセンターまでの距離の中央値は 4.7 km、33 店のうち 19 店が 5 km 以内にある。大型店どうしの最寄りの距離の中央値 7.6 km より短い。ただし TRIAL GO は 1 割に満たないので、固まりのどれだけをこの運営で説明できるかは、この解析では確かめていない。

4.2.3　西友：数キロの距離で強く固まる

選択関数を入れると、西友の $\hat{\xi}_{\rm LS}$ は短い距離でだけ正になる（図 10(a)）。$R = 50\,$km では 2.0–3.2 km のビンで 1.10 ± 0.47、5.0–8.0 km のビンで 0.56 ± 0.29 で、12.6–20.0 km のビンでは 0.10 ± 0.11 まで下がる。

8 km より短い距離では、西友は強く固まっている（表 8）。3.2–8.0 km で $+0.648$（$R = 50\,$km）と $+0.870$（$R = 100\,$km）、2.0–3.2 km で $+1.100$ と $+1.447$ だった。密集地への寄り方もそろえた模擬データでも平均は $+0.065$–$+0.151$ にとどまり、四つの値のどれも 100 個のうち一つも届かなかった（平均から 5.6–7.4σ）。西友の店は人の多い所に寄るだけでなく、互いに数キロの距離に

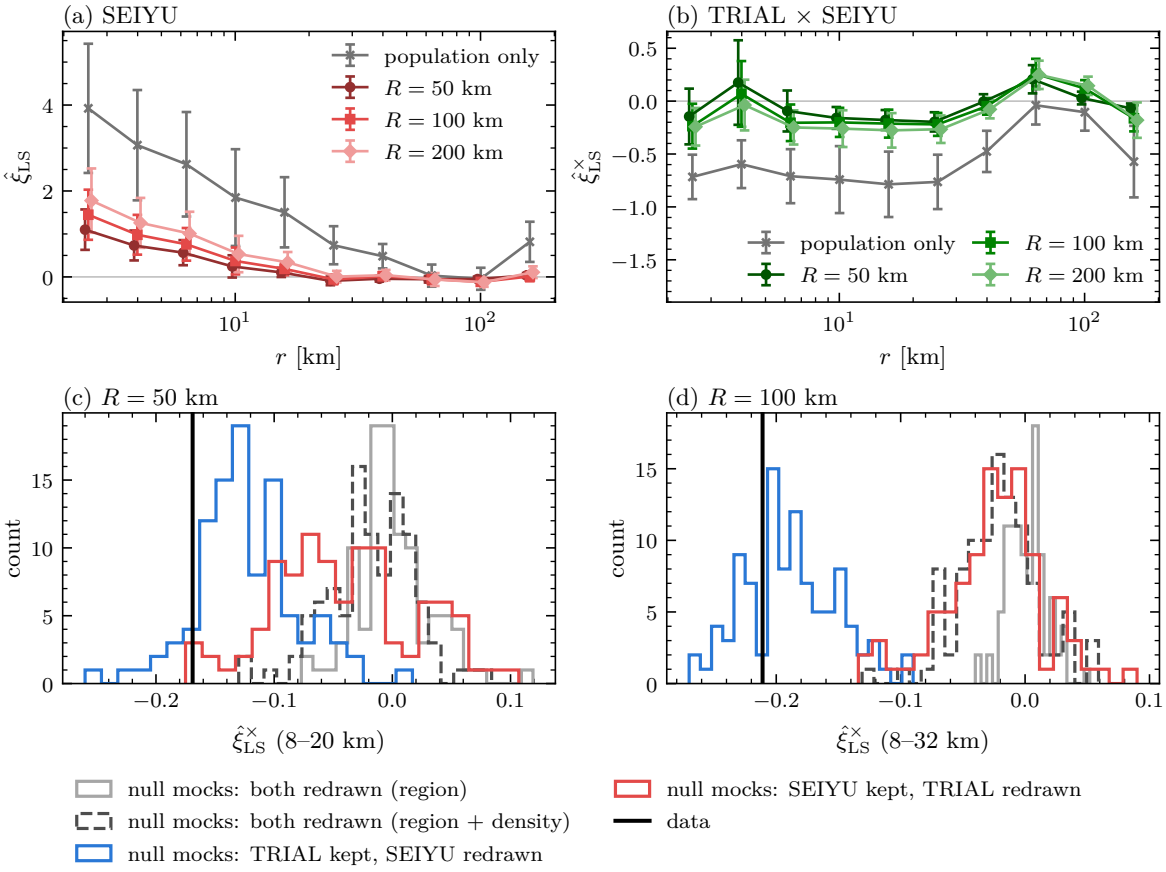


図 10　西友と、トライアルと西友の相互相関。誤差棒は jackknife（64 領域）。(a) 西友の $\hat{\xi}_{\rm LS}$。灰色は人口だけを基準にしたもの、色つきは選択関数を入れたもので、濃いほど R が小さい。(b) トライアルと西友の相互相関 $\hat{\xi}_{\times}$（式 (11)）。(c) $R = 50\,$km で 8.0–20.0 km の 2 ビンをまとめた $\hat{\xi}_{\times}$ の、帰無仮説の模擬データ 100 個での分布と、実際の値（縦線）。薄い灰色は二社とも地方だけをそろえて撒き直した模擬データ、濃い灰色の破線は密集地への寄り方もそろえたもの、青はトライアルを実際の位置のまま残して西友だけを撒き直したもの、赤は西友を残してトライアルだけを撒き直したもの。(d) (c) と同じものを $R = 100\,$km、8.0–31.7 km の 3 ビンで。

並んでいる。関東では駅に近い店が多いので [3]、駅どうしの間隔で店が並ぶことが一因と考えられるが、確かめていない。模擬データの店は 1 km 格子の中で一様に撒くので、格子より細かい人の住み方（駅の近くへの集まり）は模擬データに入らないことにも注意が要る。

8 km 以上では、判断がビンのまとめ方に左右される。8 km から $R/2$ までのビンをまとめると、$R = 50\,$km で 0.175 ± 0.174、$R = 100\,$km で 0.166 ± 0.130、$R = 200\,$km で 0.230 ± 0.148 だった。$R = 50\,$km の地方だけをそろえた模擬データは平均 -0.007、標準偏差 0.021、最大 0.053、密集地への寄り方もそろえた模擬データは平均 $+0.045$、標準偏差 0.037、最大 0.132 で、どちらも西友の値に届かなかった（後者の平均から 3.5σ）。$R = 100\,$km では、地方だけをそろえた模擬データの最大は 0.038 で、密集地への寄り方もそろえた模擬データ（平均 $+0.085$、標準偏差 0.038）

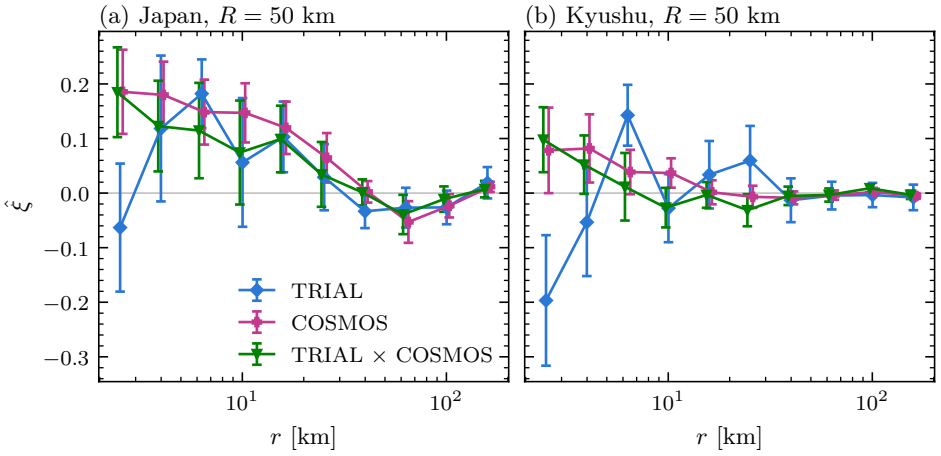


図 11　トライアルとコスモスの選択関数つきの $\hat{\xi}_{\rm LS}$ と、二社の相互相関 $\hat{\xi}_{\times}$（$R = 50\,\mathrm{km}$）。(a) 日本全体、(b) 九州 7 県の格子と店だけで計算した値。誤差は jackknife。

は 100 個のうち 2 個が西友の 0.166 以上になった（2.1σ）。密集地への寄り方で説明できるのは一部で、残りは模擬データにほとんど届かない。ところが共分散で重みをつけた平均（GLS）でまとめていたときは、密集地への寄り方もそろえた模擬データのうち 20 個（$R = 50\,\mathrm{km}$）と 69 個（$R = 100\,\mathrm{km}$）が西友の値以上になり、寄り方だけで説明できると読めた。差は 8.0–12.6 km のビン（$R = 50\,\mathrm{km}$ で 0.24 ± 0.26）から来ており、GLS はこのビンに負の重みをつけていた（2.5 節）。西友が密集地への寄り方を超えて 8 km 以上で固まっているかは、このビンの誤差が大きいので、まだはっきりとは言えない。

4.2.4　コスモス：固まりの大半は九州の外から来る

コスモスと、九州の中だけのトライアルについても同じ比較を行った（図 11、表 9）。九州の中だけの計算は 4.1 節と同じ格子と店を使い、模擬データの選択関数は 10 km 格子の表で求めた（付録 C）。

全国で選択関数を入れると、コスモスの店は強く固まっている。8–25 km（$R = 50\,\mathrm{km}$）で +0.133、8–50 km（$R = 100\,\mathrm{km}$）で +0.212 で、どちらの模擬データの 100 個も届かず、平均から 17σ 以上離れている。模擬データのばらつきが小さいのは、店が 1,735 店と多いからである。

九州の中だけで見ると、この固まりの大半は消える。$R = 50\,\mathrm{km}$ では、トライアルは +0.003 で模擬データの分布の真ん中にあり、コスモスは +0.019 で 100 個のうち 1–2 個が届いた（平均から 2.2–2.4σ）。コスモスは 3.2–8.0 km では +0.060 で、どちらの模擬データも届かなかった（4σ）。$R = 100\,\mathrm{km}$ では、トライアルは +0.098（100 個のうち 1 個が届いた）、コスモスは +0.034（4.5–6.0σ）だった。

全国で見えた二社の固まりの大半は、九州の外から来ている。九州の外では、二社とも店の少ない地方に新しく出店している（1、3.2.3 節）。新しく出る地方では、出店した市や町の近くに店を重ねていくので、選択関数の幅 R より短い距離で店が固まると考えられる。地盤の九州では店が

表 9　トライアルとコスモスの選択関数つきの $\hat{\xi}_{\mathrm{LS}}$ と相互相関 $\hat{\xi}_{\times}$ を、帰無仮説の模擬データ 100 個と比べた値。「全国」は日本全体、「九州」は九州 7 県の格子と店だけで計算した値。距離は 8 km から $R/2$ までのビンと、3.2–8.0 km の 2 ビンの単純な平均。列の意味は表 8 と同じ。

範囲	R	模擬データ	標本	距離 [km]	実際	模擬データ	以上	以下
全国	50	地方だけ	トライアル	8.0–25.0	+0.080	−0.027 ± 0.024	0	100
全国	50	地方だけ	トライアル	3.2–8.0	+0.150	−0.043 ± 0.047	0	100
全国	50	地方だけ	コスモス	8.0–25.0	+0.133	+0.000 ± 0.007	0	100
全国	50	地方だけ	コスモス	3.2–8.0	+0.164	−0.002 ± 0.013	0	100
全国	50	地方だけ	相互相関	8.0–25.0	+0.087	+0.001 ± 0.010	0	100
全国	50	地方だけ	相互相関	3.2–8.0	+0.119	+0.001 ± 0.017	0	100
全国	50	密集地も	トライアル	8.0–25.0	+0.080	−0.028 ± 0.028	0	100
全国	50	密集地も	トライアル	3.2–8.0	+0.150	−0.040 ± 0.047	0	100
全国	50	密集地も	コスモス	8.0–25.0	+0.133	+0.000 ± 0.008	0	100
全国	50	密集地も	コスモス	3.2–8.0	+0.164	−0.001 ± 0.013	0	100
全国	50	密集地も	相互相関	8.0–25.0	+0.087	+0.004 ± 0.010	0	100
全国	50	密集地も	相互相関	3.2–8.0	+0.119	+0.003 ± 0.020	0	100
全国	100	地方だけ	トライアル	8.0–50.0	+0.196	−0.015 ± 0.024	0	100
全国	100	地方だけ	トライアル	3.2–8.0	+0.290	−0.021 ± 0.057	0	100
全国	100	地方だけ	コスモス	8.0–50.0	+0.212	−0.001 ± 0.004	0	100
全国	100	地方だけ	コスモス	3.2–8.0	+0.266	−0.000 ± 0.014	0	100
全国	100	地方だけ	相互相関	8.0–50.0	+0.135	−0.001 ± 0.008	0	100
全国	100	地方だけ	相互相関	3.2–8.0	+0.175	−0.002 ± 0.022	0	100
全国	100	密集地も	トライアル	8.0–50.0	+0.196	−0.015 ± 0.020	0	100
全国	100	密集地も	トライアル	3.2–8.0	+0.290	−0.022 ± 0.051	0	100
全国	100	密集地も	コスモス	8.0–50.0	+0.212	+0.004 ± 0.006	0	100
全国	100	密集地も	コスモス	3.2–8.0	+0.266	+0.004 ± 0.013	0	100
全国	100	密集地も	相互相関	8.0–50.0	+0.135	+0.001 ± 0.007	0	100
全国	100	密集地も	相互相関	3.2–8.0	+0.175	+0.001 ± 0.019	0	100
九州	50	地方だけ	トライアル	8.0–25.0	+0.003	−0.003 ± 0.036	38	62
九州	50	地方だけ	トライアル	3.2–8.0	+0.045	−0.005 ± 0.061	24	76
九州	50	地方だけ	コスモス	8.0–25.0	+0.019	−0.004 ± 0.010	2	98
九州	50	地方だけ	コスモス	3.2–8.0	+0.060	−0.003 ± 0.016	0	100
九州	50	地方だけ	相互相関	8.0–25.0	−0.015	−0.003 ± 0.012	86	14
九州	50	地方だけ	相互相関	3.2–8.0	+0.032	−0.001 ± 0.021	6	94
九州	50	密集地も	トライアル	8.0–25.0	+0.003	−0.006 ± 0.033	42	58
九州	50	密集地も	トライアル	3.2–8.0	+0.045	−0.009 ± 0.060	16	84
九州	50	密集地も	コスモス	8.0–25.0	+0.019	−0.002 ± 0.009	1	99
九州	50	密集地も	コスモス	3.2–8.0	+0.060	−0.001 ± 0.015	0	100
九州	50	密集地も	相互相関	8.0–25.0	−0.015	−0.001 ± 0.013	86	14
九州	50	密集地も	相互相関	3.2–8.0	+0.032	−0.002 ± 0.023	7	93
九州	100	地方だけ	トライアル	8.0–50.0	+0.098	+0.003 ± 0.033	1	99
九州	100	地方だけ	トライアル	3.2–8.0	+0.094	+0.017 ± 0.073	15	85
九州	100	地方だけ	コスモス	8.0–50.0	+0.034	−0.001 ± 0.006	0	100
九州	100	地方だけ	コスモス	3.2–8.0	+0.110	+0.001 ± 0.014	0	100
九州	100	地方だけ	相互相関	8.0–50.0	−0.051	−0.002 ± 0.010	100	0
九州	100	地方だけ	相互相関	3.2–8.0	+0.003	−0.004 ± 0.022	37	63
九州	100	密集地も	トライアル	8.0–50.0	+0.098	+0.004 ± 0.033	1	99
九州	100	密集地も	トライアル	3.2–8.0	+0.094	+0.019 ± 0.075	12	88
九州	100	密集地も	コスモス	8.0–50.0	+0.034	+0.002 ± 0.007	0	100
九州	100	密集地も	コスモス	3.2–8.0	+0.110	+0.003 ± 0.016	0	100
九州	100	密集地も	相互相関	8.0–50.0	−0.051	−0.002 ± 0.011	100	0
九州	100	密集地も	相互相関	3.2–8.0	+0.003	−0.006 ± 0.024	33	67

すでに人口に沿って行き渡っており、固まりはほとんど残らない。ただし、店を開いた年は調べていないので、この説明は確かめていない。

4.2.5　Walmart：数十 km で弱く固まり、数 km では離れる

Walmart の模擬データは $R = 50$、100 km で 100 個ずつ作り、選択関数は 10 km 格子の表で求めた（付録 C、図 12、表 10）。選択関数を入れると、Walmart の店は 8 km より長い距離では固ま

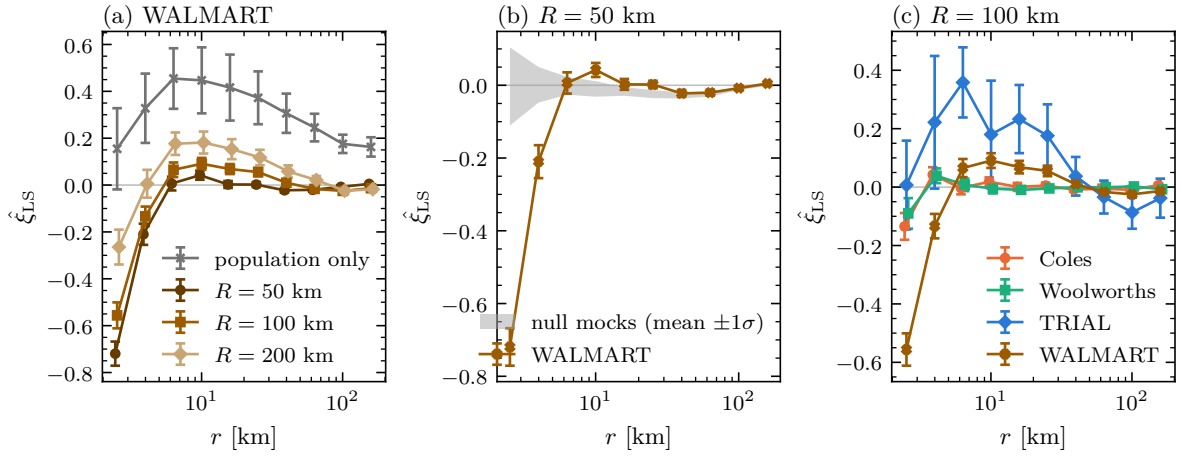


図 12　Walmart の選択関数つきの $\hat{\xi}_{\rm LS}$。(a) 人口だけを基準にしたもの（灰色）と、選択関数の幅 $R = 50$、100、200 km の基準。(b) $R = 50$ km の値と、密集地への寄り方もそろえた帰無仮説の模擬データ 100 個のビンごとの平均と標準偏差（灰色の帯）。(c) $R = 100$ km での Coles、Woolworths、トライアル、Walmart。

表 10　Walmart の選択関数つきの $\hat{\xi}_{\rm LS}$ を、帰無仮説の模擬データ 100 個と比べた値。距離は 8 km から $R/2$ までのビン、3.2–8.0 km の 2 ビンの単純な平均と、2.0–3.2 km のビン。列の意味は表 8 と同じ。

R	模擬データ	標本	距離 [km]	実際	模擬データ	以上	以下
50	地方だけ	Walmart	8.0–25.0	+0.022	-0.031 ± 0.009	0	100
50	地方だけ	Walmart	3.2–8.0	−0.102	-0.040 ± 0.022	100	0
50	地方だけ	Walmart	2.0–3.2	−0.719	-0.046 ± 0.101	100	0
50	密集地も	Walmart	8.0–25.0	+0.022	-0.014 ± 0.013	1	99
50	密集地も	Walmart	3.2–8.0	−0.102	-0.000 ± 0.029	100	0
50	密集地も	Walmart	2.0–3.2	−0.719	-0.003 ± 0.107	100	0
100	地方だけ	Walmart	8.0–50.0	+0.072	-0.015 ± 0.005	0	100
100	地方だけ	Walmart	3.2–8.0	−0.034	-0.016 ± 0.020	77	23
100	地方だけ	Walmart	2.0–3.2	−0.556	-0.017 ± 0.051	100	0
100	密集地も	Walmart	8.0–50.0	+0.072	$+0.023 \pm 0.010$	0	100
100	密集地も	Walmart	3.2–8.0	−0.034	$+0.036 \pm 0.018$	100	0
100	密集地も	Walmart	2.0–3.2	−0.556	$+0.041 \pm 0.042$	100	0

り、短い距離では互いに離れている。8–25 km（$R = 50$ km）で +0.022、8–50 km（$R = 100$ km）で +0.072 で、地方だけをそろえた模擬データの 100 個はどちらも届かなかった。密集地への寄り方もそろえた模擬データでも、$R = 50$ km で 100 個のうち 1 個しか届かず（2.8σ）、$R = 100$ km では届かなかった（5.0σ）。ただし値は、トライアル（$R = 50$ km で +0.080）やコスモス（+0.133）より小さい。

8 km より短い距離では、逆に店どうしが離れている。$R = 50$ km のビンごとの値は、2.0–3.2 km

で -0.719 ± 0.052、3.2–5.0 km で -0.210 ± 0.045、5.0–8.0 km で $+0.006\pm0.029$ だった。2.0–3.2 km と 3.2–8.0 km の値は、どちらの模擬データの 100 個も下回った（$R = 50$ km で -2.8 から -6.7σ）。近くの店どうしが客を奪い合うので、Walmart は約 5 km より近くには店を置かないようにしていると考えられる。トライアルは同じ 3.2–8.0 km で $+0.150$ と固まっていたので、短い距離での店の置き方は二社で逆である。

Walmart の店が数 km の距離では離れ、数十 km の距離では固まるという結果は、Walmart の出店を、店の密集による費用の節約（物流拠点からの配送など）と、近くの店どうしの客の奪い合いの釣り合いとして説明した研究 [38] と合う。ただし、この解析では物流拠点の位置も開店の年も使っておらず、この説明は確かめていない。

4.3 二社の関係

4.3.1 トライアルと西友：負だが、トライアルの置き方で説明できる

人口だけを基準にすると、二社の相互相関は大きく負になる（図 10(b) の灰色）。2–32 km のビンはどれも -0.6 から -0.8 で、8.0–50.2 km の 4 ビンをまとめると -0.692 ± 0.249 だった。トライアルは九州に、西友は関東と中部に多いので、二社の店の組は人口から見込まれるより少ない。

選択関数を入れても、相互相関は負のまま残る。$R = 50$ km では 8.0–12.6 と 12.6–20.0 km のビンがそれぞれ -0.159 ± 0.103 と -0.179 ± 0.094 だった。$R = 100$ km では 20.0–31.7 km のビンも -0.220 ± 0.108 で、31.7–50.2 km のビンは -0.049 ± 0.079 とほぼ 0 になる。8 km から $R/2$ までのビンをまとめると、$R = 50$ km で -0.169 ± 0.090、$R = 100$ km で -0.211 ± 0.115、$R = 200$ km で -0.220 ± 0.122 だった。-0.2 前後の値は、8–30 km の距離にある二社の店の組が、出店している地方から見込まれるより 2 割ほど少ないことを表す。

二社とも撒き直した模擬データでは、この負の値は説明できない（図 10(c)(d)）。$R = 50$ km の地方だけをそろえた模擬データは平均 $+0.001$、標準偏差 0.031、最小 -0.077、密集地への寄り方もそろえた模擬データは平均 -0.018、標準偏差 0.037、最小 -0.130 で、どちらの 100 個も実際の -0.169 に届かなかった。$R = 100$ km でも、最小はそれぞれ -0.040 と -0.131 で、実際の -0.211 に届いたものはない。西友が人の多い所に寄り、トライアルが寄らないことは、相互相関を -0.02 ほどしか下げない。

ただし、これらの模擬データの店は互いに固まっていないが、実際のトライアルの店は固まっている（4.2 節）。実際の値の jackknife の標準偏差は 0.090 で、密集地への寄り方もそろえた模擬データのばらつき 0.037 の 2.4 倍ある。jackknife の誤差で見ると -0.169 は 0 から 1.9σ で、3–20 km の 4 ビンの $\chi^2 = 5.0$ 以上になった地方だけをそろえた模擬データは 100 個のうち 21 個だった（表 2）。

そこで、一方のチェーンを実際の位置のまま残した模擬データとも比べた（2.5 節）。撒き直す側は選択関数と密集地への寄り方に従って撒き、選択関数は 10 km 格子の表で求めた（付録 C。結果の差はランダムカタログの揺らぎより小さい）。トライアルを実際の位置のまま残し、西友を撒き直すと、相互相関は大きく負になる（図 10(c)(d) の青）。$R = 50$ km の 100 個では平均 -0.124、標

準偏差 0.043、最小 −0.260 で、100 個のうち 10 個が実際の −0.169 以下になった。$R = 100\,\mathrm{km}$ でも平均 −0.189、最小 −0.270 で、28 個が実際の −0.211 以下になった。逆に、西友を実際の位置のまま残し、トライアルを撒き直すと、負の値はほとんど再現されない（赤）。$R = 50\,\mathrm{km}$ では平均 −0.040、最小 −0.175 で、実際の値以下になったのは 100 個のうち 2 個だった。$R = 100\,\mathrm{km}$ では平均 −0.025、最小 −0.134 で、実際の値以下になったものはない。

したがって、二社の相互相関が負になることは、トライアルの実際の店の置き方と、西友が地方と密集地への寄り方に従って店を置くことで説明できる。二社が互いを避けて店を置いていると考える必要はない。トライアルを撒き直すと再現しないので、負の値を生んでいるのは、トライアルの店の置き方のうち、選択関数と γ では表せない性質である。たとえば、西友の多い地方の中でトライアルが人の多い所をとくに避けている可能性があるが、この解析では確かめていない。

8 km より短い距離では、二社の相互相関はどの模擬データとも合う（表 8）。3.2–8.0 km で +0.041（$R = 50\,\mathrm{km}$）と −0.068（$R = 100\,\mathrm{km}$）、2.0–3.2 km で −0.145 と −0.237 で、二社とも撒き直した模擬データとの差は 1.3σ 以内だった。8 km 以上で見えた負の値は、これより短い距離には続いていない。トライアルを残して西友を撒き直した模擬データでは、3.2–8.0 km の平均が −0.140 と −0.204 で、実際の値以上になったのは 100 個のうち 3 個と 5 個だった。この距離では、トライアルの店の置き方から見込まれるより二社の店が近くに並んでいるが、差は 2σ 足らずで、本稿ではこれ以上の結論は出さない。

4.3.2　トライアルとコスモス：全国では正、九州の中では負

全国では、トライアルとコスモスの相互相関は +0.087（$R = 50\,\mathrm{km}$）と +0.135（$R = 100\,\mathrm{km}$）で、模擬データから 8σ 以上離れて正である（表 9）。西友との相互相関が負だったのとは逆に、二社の店は出店している地方から見込まれるより近くに並んでいる。

九州の中だけで見ると、相互相関は $R = 50\,\mathrm{km}$ では模擬データと合い、$R = 100\,\mathrm{km}$ では −0.051 と負になって、100 個のどれもこの値以下にならなかった（-4.4–-4.8σ）。九州の中では、二社の店はむしろ互いに少し離れて並んでいる。全国での正の値は、二社の固まりと同じく九州の外から来ている（4.2 節）。二社の新しい店が同じ市や町に向かえば、相互相関も正になる。

5　結論と議論

本稿ではオーストラリア、日本、米国のスーパーとドラッグストアの店舗の座標に、二点相関関数の Landy–Szalay 推定量を当てた。結果を三つの問いに分けてまとめる。

1. **人口で説明できるか。** Coles と Woolworths の店の並び方は人口の分布でほぼ説明でき、人口に比例する基準で $\hat{\xi}_{\mathrm{LS}}$ は 0 と合う。トライアル、コスモス、Walmart は人の密集地に店が少なく、$dr/rr - 1$ は 8.0–50.2 km で −0.634、−0.657、−0.475 と、オーストラリアの二社の 8–10 倍ある。西友は逆に +0.620 で密集地に寄る。この差の大半は出店している地方の偏りで、九州の中だけで測るとトライアルとコスモスの並び方は人口でほぼ説明でき

る。地方の中まで寄り方が残るのは、密集地に寄る西友（$\gamma = 0.417$）と、避ける Walmart（$\gamma = -0.18$ から -0.20）である。

2. **地方を取り除いても固まるか。** Coles と Woolworths は固まらない。トライアルは 3–25 km で固まり、8–20 km の値は模擬データ 100 個のどれも届かなかった（平均から 4.5σ）。それより近い 2.0–3.2 km には店を重ねていない。西友は 8 km より短い距離で強く固まる（寄り方もそろえた模擬データから 5.6–7.4σ）。8 km 以上の判断はビンのまとめ方に左右される。コスモスは全国では強く固まるが（17σ 以上）、九州の中だけでは固まりの大半が消え、トライアルも同じである。Walmart は 8–50 km で弱く固まり（$R = 100$ km で 5.0σ）、8 km より短い距離ではトライアルとは逆に店どうしが離れる（2.0–3.2 km で -0.719、-6.7σ）。
3. **二社の関係。** トライアルと西友の相互相関は、地方を取り除いても 8–30 km で -0.2 前後と負になる。ただし、トライアルを実際の位置のまま残して西友だけを撒き直すと、100 個のうち 10 個（$R = 100$ km では 28 個）が実際の値以下になり、二社が互いを避けていると考える必要はない。トライアルとコスモスの相互相関は全国では正（$+0.087$、8σ 以上）だが、九州の中だけでは $R = 100$ km で -0.051 と負になる。

この結果には次の限界がある。

1. 比べているのは国ではなくチェーンである。日本の小売全体の並び方を知るには、ほかのチェーンも調べる必要がある。
2. データの質がチェーンと国で違う。コスモスの座標は住所から推定し、435 店は町か字までしか当たらない（3.2.3 節）。米国の店は日付を固定して取り出せず、人口の GHS-POP は国勢調査の区域ごとの人口を建物の分布で格子に割り振った推計である（3.3 節）。人口は 2020 年（オーストラリアは 2025 年）、店は 2026 年で、時期がそろっていない。OSM から取った店の限界は 3.1.3 節にまとめた。
3. ランダムカタログも模擬データも 1 km 格子の中では一様なので、格子より細かい人の住み方を表せない。2.5 km のビンは χ^2 から除いたが、西友の 8 km より短い距離の固まりには、この分が含まれうる。
4. jackknife の誤差は一つの国の中での揺らぎしか表さない。国どうしを比べるときは、この誤差の意味を確かめる必要がある。
5. ビンをまとめた値は、まとめ方に左右される。単純な平均ではなく GLS でまとめると、西友が密集地への寄り方を超えて 8 km 以上で固まっているという判断は出てこない。
6. 人口あたりの店が少ないこと、店が固まること、二社の相互相関が負になることの原因（店の大きさ、土地の値段、ほかのチェーンとの競争、出店の方針など）は、この解析からはわからない。トライアルの固まりにはドミナント出店と TRIAL GO の運営が、コスモスの固まりには九州の外への新しい出店が合うが、確かめるには店の業態や開店の年を分けて $\hat{\xi}$ を測る必要がある。

宇宙論の方法を小売に当てはめるときは、次の三点に注意が要る。一つ目は、何を基準に撒くか

で $\hat{\xi}$ に出るものが変わることである。陸地に一様な基準では人の住み方が、人口に比例する基準では出店している地方の偏りが、$\hat{\xi}$ に出る。二つ目は、選択関数を店そのものから作ると、店が固まっている分の一部も基準に取り込まれ、$\hat{\xi}$ が小さめに出ることである。三つ目は、店の数が銀河よりずっと少ないことである。数百の点では jackknife の共分散が粗く、χ^2 分布から求めた上側確率は、場合によって大きくも小さくも出すぎた（表 2）。有意かどうかは帰無仮説の模擬データと比べて判断する必要がある。なお、店の数を固定しても人口の濃淡があると DD の分散は大きくなるが、Landy–Szalay 推定量ではその分がほぼ打ち消される（付録 D）。

謝辞

店舗の座標には OpenStreetMap の寄稿者が描いたデータを使った。人口にはオーストラリア統計局の格子人口と、e-Stat で公開されている国勢調査の 1 km メッシュを使った。

参考文献

[1] Australian Competition and Consumer Commission. Supermarkets inquiry: Final report. https://www.accc.gov.au/system/files/supermarkets-inquiry_1.pdf, 2025. February 2025, released 2025-03-21. Executive Summary, p. 1.

[2] 株式会社トライアルホールディングス. 会社沿革. https://trial-holdings.inc/company/history/, 2026. 2026-11-10 閲覧.

[3] 高野真維. 西友を買収するトライアルホールディングスとは？ 完全子会社化の狙いは？ 相乗効果は？. https://netshop.impress.co.jp/node/13627, 2025. ネットショップ担当者フォーラム、2025-03-11.

[4] B. D. Ripley. The second-order analysis of stationary point processes. *Journal of Applied Probability*, 13(2):255–266, 1976. doi: 10.2307/3212829.

[5] A. J. Baddeley, J. Møller, and R. Waagepetersen. Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54(3):329–350, 2000. doi: 10.1111/1467-9574.00144.

[6] Gilles Duranton and Henry G. Overman. Testing for localization using micro-geographic data. *The Review of Economic Studies*, 72(4):1077–1106, 2005. doi: 10.1111/0034-6527.00362.

[7] E. Marcon and F. Puech. Measures of the geographic concentration of industries: Improving distance-based methods. *Journal of Economic Geography*, 10(5):745–762, 2010. doi: 10.1093/jeg/lbp056.

[8] P. J. E. Peebles. *The Large-Scale Structure of the Universe*. Princeton University Press, Princeton, 1980.

[9] Marc Davis and P. J. E. Peebles. A survey of galaxy redshifts. v. the two-point position

and velocity correlations. *Astrophys. J.*, 267:465, 1983. doi: 10.1086/160884.

[10] A. J. S. Hamilton. Toward better ways to measure the galaxy correlation function. *Astrophys. J.*, 417:19, 1993. doi: 10.1086/173288.

[11] Stephen D. Landy and Alexander S. Szalay. Bias and variance of angular correlation functions. *Astrophys. J.*, 412:64, 1993. doi: 10.1086/172900.

[12] John von Neumann. Various techniques used in connection with random digits. In *Monte Carlo Method*, volume 12 of *National Bureau of Standards Applied Mathematics Series*, pages 36–38. U.S. Government Printing Office, Washington, DC, 1951.

[13] M. H. Quenouille. Notes on bias in estimation. *Biometrika*, 43:353–360, 1956. doi: 10.1093/biomet/43.3-4.353.

[14] John W. Tukey. Bias and confidence in not-quite large samples (abstract). *Ann. Math. Statist.*, 29:614, 1958. doi: 10.1214/aoms/1177706647.

[15] P. Norberg, C. M. Baugh, E. Gaztañaga, and D. J. Croton. Statistical analysis of galaxy surveys – i. robust error estimation for two-point clustering statistics. *Mon. Not. R. Astron. Soc.*, 396:19–38, 2009. doi: 10.1111/j.1365-2966.2009.14389.x.

[16] J. Hartlap, P. Simon, and P. Schneider. Why your model parameter confidences might be too optimistic. unbiased estimation of the inverse covariance matrix. *Astron. Astrophys.*, 464:399, 2007. doi: 10.1051/0004-6361:20066170.

[17] Elena Sellentin and Alan F. Heavens. Parameter inference with estimated covariance matrices. *Mon. Not. R. Astron. Soc.*, 456:L132–L136, 2016. doi: 10.1093/mnrasl/slv190.

[18] G. D'Agostini. On the use of the covariance matrix to fit correlated data. *Nuclear Instruments and Methods in Physics Research Section A*, 346:306–311, 1994. doi: 10.1016/0168-9002(94)90719-6.

[19] OpenStreetMap contributors. OpenStreetMap. https://www.openstreetmap.org, 2026. Overpass API で 2026-11-01T00:00:00Z 時点のデータを取得。ODbL.

[20] Natural Earth. 1:10m physical vectors: Coastline (version 5.1.2). https://www.naturalearthdata.com/, 2022. パブリックドメイン.

[21] Coles Group Limited. 2026 full year results release. https://www.colesgroup.com.au/DownloadFile.axd?file=%2FReport%2FComNews%2F20260825%2F03125402.pdf, 2026. 2026-08-25. Appendix 1: Number of retail stores.

[22] Woolworths Group Limited. F26 results announcement. https://www.woolworthsgroup.com.au/content/dam/wwg/investors/asx-announcements/2026/3108623.pdf, 2026. 2026-08-26. Appendix Six: Five-year store and trading area analysis.

[23] Australian Bureau of Statistics. Regional population, 2024–25 financial year: Australian population grid 2025. https://www.abs.gov.au/statistics/people/population/regional-population/2024-25, 2026. GeoTIFF, 1 km grid, EPSG:3577. CC BY 4.0.

[24] John P. Snyder. *Map Projections: A Working Manual*. Number 1395 in U.S. Geological

Survey Professional Paper. U.S. Government Printing Office, Washington, DC, 1987. doi: 10.3133/pp1395.

[25] 株式会社トライアルホールディングス. 本決算説明資料 2026 年 6 月期. https://pdf.irpocket.com/C141A/xoA3/MuFf/T9TP/zgQJ.pdf, 2026. 2026-08-13。地域別の出店状況（2026 年 6 月末時点）.

[26] 株式会社トライアルカンパニー. 店舗検索. https://www.trial-net.co.jp/shops, 2026. 店舗の業態・緯度・経度・都道府県を 2026-11-03 に取得.

[27] 株式会社西友. 店舗検索. https://www.seiyu.co.jp/searchshop/, 2026. 店舗の名前・住所・緯度・経度を 2026-11-04 に取得.

[28] 株式会社コスモス薬品. 2026 年 5 月期 決算短信〔日本基準〕（連結）. https://www2.jpx.co.jp/disc/33490/140120260625579741.pdf, 2026. 2026-07-13、TDnet。経営成績等の概況の地区別店舗数.

[29] 株式会社コスモス薬品. 店舗検索. https://www.cosmospc.co.jp/shop/, 2026. 全店舗の一覧（店名・住所・地区）をブラウザで 2026-11-12 に保存.

[30] 国土地理院. 住所検索 api. https://msearch.gsi.go.jp/address-search/AddressSearch, 2026. 2026-11-13 に住所から座標を引いた.

[31] 総務省統計局. 令和 2 年国勢調査 地域メッシュ統計（世界測地系 1 km メッシュ、人口及び世帯）. https://www.e-stat.go.jp/gis/statmap-search?type=1, 2022. 統計表 T001140。e-Stat から 1 次メッシュごとに取得（2026-11-03）。政府標準利用規約 第 2.0 版.

[32] 総務省統計局. 令和 2 年国勢調査 人口等基本集計 結果の要約. https://www.stat.go.jp/data/kokusei/2020/kekka/pdf/summary_01.pdf, 2021.

[33] Walmart Inc. Form 10-k for the fiscal year ended january 31, 2026. https://www.sec.gov/Archives/edgar/data/104169/000010416926000055/wmt-20260131.htm, 2026. 2026-03-13 提出。売上高は Item 7、店数は Item 2. Properties.

[34] Natural Earth. 1:10m cultural vectors: Admin 1 – states, provinces (version 5.1.2). https://www.naturalearthdata.com/, 2022. パブリックドメイン.

[35] Marcello Schiavina, Sergio Freire, Alessandra Carioli, and Kytt MacManus. GHS-POP R2023A – GHS population grid multitemporal (1975–2030). European Commission, Joint Research Centre. https://data.jrc.ec.europa.eu/dataset/2ff68a52-5b5b-4a22-8f40-c41da8332cfe, 2023. 2020 年、1 km 格子（ESRI:54009）。CC BY 4.0.

[36] U.S. Census Bureau. Table 2. resident population for the 50 states, the district of columbia, and puerto rico: 2020 census. https://www2.census.gov/programs-surveys/decennial/2020/data/apportionment/apportionment-2020-table02.xlsx, 2021.

[37] 総務省統計局. 市区町村別メッシュ・コード一覧. https://www.stat.go.jp/data/mesh/m_itiran.html, 2026. 都道府県ごとの CSV を 2026-11-13 に取得。政府標準利用規約 第

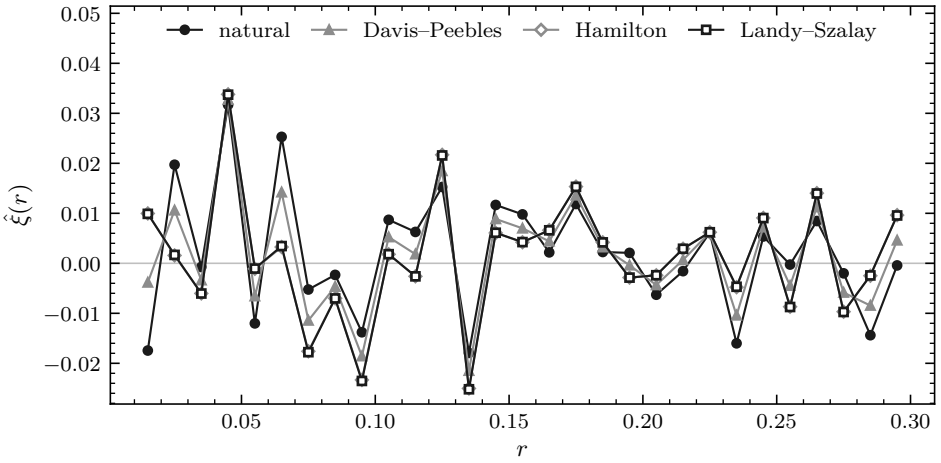


図 13　単位正方形のポアソン点に対する 4 つの推定量。

2.0 版.

[38] Thomas J. Holmes. The diffusion of wal-mart and economies of density. *Econometrica*, 79:253–302, 2011. doi: 10.3982/ECTA7699.

[39] 株式会社トライアルホールディングス. 福岡で実験を重ねた“次世代ローコスト小売モデル”がついに東京展開 街の生活必需店「trial go」都内初出店. https://trial-holdings.inc/news/release/690c71f4b36659dd5277a8d1/, 2025. ニュースリリース、2025-11-07.

[40] 西村夢音. トライアルの小型スマートストア「trial go」都内初出店. https://www.watch.impress.co.jp/docs/news/2061411.html, 2025. Impress Watch、2025-11-07.

[41] The mathlib Community. The Lean mathematical library. In *Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2020)*, pages 367–381, 2020. doi: 10.1145/3372885.3373824.

付録 A　toy model

実装が正しければ、相関のない点に当てた推定量は 0 のまわりに散るはずである。単位正方形に一様に 2000 点を撒いてデータとし、別に 8000 点を撒いてランダムカタログとした。図 13 のとおり、4 つの推定量はどれも ± 0.04 の範囲で 0 のまわりに散る。Hamilton と Landy–Szalay の値はほぼ重なり、natural は他の 3 つと違う揺れ方をする。

jackknife の誤差が正しく出るかは、toy model で確かめた。単位正方形に相関のない点を撒き直して 60 回計算し、$\hat{\xi}_{\mathrm{LS}}$ の実際のばらつきと、jackknife で見積もった標準偏差を比べた。jackknife の値は実際のばらつきより 7–20% 大きかった。toy model では、jackknife は誤差をやや大きめに見積もったことになる。

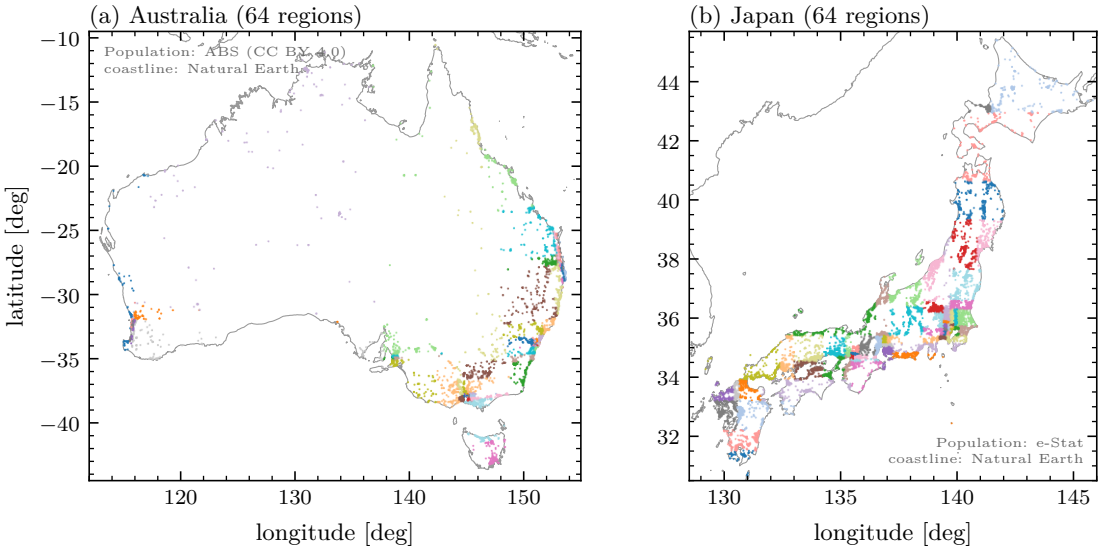


図 14　jackknife の 64 の領域。人口に比例するランダムの点を、領域ごとの色で打った。本文の誤差に使った領域と同じ。

付録 B　jackknife の領域

2.5 節の 64 の領域を図 14 に示す。点は人口に比例するランダムの点 20,000 個で、領域ごとに色を変えた。どの領域にも 312 か 313 個の点が入る。点の数をそろえたので、シドニー、メルボルン、東京、大阪のような大都市では領域が小さく、オーストラリアの内陸や北海道では大きい。色は 20 色を使い回しているので、離れた領域が同じ色になることがある。

付録 C　選択関数を 10 km 格子の表で求める

帰無仮説の模擬データは 1 個ごとに、選択関数 S_R（式 (13)）を何十万もの点で求め直す。棄却法の候補の点と、γ を決める参照の点のそれぞれで、まわりの店と参照の点を足し合わせるからである。点ごとに厳密に求めると、トライアルと西友の相互相関の模擬データ 1 個に 64.3 秒かかり、その 83% が S_R の計算だった。模擬データを 100 個ずつ、いくつもの設定で作るには時間がかかりすぎる。

S_R は幅 $R \geq 50\,\mathrm{km}$ でならした関数なので、それより十分狭い範囲ではほとんど変わらない。そこで、人が住む 10 km 格子（2 次メッシュ、4,300 個）の中心で S_R を一度だけ求めて表にし、各点ではその点を含む格子の値を使った。S_R を求める回数は約 40 万回から 8,600 回（二社）になり、模擬データ 1 個の時間は 9.9 秒になった。格子の中では S_R を一定とみなすので、ずれは格子の大きさ（約 10 km）と S_R の傾きの積の程度で、店の分布が急に変わる所で大きくなる。

表の値と厳密な値の差を、人口に比例して撒いた 4 万点で調べた（表 11）。相対誤差の中央値は

表 11　10 km 格子の表で求めた S_R と厳密な S_R の差。人口に比例して撒いた 4 万点で求めた。

R	チェーン	相対誤差の中央値	相対誤差の 95% 点	誤差の最大 $/S_{\max}$
50 km	トライアル	0.015	0.055	0.114
50 km	西友	0.009	0.114	0.080
100 km	トライアル	0.005	0.019	0.026
100 km	西友	0.005	0.050	0.030

表 12　選択関数を入れた $\hat{\xi}_{\rm LS}$ と $\hat{\xi}_{\times}$ の、8 km から $R/2$ までのビンの単純な平均（本文と同じまとめ方）。種 1–5 の平均 ± 標準偏差と、表の方法から厳密な方法を引いた差の平均。

R	量	厳密	表	差の平均
50 km	トライアル	$+0.082 \pm 0.006$	$+0.082 \pm 0.006$	-0.0006
50 km	西友	$+0.175 \pm 0.008$	$+0.175 \pm 0.008$	-0.0006
50 km	相互相関	-0.169 ± 0.005	-0.168 ± 0.004	$+0.0012$
100 km	トライアル	$+0.216 \pm 0.021$	$+0.218 \pm 0.016$	$+0.0016$
100 km	西友	$+0.165 \pm 0.006$	$+0.164 \pm 0.008$	-0.0002
100 km	相互相関	-0.209 ± 0.006	-0.213 ± 0.009	-0.0040

0.5–1.5%、95% 点は 2–11% だった。相対誤差は S_R がほぼ 0 の所で大きくなるので、S_R の最大値 $S_{\max}$ で割った誤差の最大も示した。棄却法で候補を残す確率は $S_R/S_{\max}$ なので、こちらが撒く点への影響の目安になる。

結果への影響は、実際の店で乱数の種を 1 から 5 に変え、厳密な方法と表の方法の両方で計算して調べた（表 12）。二つの方法の差の平均はどの量でも 0.005 より小さかった。これは種を変えたときの揺らぎ（0.004–0.021）と同じか、それより小さく、本文の jackknife の誤差（0.09–0.17）より一桁小さい。

模擬データの分布も比べた。$R = 50$ km で、密集地への寄り方もそろえた模擬データを表の方法で 100 個作った。相互相関の値は平均 -0.017、標準偏差 0.036、最小 -0.146 で、厳密な方法の 100 個（平均 -0.018、標準偏差 0.037、最小 -0.130）とほぼ同じだった。西友とトライアルの $\hat{\xi}_{\rm LS}$ も、平均と標準偏差が 0.001 以内で一致した。同じ種でも、二つの方法の模擬データ 1 個ずつの値は 0.01–0.02 ほど違う。表の値で候補を残すかを決めるので、撒かれる点そのものが変わるためである。模擬データでは分布だけを使うので、この違いは問題にならない。

以上から、模擬データの一部では表の方法を使った。使ったのは、表 2 のうち日本の $R = 50$、100 km 以外の模擬データと、一方のチェーンを実際の位置のまま残す模擬データ（4.3 節）である。実際の店の値は、本文のどこでも厳密な方法で求めている。時間は Apple M1 の Mac で、1 つのスレッドで測った。

付録 D　DD の分散

この付録ではデータ同士の組の数 DD の分散を、店の数を固定した場合と、店の数もポアソン分布で揺らした場合とで導く。店の数を固定すると、分散のうち店の数の 3 乗に比例する項は、もう一点がそのビンの距離に入る確率が点の位置によって違うことだけから生じる。ポアソン分布で揺らした場合の 3 乗の項には、これに店の数の揺らぎによる分が足される。本稿の模擬データは店と同じ数の点を撒くので（2.5 節）、前者がそのばらつきに当たる。

D.1　設定

n 個の点が領域の上の確率密度 $p(\boldsymbol{x})$（$\int p\, d^2x = 1$）に従って互いに独立に撒かれているとする。人口に比例するランダムカタログなら p は人口の密度に比例し、選択関数を入れた模擬データなら $S_R\,\rho^\gamma$ に比例する。ある距離のビンについて、二点 $\boldsymbol{x}, \boldsymbol{y}$ の距離がそのビンに入れば 1、入らなければ 0 をとる関数を $\theta(\boldsymbol{x}, \boldsymbol{y})$ と書く。点 i と点 j について $\theta_{ij} = \theta(\boldsymbol{x}_i, \boldsymbol{x}_j)$ とすると、

$$DD = \sum_{i<j} \theta_{ij} \tag{23}$$

である。

点 $\boldsymbol{x}$ に置いた点から見て、もう一点がそのビンの距離に入る確率を

$$f(\boldsymbol{x}) = \int \theta(\boldsymbol{x}, \boldsymbol{y})\, p(\boldsymbol{y})\, d^2y \tag{24}$$

と書く。f を使って三つの量を定義する。記号は Landy と Szalay [11] にならう。

$$\begin{aligned} G_p &= \mathbb{E}[\theta_{12}] = \int p\, f\, d^2x, \\ G_t &= \mathbb{E}[\theta_{12}\,\theta_{13}] = \int p\, f^2\, d^2x, \\ G_q &= \mathbb{E}[\theta_{12}\,\theta_{34}] = G_p^2. \end{aligned} \tag{25}$$

G_p は一つの組がビンに入る確率、G_t は一点を共有する二つの組がともにビンに入る確率、G_q は点を共有しない二つの組がともにビンに入る確率である。G_t では共有する点 $\boldsymbol{x}_1$ を固定すると、残りの二点は独立にそれぞれ確率 $f(\boldsymbol{x}_1)$ でビンに入る。G_q では二つの組が独立なので、G_p の 2 乗になる[*3]。

[*3] Landy と Szalay は領域を K 個のセルに分け、点が同じセルに重ならないように数えたので、G_q と G_p^2 は $1/K$ の程度だけ違う。ここでは $K \to \infty$ の連続の場合を考える。

D.2　店の数を固定した場合

式 (23) の期待値は、組が $n(n-1)/2$ 個あるので

$$\mathbb{E}[DD \mid n] = \frac{n(n-1)}{2}\, G_p \tag{26}$$

である。

2 乗の期待値は二つの組 $\{i,j\}$ と $\{k,l\}$ のすべての順序つきの組み合わせについての和になる。二つの組の関係は三通りに分かれる。同じ組である場合は $n(n-1)/2$ 通りあり、それぞれの期待値は $\mathbb{E}[\theta_{12}^2] = G_p$ である。一点を共有する場合は、共有する点の選び方が n 通り、残りの二点の順序つきの選び方が $(n-1)(n-2)$ 通りで、合わせて $n(n-1)(n-2)$ 通りある。点を共有しない場合は、一つ目の組の選び方が $n(n-1)/2$ 通り、二つ目の組の選び方が $(n-2)(n-3)/2$ 通りある。したがって

$$\mathbb{E}[DD^2 \mid n] = \frac{n(n-1)(n-2)(n-3)}{4}\, G_q + n(n-1)(n-2)\, G_t + \frac{n(n-1)}{2}\, G_p \tag{27}$$

である。三通りの数を足すと組み合わせの総数になる。

$$\left[\frac{n(n-1)}{2}\right]^2 = \frac{n(n-1)(n-2)(n-3)}{4} + n(n-1)(n-2) + \frac{n(n-1)}{2}. \tag{28}$$

式 (26) の 2 乗をこの恒等式で三つに分けて式 (27) から引き、$G_q = G_p^2$ を使うと、

$$\mathrm{Var}[DD \mid n] = n(n-1)(n-2)\left(G_t - G_p^2\right) + \frac{n(n-1)}{2}\, G_p\left(1 - G_p\right) \tag{29}$$

を得る。点を共有しない二つの組は独立なので、その共分散は 0 で、分散に効かない。

n^3 の項の係数は式 (25) から

$$G_t - G_p^2 = \int p\left(f - G_p\right)^2 d^2x \tag{30}$$

と書ける。これは p に従って選んだ点で見た f の分散で、負にならない。どの点から見ても f が同じなら 0 になり、分散には n^2 の項だけが残る。f が点によって違う理由は二つある。一つは、境界の近くではビンの距離の輪の一部が領域の外にはみ出すことである。もう一つは、p が場所によって違うことである。ビンの距離 r が p の変わる尺度より短く、点が境界から r 以上離れていれば、輪の面積を A として $f(\boldsymbol{x}) \simeq A\, p(\boldsymbol{x})$ である。このとき式 (30) は A^2 と、点の位置で見た p の分散の積になる。人口に比例して撒く場合は、都市とそれ以外で p が大きく違うので、この項は小さいとは限らない。

D.3　店の数もポアソン分布で揺らした場合

店の数 n が平均 N のポアソン分布に従うとする。ポアソン分布の階乗モーメントは

$$\mathbb{E}\big[n(n-1)\cdots(n-k+1)\big] = \sum_{n=k}^{\infty} \frac{n!}{(n-k)!}\, \frac{N^n e^{-N}}{n!} = N^k \tag{31}$$

である。全分散の公式

$$\mathrm{Var}[DD] = \mathbb{E}\big[\mathrm{Var}[DD \mid n]\big] + \mathrm{Var}\big[\mathbb{E}[DD \mid n]\big] \tag{32}$$

の右辺を順に求める。第 1 項は式 (29) に式 (31) を当てて

$$\mathbb{E}\big[\mathrm{Var}[DD \mid n]\big] = N^3\big(G_t - G_p^2\big) + \frac{N^2}{2}\,G_p\big(1 - G_p\big) \tag{33}$$

である。第 2 項は、店の数の揺らぎによって式 (26) が揺らぐ分である。式 (28) の両辺に 4 をかけると $[n(n-1)]^2$ が階乗の積の和に分かれるので、式 (31) から $\mathbb{E}\{[n(n-1)]^2\} = N^4 + 4N^3 + 2N^2$ となる。よって

$$\mathrm{Var}\big[\mathbb{E}[DD \mid n]\big] = \frac{G_p^2}{4}\Big(\mathbb{E}\big\{[n(n-1)]^2\big\} - N^4\Big) = \Big(N^3 + \frac{N^2}{2}\Big)\,G_p^2 \tag{34}$$

である。二つを足すと G_p^2 の項が打ち消し合い、

$$\mathrm{Var}[DD] = N^3\,G_t + \frac{N^2}{2}\,G_p \tag{35}$$

を得る。

式 (29) と式 (35) の違いは、G_p^2 の項があるかどうかである。ポアソン分布で揺らした場合の $N^3 G_t$ には、f が一様でも $N^3 G_p^2$ が残る。これは店の数の揺らぎが DD の期待値を揺らす分で、店の数を固定すれば消える。式 (35) は店の数の揺らぎも含めた全分散で、条件つき分散をポアソン分布で平均した $\mathbb{E}[\mathrm{Var}[DD \mid n]]$ より $(N^3 + N^2/2)\,G_p^2$ だけ大きい。

D.4 Lean での検証

この付録の導出を、Lean 4 と数学ライブラリ Mathlib で検証した [41]。出発点は、n 個の点が同じ確率分布 $p(\boldsymbol{x})\,d^2x$ から独立に撒かれ、θ が 0 か 1 をとり二点について対称、という模型である。G_p と G_t は式 (25) のとおり積分で定義し、変数としては受け取らない。証明は `lean/RetailCosmology/` にあり、`cd lean && lake build` で検査できる。式と定理の対応は次のとおりである。

- 式 (25)：`integral_pair`、`integral_triple`、`integral_quad`。
- 式 (26)：`integral_DD`。
- 式 (27)：`integral_DD_sq`。二つの組の関係ごとの数え上げは `dd_second_moment` で示す。
- 式 (28)：`count_identity`。
- 式 (29)：`variance_DD`。
- 式 (30)：`Gtriple_sub_sq`。
- 式 (31)：`poisson_factorial_moment`。Mathlib のポアソン分布での期待値として示す。
- 式 (35)：`variance_DDmix`。店の数をポアソン分布で選び、点を独立に撒く確率空間の上で、DD の分散を求める。

- 式 (32)：`variance_DDmix_eq_total`。同じ確率空間で、全分散が式 (32) の右辺と等しいことを示す。
- 式 (36)：`integral_sum_binCount`、`integral_sum_binCount_mul`。

G_p と G_t の数値と、模型が店の分布に合うかどうかは、Lean では扱わず、Rust で確かめた（この付録の toy model での検算とオーストラリアでの値）。

D.5　toy model での検算

二つの式を単位正方形に一様に撒いた点で確かめた。ビンは $0.05 \le r < 0.10$ とし、G_p と G_t は式 (24) の f を角度について積分して求めた。$G_p = 2.1276 \times 10^{-2}$ は単位正方形の二点の距離の累積分布 $\pi r^2 - 8r^3/3 + r^4/2$ から求めた値と一致する。$G_t - G_p^2 = 1.52 \times 10^{-5}$ は、f が正方形の辺の近くで小さくなることから生じる。点の数を $n = 200$ に固定して 4000 回撒き直すと、DD の標本分散は 541 ± 12 で、式 (29) の 534 と合う。境界の項を落とした式（$G_t = G_p^2$）では 414 で、模擬データから標準誤差の 10 倍離れる。点の数を平均 200 のポアソン分布から選んで 4000 回撒き直すと、標本分散は 4231 ± 95 で、式 (35) の 4168 と合う。店の数を固定した模擬データに式 (35) を当てると、分散を 8 倍近く大きく見積もることになる。計算は `cargo test --lib variance -- --nocapture` で行った（`crates/retail-cosmology/src/engine/variance.rs`）。

D.6　オーストラリアでの値

オーストラリアの G_p と G_t を、人口に比例して撒いたランダムカタログから求めた。点 i から見てビンに入る点の数を c_i とすると、N 点のランダムカタログから

$$G_p = \frac{\sum_i c_i}{N(N-1)}, \qquad G_t = \frac{\sum_i c_i(c_i-1)}{N(N-1)(N-2)} \tag{36}$$

で偏りなく求まる。どちらも、異なる点の組または三つ組すべてについて、式 (25) の期待値の中身を平均したものである。距離のビンは二点相関関数と同じで、$N = 40{,}000$ とした（表 13）。N を 10,000、20,000、40,000 と変え、$N = 20{,}000$ で乱数の種を三通りに変えても、$(G_t - G_p^2)/G_p^2$ は 5% 以内しか動かない。

500 km より短い距離では、$(G_t - G_p^2)/G_p^2$ は 0.8–2.4 である。点の位置によって f が平均と同じ程度の幅で違うということである。陸地に一様に撒いたランダムカタログでは、同じ量は 0.004–0.12 にとどまる。したがって、この大きさは海岸線ではなく人口の濃淡から来る。店の数を固定した分散のうち、n^3 の項は n^2 の項の、Coles で 1.6–250 倍、Woolworths で 2.0–320 倍になる。n^2 の項より n^3 の項が大きいので、店の数を固定しても DD の相対標準偏差は 500 km より短い距離で 5–11% ある。ポアソン分布で揺らした場合の標準偏差は、500 km より短い距離で固定した場合の 1.2–1.5 倍、500–800 km で 2.0 倍である。

この n^3 の項は、人の多い所に点がたまたま多く落ちたかどうかで DD が揺らぐ分である。Landy–Szalay の推定量は DR を引くので、この揺らぎの大部分が打ち消され、分散はポアソン分

表 13　オーストラリアの幾何の因子と、DD の相対標準偏差 $\sqrt{\mathrm{Var}[DD]}/\mathbb{E}[DD]$。「固定」は店の数を固定した式 (29)、「ポアソン」は店の数をポアソン分布で揺らした式 (35) による。G_p と G_t は人口に比例して撒いた 40,000 点のランダムカタログから式 (36) で求めた。

			Coles（865 店）		Woolworths（1116 店）	
r [km]	$10^3 G_p$	$(G_t - G_p^2)/G_p^2$	固定	ポアソン	固定	ポアソン
2.0–3.2	1.02	0.89	0.082	0.107	0.069	0.091
3.2–5.0	2.26	0.96	0.075	0.101	0.064	0.088
5.0–8.0	4.90	1.00	0.072	0.099	0.063	0.087
8.0–12.6	10.04	1.03	0.071	0.098	0.062	0.086
12.6–20.0	18.86	0.94	0.067	0.096	0.059	0.084
20.0–31.7	27.34	0.80	0.062	0.092	0.054	0.081
31.7–50.2	27.60	0.89	0.065	0.094	0.057	0.083
50.2–79.6	25.13	1.30	0.078	0.104	0.069	0.091
79.6–126.2	24.68	1.54	0.085	0.109	0.075	0.096
126.2–200.0	24.63	1.94	0.095	0.117	0.084	0.103
200.0–317.0	31.50	2.28	0.103	0.124	0.091	0.109
317.0–502.4	44.75	2.42	0.106	0.126	0.093	0.111
502.4–796.2	241.53	0.31	0.038	0.078	0.034	0.069
796.2–1261.9	131.14	0.63	0.054	0.087	0.048	0.077
1261.9–2000.0	158.00	0.76	0.060	0.091	0.052	0.080

布の n^2 の項に近くなる [11]。図 15 で、4.1 節の jackknife の誤差と比べた。$\hat{\xi}_{\mathrm{N}} = dd/rr - 1$ の誤差は式 (29) から求めた DD の相対標準偏差の 0.6–1.8 倍で、$\hat{\xi}_{\mathrm{LS}}$ の誤差は n^2 の項だけから求めた値の 0.9–2.7 倍である。二つの式の値は 30–500 km で 7–16 倍離れており、二つの推定量の誤差はそれぞれの式の近くにある。

付録 E　推定量の恒等式

式 (10) の四つの推定量は、natural の値 $\hat{\xi}_{\mathrm{N}}$ と

$$\epsilon = \frac{dr}{rr} - 1 \tag{37}$$

の二つで書ける。ϵ はデータの点がランダムカタログの点の多い所にどれだけ寄っているかを表し、本文で密集地への寄り方を表すのに使った量である。式 (10) に $dd = rr(1+\hat{\xi}_{\mathrm{N}})$ と $dr = rr(1+\epsilon)$ を入れると、

$$\hat{\xi}_{\mathrm{LS}} = \hat{\xi}_{\mathrm{N}} - 2\epsilon, \tag{38}$$

$$\hat{\xi}_{\mathrm{DP}} = \frac{1+\hat{\xi}_{\mathrm{N}}}{1+\epsilon} - 1, \tag{39}$$

$$\hat{\xi}_{\mathrm{H}} = \frac{1+\hat{\xi}_{\mathrm{N}}}{(1+\epsilon)^2} - 1 \tag{40}$$

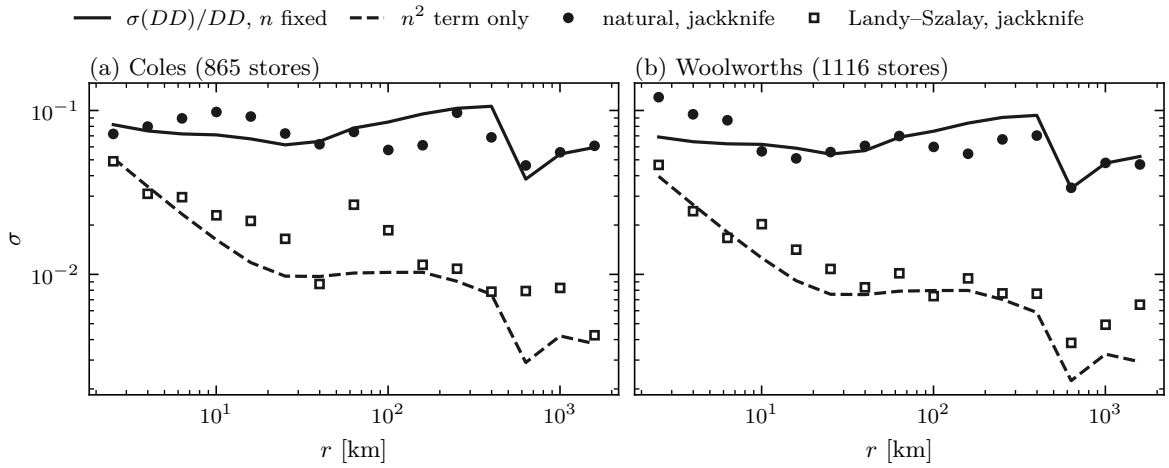


図 15　オーストラリアの店の $\hat{\xi}$ の jackknife の誤差（点）と、DD の相対標準偏差の式（線）。基準は人口に比例するランダムカタログ。実線は店の数を固定した式 (29)、破線はその n^2 の項だけ。黒丸は natural、白抜きの四角は Landy–Szalay の推定量の誤差。

を得る。式 (38) は本文の式 (22) と同じものである。どれも近似を使っていない。

ϵ が小さいとして $\hat{\xi}_{\mathrm{LS}}$ との差を ϵ で展開すると、

$$\begin{aligned} \hat{\xi}_{\mathrm{N}} - \hat{\xi}_{\mathrm{LS}} &= 2\epsilon, \\ \hat{\xi}_{\mathrm{DP}} - \hat{\xi}_{\mathrm{LS}} &= \bigl(1 - \hat{\xi}_{\mathrm{N}}\bigr)\,\epsilon + O(\epsilon^2), \\ \hat{\xi}_{\mathrm{H}} - \hat{\xi}_{\mathrm{LS}} &= -2\hat{\xi}_{\mathrm{N}}\,\epsilon + 3\bigl(1 + \hat{\xi}_{\mathrm{N}}\bigr)\,\epsilon^2 + O(\epsilon^3) \end{aligned} \tag{41}$$

となる。相関のない点では $\hat{\xi}_{\mathrm{N}}$ も ϵ も 0 のまわりで揺れるので、Hamilton と Landy–Szalay の差は揺れの 2 乗の大きさしかない。natural と Davis–Peebles は ϵ の 1 乗の大きさだけ Landy–Szalay から離れ、その差は natural の方が約 2 倍大きい。toy model で Hamilton と Landy–Szalay がほぼ重なり、natural がほかの三つと違う揺れ方をしたのはこのためと考えられる（図 13）。

ϵ の揺れは、データの点がたまたまランダムカタログの点の多い所に落ちたかどうかで生じる。そのとき dd も同じ向きに揺れるので、$\hat{\xi}_{\mathrm{N}}$ と 2ϵ の揺れの大部分は共通である。式 (38) で 2ϵ を引く Landy–Szalay は、この共通の揺れを打ち消す。付録 D の n^3 の項がこの揺れに当たり、オーストラリアの 30–500 km で $\hat{\xi}_{\mathrm{LS}}$ の誤差が $\hat{\xi}_{\mathrm{N}}$ より小さかったのはこのためである（図 15）。式 (38)–(40) は実装の関数で成り立つことをテストで確かめた（`cargo test --lib estimator_identities`、`crates/retail-cosmology/src/engine/estimator.rs`）。

付録 F　結果の確かめ

F.1　設定を変えた計算

オーストラリアの結果（4.1 節）が計算の設定に左右されないかを、設定を一つずつ変えて確かめた（表 14）。変えたのは、ランダムカタログの点の数（20,000 → 50,000）、人口の $\hat{\xi}$ に使う点の

表 14　8.0–50.2 km の 4 ビンをまとめた値と、$\hat{\xi}_{\mathrm{LS}}$（人口が基準）の χ^2 の上側確率。16 領域では領域の数がビンの数に近く、χ^2 を求められない。

設定	Woolworths		Coles		上側確率
	比	$dr/rr-1$	比	$dr/rr-1$	W / C
基本	0.857 ± 0.052	-0.061 ± 0.025	0.873 ± 0.069	-0.061 ± 0.040	0.72 / 0.73
ランダム 5 万点	0.881 ± 0.063	-0.061 ± 0.024	0.897 ± 0.060	-0.063 ± 0.043	0.89 / 0.71
人口の点 1 万点	0.872 ± 0.047	-0.061 ± 0.025	0.887 ± 0.071	-0.061 ± 0.040	0.72 / 0.73
32 領域	0.857 ± 0.060	-0.061 ± 0.028	0.873 ± 0.078	-0.061 ± 0.050	0.92 / 0.95
16 領域	0.857 ± 0.048	-0.061 ± 0.022	0.873 ± 0.097	-0.061 ± 0.065	—
種 2	0.915 ± 0.058	-0.054 ± 0.025	0.929 ± 0.088	-0.056 ± 0.045	0.69 / 0.65
種 3	0.848 ± 0.063	-0.067 ± 0.025	0.863 ± 0.096	-0.067 ± 0.041	0.95 / 0.61

数（2,000 → 10,000）、jackknife の領域の数（64 → 32, 16）、乱数の種（1 → 2, 3）である。上側確率は χ^2 分布と比べた値である。どの設定でも、4.1 節の結論は変わらない。16 領域では、一つの領域に入る人口が 64 領域のときの 4 倍になり、都市の中を細かく切らなくなる。それでも誤差は大きくても 1.6 倍（Coles の $dr/rr-1$）にとどまるので、領域が小さすぎて誤差を小さく見積もっている心配は少ない。32 領域と 16 領域でまとめた値が基本と同じなのは、ランダムカタログが同じで、まとめ方の重みが領域の分け方に依らないためである。

F.2　模擬データの分布の形

本文では、実際の値が模擬データの平均から何 σ 離れているかを目安に書いた。100 個では「0.01 未満」までしか言えないので、どれだけ離れているかを模擬データの平均と標準偏差で測ったものである。まとめた値は多くの店の組の数から作る量なので、中心極限定理から正規分布に近いと見込める。

トライアルの $R = 50\,\mathrm{km}$ の模擬データ（4.2 節）で確かめた。100 個を小さい順に並べると、25、50、75、95 番目の値は平均から -0.75σ、-0.04σ、$+0.59\sigma$、$+1.60\sigma$ で、正規分布の分位点 -0.69σ、-0.01σ、$+0.66\sigma$、$+1.60\sigma$ とよく合う。ただし最大は $+3.2\sigma$ で、正規分布で 100 個の最大として見込まれる $+2.6\sigma$ ほどより外にあり、裾はやや重い（歪度 0.34、超過尖度 0.81）。トライアルの 4.5σ は正規分布を外挿した目安で、これに対応する上側確率（4×10^{-6}）までは主張しない。ほかの σ の値も同じ意味の目安である。

模擬データの jackknife の標準偏差の中央値は 0.035 で、模擬データの実際のばらつき 0.024 より大きかった。toy model（付録 A）と同じく、jackknife は誤差を大きめに見積もっている。

付録 G　再現の手順

本文の数値と図表は、リポジトリ `https://github.com/inuverse44/retail-cosmology` の次のコマンドで作った。計算は Rust（`cargo run --release --bin …`）、描画は Python（`uv`

run scripts/…）で行う。日付の引数は、店のデータを取った日である（オーストラリアの OSM は 2026-11-01、トライアルのサイトは 2026-11-03、西友のサイトは 2026-11-04、米国の OSM は 2026-11-13）。描画のスクリプトの冒頭には、その図に要る計算のコマンドを順に書いてある。

■二点相関関数

- オーストラリア（図 7）：`xi -- 2026-11-01`
- トライアル（図 8、9）：`xi -- 2026-11-03 --country jp --footprint 50,100,200`
- 西友と相互相関（図 10）：`xi -- 2026-11-03 --country jp --stores $T,$S --cross TRIAL,SEIYU --footprint 50,100,200`
  `$T` はトライアルの店のファイル `data/trial_site/trial_site_clean_2026-11-03.csv`
  `$S` は西友の店のファイル `data/seiyu_site/seiyu_site_clean_2026-11-04.csv`
- 設定を変えた計算（表 14）：上に `--n-random 50000`、`--n-pop 10000`、`--depth 5`（32 領域）か `4`（16 領域）、`--seed 2` か `3` を足す。

■帰無仮説の模擬データ

- トライアルの $R = 50\,\mathrm{km}$（図 9(c)）：`scripts/null_mocks.sh 2026-11-03 jp 50 100 6`
- 西友と相互相関：`scripts/null_mocks_cross.sh`（引数は `scripts/plot_cross.py` の冒頭）
- χ^2 の校正（表 2）：模擬データを `scripts/chi2_mocks.sh` と上の二つで作り、`chi2_calibrate -- 2026-11-01 2026-11-03` で表にする。

■表

- 表 3：`osm_census -- 2026-11-01`
- 表 4：`trial_census -- 2026-11-03 2026-11-01`
- 表 5：`seiyu_census -- 2026-11-04 2026-11-03`
- 表 13：`var_dd -- 2026-11-01`
- 表 11：`cost -- 50 40000` と `cost -- 100 40000`
- 表 6：保存した一覧から `cosmos_census -- 2026-11-12 2026-11-01` が住所の一覧を書き、`scripts/cosmos_site/geocode.sh 2026-11-12` で座標を引いてから、もう一度 `cosmos_census` を走らせる。OSM の店は `scripts/osm/fetch.sh 2026-11-01 cosmos` と `osm_census -- 2026-11-01 cosmos`。
- 表 9：`xi -- 2026-11-03 --country jp --stores $T,$C --cross TRIAL,COSMOS --footprint 50,100`
  に `--tag cosmosfp`（全国）か `--prefectures 40-46 --tag kyushu`（九州）を付ける。

$C はコスモスの店のファイル data/cosmos_site/cosmos_site_clean_2026-11-12.csv。
九州の格子は scripts/estat/fetch_mesh_prefecture.sh で取る。模擬データは scripts/null_mocks_cosmos.sh、表は short_range -- 2026-11-03 cosmos。

- 表 7：scripts/osm/fetch.sh 2026-11-13 walmart_outside のあと osm_census -- 2026-11-13 walmart。
- 表 10：店は scripts/osm/fetch.sh 2026-11-13 walmart と osm_census -- 2026-11-13 walmart、
  人口は scripts/ghsl/fetch.sh、州境は scripts/naturalearth/fetch.sh で取る。
  実際の値は xi -- 2026-11-13 --country us --footprint 50,100,200 --tag fp、
  模擬データは scripts/null_mocks_walmart.sh 2026-11-13 100 6、
  表は short_range -- 2026-11-13 walmart。
- 表 8：scripts/null_mocks_short.sh で模擬データをビンごとの値つきで作り直し、short_range -- 2026-11-03 で表にする。
- 表 12：上の西友と相互相関の xi を --footprint 50,100 --seed s（$s = 1, \ldots, 5$）で、--selection-grid on を付けた場合と付けない場合について走らせる。

■図

- 図 1：scripts/plot_patterns.py
- 図 2：random_points のあと scripts/plot_random.py
- 図 3：selection_map -- 2026-11-03 のあと scripts/plot_selection.py
- 図 4：scripts/plot_stores.py
- 図 5：scripts/plot_stores_jp.py
- 図 6：scripts/plot_stores_us.py
- 図 11：scripts/plot_cosmos.py
- 図 12：scripts/plot_walmart.py
- 図 7：scripts/plot_xi.py
- 図 8：scripts/plot_compare.py
- 図 9：scripts/plot_footprint.py
- 図 10：scripts/plot_cross.py
- 図 13：toy のあと scripts/plot_toy.py
- 図 14：jk_regions -- --country au（jp）のあと scripts/plot_jk_regions.py
- 図 15：scripts/plot_var_dd.py