

Start-up 第1回課題発表 ～離散選択モデル～

2019/05/16

交通研B4 高谷和弘

離散選択モデル

- ▶ 松山、2007年のPPデータを利用
- ▶ 以下の8店舗を訪れた買い物トリップを対象に、目的地選択モデルを構築
- ▶ MNL(多項ロジットモデル)によりパラメータを推定

推定の対象とした目的地8店舗

	到着地コード	店舗名
1	5202	フジグラン松山店
2	5204	フジグラン重信SC
3	5206	パーティ姫原SC
4	5208	パーティ衣山SC
5	5212	伊予鉄高島屋
6	5213	三越松山店
7	5272	ジャスコ松山店
8	5303	ジョー・プラ



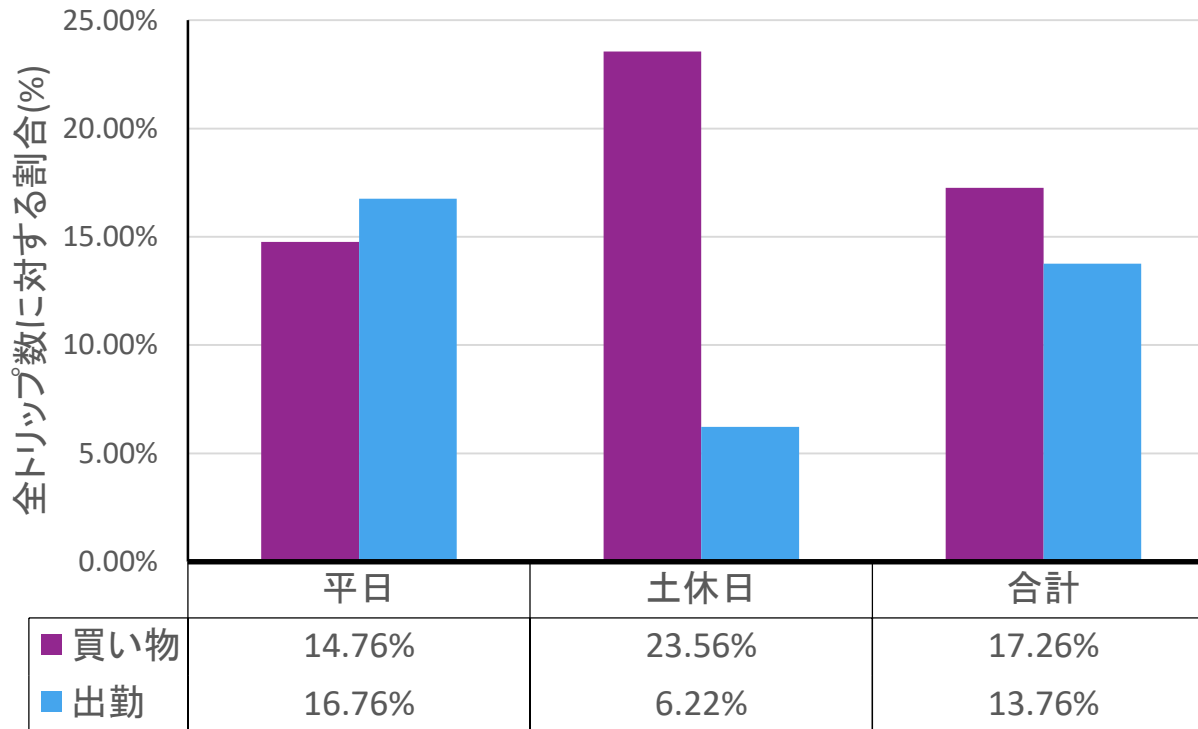
基礎集計

～大まかなデータの傾向をつかむ～

基礎集計(1)

- 2007松山PPの全トリップデータ(trip.csv)から、「目的」=「買い物」を抽出
- 平日か土休日かで場合分けをしてトリップ数を集計
- 比較用に、「目的」=「出勤・登校」でも同様の集計を行った

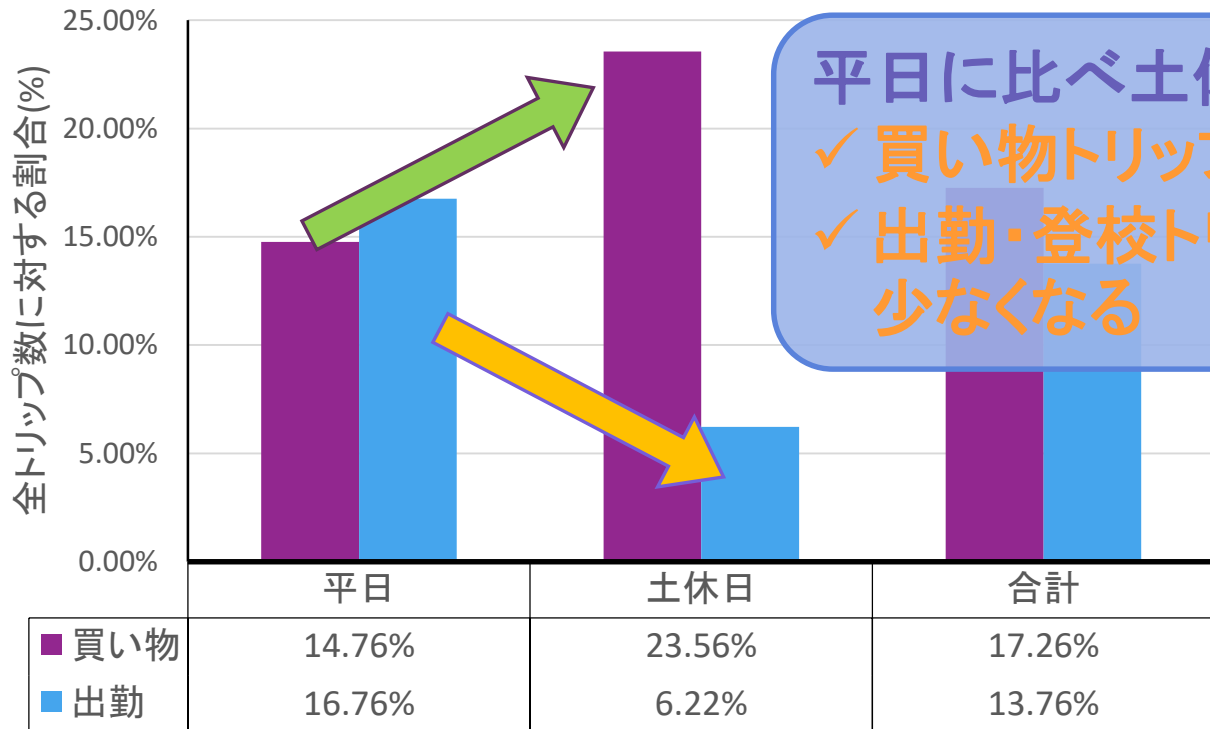
移動目的別割合 平日vs土休日



基礎集計(1)

- 2007松山PPの全トリップデータ(trip.csv)から、「目的」=「買い物」を抽出
- 平日か土休日かで場合分けをしてトリップ数を集計
- 比較用に、「目的」=「出勤・登校」でも同様の集計を行った

移動目的別割合 平日vs土休日

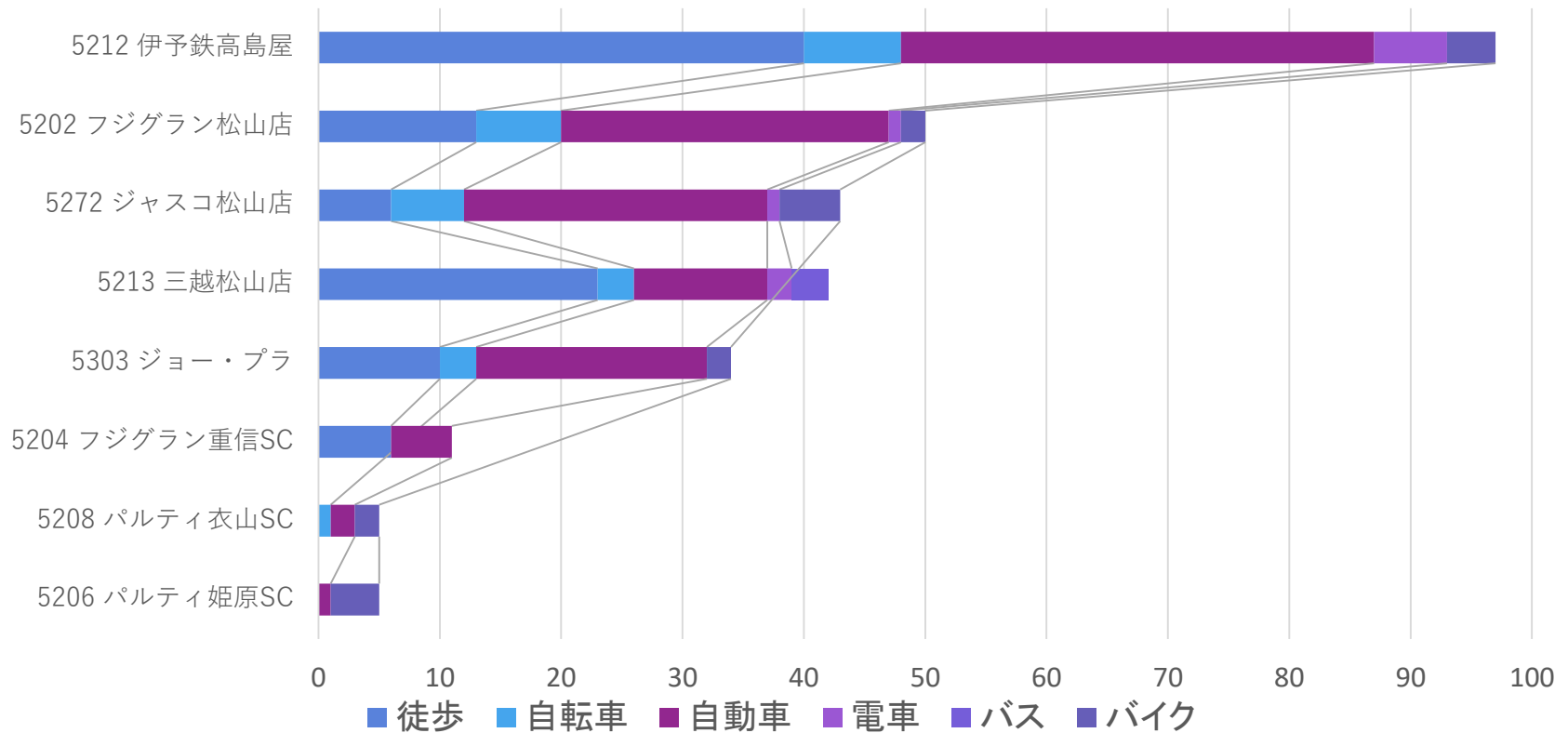


平日に比べ土休日は
✓ 買い物トリップの割合が多く
✓ 出勤・登校トリップの割合が少なくなる

基礎集計(2)

- 推定対象の大型商業施設8店舗を到着地とするトリップについて
- 店舗別・代表交通手段別に集計

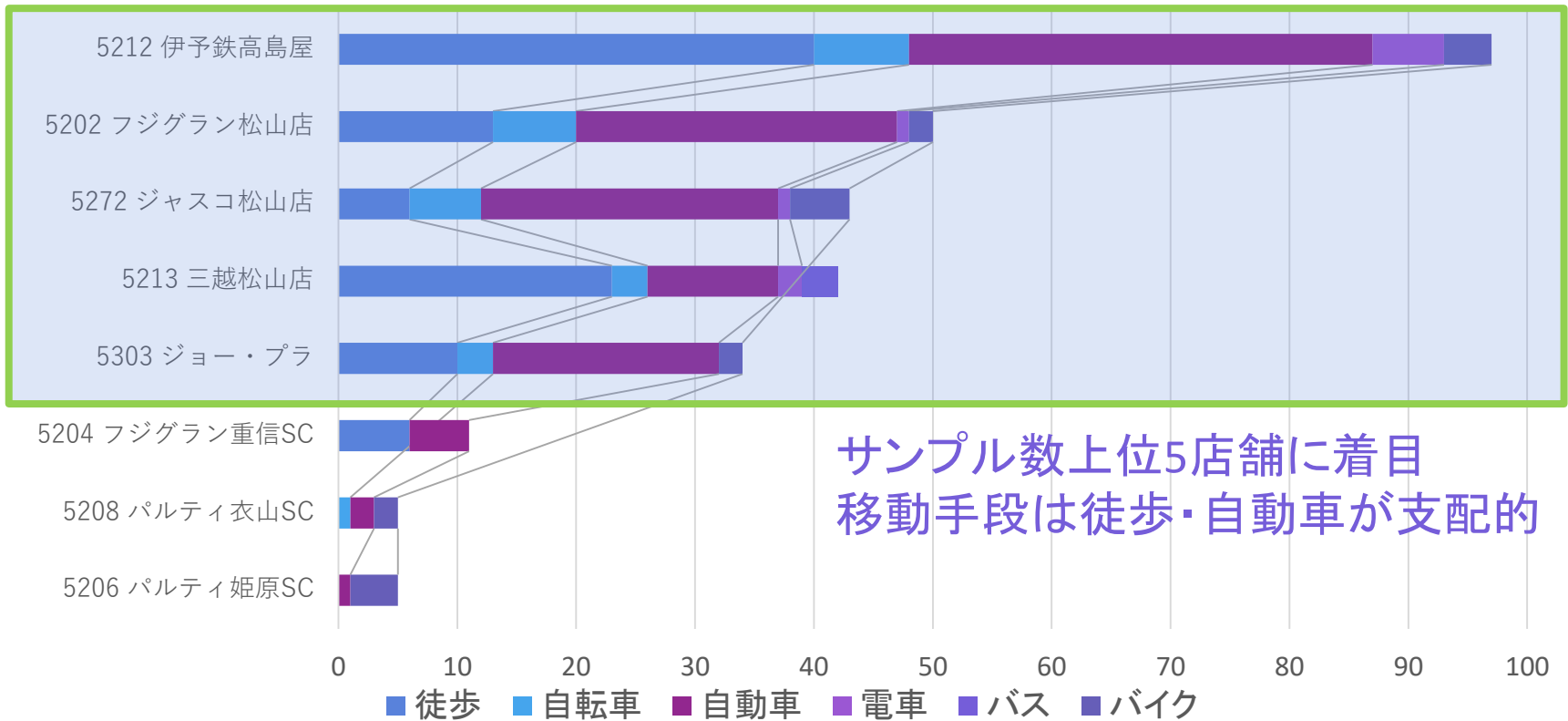
店舗別交通手段別買い物トリップ数



基礎集計(2)

- 推定対象の大型商業施設8店舗を到着地とするトリップについて
- 店舗別・代表交通手段別に集計

店舗別交通手段別買い物トリップ数



基礎集計(2)

▶ 上位5店舗について、徒歩利用と自動車利用を比較してみる

	店舗名	自動車	徒歩
1	フジグラン松山店	27	13
5	伊予鉄高島屋	39	40
6	三越松山店	11	23
7	ジャスコ松山店	25	6
8	ジョー・プラ	19	10



基礎集計(2)

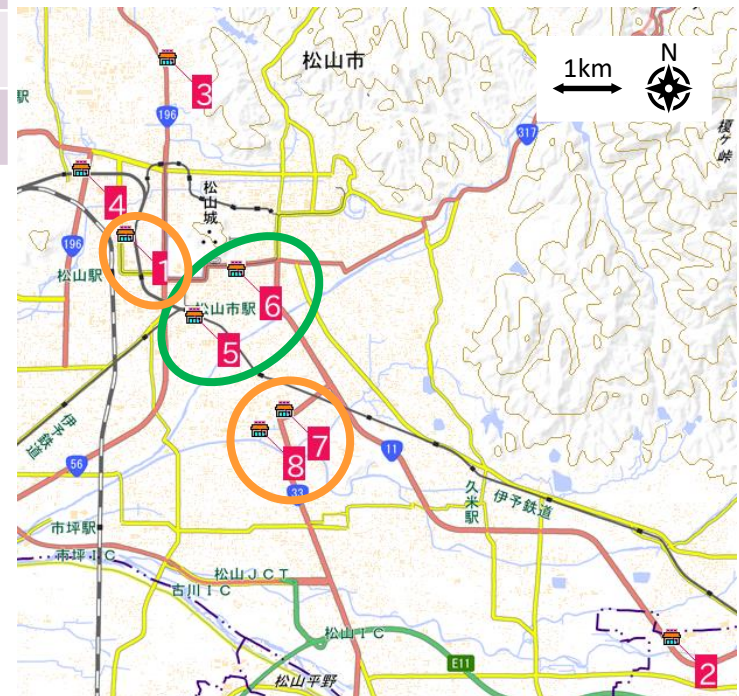
▶ 上位5店舗について、徒歩利用と自動車利用を比較してみる

	店舗名	自動車	徒歩
1	フジグラン松山店	27	13
5	伊予鉄高島屋	39	40
6	三越松山店	11	23
7	ジャスコ松山店	25	6
8	ジョー・プラ	19	10

高島屋・三越は徒歩が多い
← 中心市街地にあるため？

自動車が優勢の3店舗は
比較的郊外に立地

- ✓ 高島屋は利用者が8店舗中最多
- ✓ 自動車・徒歩どちらも同程度に多い
→ 中心市街地・郊外の双方を
商圏に持つ？



$$P(i) = \frac{\exp V_i}{\sum_{j \in C_n} \exp V_j}$$

目的地選択モデル

～MNLを用いたパラメータ推定～

目的地選択モデル

- 対象8店舗を目的地とする離散選択モデル
- 出発地からの距離等を説明変数として多項ロジットモデル(MNL)により推定

推定の手順

1. 効用関数(確定項)の形を設定
 - ✓ 説明変数の取り方を3パターン検証
 - ✓ 出発地からの距離/店舗の敷地面積/交通手段
2. 選択確率の式から対数尤度関数を計算
3. 対数尤度関数の最大化問題を解く(最尤推定法)
 - ✓ Rの最適化関数`optim()`を使用
 - ✓ 計算アルゴリズム: Nelder-Mead法、BFGS法、etc.
4. 推定結果が信頼できるかどうかをチェック
 - ✓ パラメータ推定値の正負
 - ✓ 各パラメータのt値の大小

効用関数(1)

➤ 設定1: 説明変数として「出発地からの距離」のみを考える

効用関数

個人jが店舗iを選択したときの効用の確定項

$$V_{i,j} = \alpha_i + \beta x_{i,j}$$

選択肢固有定数項

- 個人jに依存しない
- 各選択肢そのものが持つ特性を反映
- 基準化のため店舗8は $\alpha_8 = 0$ とする

説明変数

- 個人jの出発地から店舗iまでの直線距離

パラメータ

- 説明変数に対してかかる項
- 線形の関数を仮定
- 全選択肢において共通

選択確率の式

選択肢集合 C_n から選択肢 i を選択する確率、誤差項にガンベル分布を仮定

$$P(i) = \frac{\exp V_i}{\sum_{j \in C_n} \exp V_j}$$

推定結果(1)

BFGS法で推定

説明変数	パラメータ	t値
定数項(店舗1)	0.320	1.432
定数項(店舗2)	-1.193	-3.431**
定数項(店舗3)	-1.916	-3.997**
定数項(店舗4)	-1.458	-2.938**
定数項(店舗5)	1.225	5.770**
定数項(店舗6)	0.309	1.316
定数項(店舗7)	0.221	0.963
出発地からの距離[m/5000]	-0.609	-2.608**
サンプル数	287	
初期尤度	-596.80	
最終尤度	-500.59	
尤度比	0.1612	
修正済尤度比	0.1478	

* 5%有意 ** 1%有意

補足:t値について

- 推定値が初期値[0,...,0]からどの程度差があるかを示す値、絶対値が大きいほどそのパラメータの有意性は高い
- $|t| > 1.96$ なら有意水準5%で、 $|t| > 2.58$ なら有意水準1%で有意とみなす(自由度 ∞ のt分布で $p=0.05, 0.01$ となるt値)

効用関数(2)

➤ 設定2: 説明変数として「出発地からの距離」及び「店舗の敷地面積」を考える

cf. ハフモデル: 商圈分析に利用、店舗の魅力度として(売場)面積を採用

$$P_{i,j} = \frac{S_j / d_{i,j}^\lambda}{\sum_j S_j / d_{i,j}^\lambda}$$

効用関数

個人jが店舗iを選択したときの効用の確定項

$$V_{i,j} = \alpha_i + \beta_1 x_{1,i,j} + \beta_2 x_{2,i}$$

選択肢固有定数項

- 基準化のため店舗8は0とする

説明変数1

- 個人jの出发点から店舗iまでの直線距離

説明変数2

- 店舗iの敷地面積

推定結果(2)

BFGS法で推定

説明変数	パラメータ	t値
定数項(店舗1)	0.519	0.000
定数項(店舗2)	-1.076	-0.001
定数項(店舗3)	-1.749	-0.001
定数項(店舗4)	-1.155	0.000
定数項(店舗5)	1.501	0.001
定数項(店舗6)	0.363	0.001
定数項(店舗7)	0.252	0.001
出発地からの距離[m/5000]	-0.609	-2.608**
店舗の敷地面積[m ² /10000]	-0.473	0.000
サンプル数	287	
初期尤度	-596.80	
最終尤度	-500.59	
尤度比	0.1612	
修正済尤度比	0.1461	

* 5%有意 ** 1%有意

- t値を見る限り、説明変数として有意なのは距離のみ
 - 店舗の面積は選択決定に影響しない
- 1回目と比べ尤度比が変わっていない＝モデルの精度は上がっていない

効用関数(3)

- 設定3: 説明変数として「出発地からの距離」と「交通手段ダミー」を考える
- ✓ 2回目の推定から、店舗の敷地面積の影響は少ないと判断できる→除外
 - ✓ 集計結果では、交通手段が自動車か徒歩かで行き先が異なる傾向に
 - ✓ 主要な交通手段の違いは店舗の立地条件と相関がある

	到着地コード	店舗名	立地	主な交通手段
1	5202	フジグラン松山店	郊外	自動車
2	5204	フジグラン重信SC	郊外	
3	5206	パルティ姫原SC	郊外	
4	5208	パルティ衣山SC	郊外	
5	5212	伊予鉄高島屋	中心市街地	自動車・徒歩
6	5213	三越松山店	中心市街地	徒歩
7	5272	ジャスコ松山店	郊外	自動車
8	5303	ジョー・プラ	郊外	自動車

効用関数(3)

効用関数

個人jが店舗iを選択したときの効用の確定項

店舗2,3,4: $V_{i,j} = \alpha_i + \beta x_{i,j}$

店舗1,7,8: $V_{i,j} = \alpha_i + \beta x_{1,i,j} + \gamma_1 y_{1,j}$

店舗5,6: $V_{i,j} = \alpha_i + \beta x_{1,i,j} + \gamma_1 y_{1,j} + \gamma_2 y_{2,j}$

ダミー変数2

- 徒歩利用に関するダミー
- 徒歩で行っていれば1、それ以外なら0(実績値)
- 徒歩利用者の多かった店舗5,6で考慮

選択肢固有定数項

- 基準化のため店舗8は0とする

説明変数

- 個人jの出発地から店舗iまでの直線距離

ダミー変数1

- 自動車利用に関するダミー
- 自動車を利用していれば1、それ以外なら0(実績値)
- 選択者の多かった店舗1,5,6,7,8で考慮

推定結果(3)

BFGS法で推定

説明変数	パラメータ	t値
定数項(店舗1)	0.316	1.413
定数項(店舗2)	-0.902	-2.293*
定数項(店舗3)	-1.614	-3.136**
定数項(店舗4)	-1.060	-1.978*
定数項(店舗5)	0.883	3.882**
定数項(店舗6)	-0.054	-0.216
定数項(店舗7)	0.218	0.951
出発地からの距離[m/5000]	-0.749	-3.178**
自動車ダミー	0.662	1.395
徒歩ダミー	1.157	4.334**
サンプル数	287	
初期尤度	-596.80	
最終尤度	-490.56	
尤度比	0.1780	
修正済尤度比	0.1613	

* 5%有意 ** 1%有意

➤ 説明変数として距離以外に徒歩ダミーが有意水準1%で有意

◦ 歩行者にとって中心市街地にある店舗は相対的に魅力が高い

➤ 1,2回目と比べ尤度比が上がった＝モデルとしての精度が向上した

推定過程での気づき

- ▶ パラメータの初期値[0,...,0]に対し単純に尤度の最適化を行っただけでは、`optim()`の計算が収束せず、正しい推定値が得られない場合があった

→初期値をその都度変更して繰り返し最適化計算を実行

```
38 ### 対数尤度関数の最大化
39 while (TRUE) { # 収束するまで繰り返す
40   # パラメータ値の最適化: 最適化関数 optim
41   res <- optim(b0, func, gr=NULL, method = "BFGS", hessian = TRUE, control=list(fnscale=-1))
42
43   if (res$convergence == 0) break
44   else {
45     b0 <- res$par
46     print(b0)
47   }
48 }
```

- ▶ Nelder-Mead法で計算を行うとt値が正しく計算されなかった(NaNを出力)

→t値の算出の際に計算するヘッセ行列がおかしい?(未解決)

$$t_i = \frac{\hat{\beta}_i}{\sqrt{-diag(H^{-1})}}$$

ヘッセ行列の逆行列=分散共分散行列
対角成分=分散が負になってしまう??
BFGS法では問題なく計算できた

発展的考察

- 平日より土休日のほうが買い物トリップが多かったことについて、今回のモデル推定では加味できていない
 - ✓ そもそもいつ(何曜日に)買い物に出かけるか、出かけるかは目的地選択以前の問題: →Activity Model, RL

- 今回は交通手段の選択結果をダミー説明変数として考慮した
 - ✓ 本来、人間は目的地選択→交通機関選択の順で行動を決めて動く
 - ✓ 交通手段選択の推定については別途モデルを構築する必要がある

- 今回のモデルを用いて考察できそうなこと
 - ✓ 近隣に新規出店があった場合の利用変化予測

以上