松山PPデータのMNL推定

交通研B4 月田光

交通手段の離散選択モデル

- 2007年松山のPPデータを利用。
- MNLモデルによりパラメータ推定を行う。

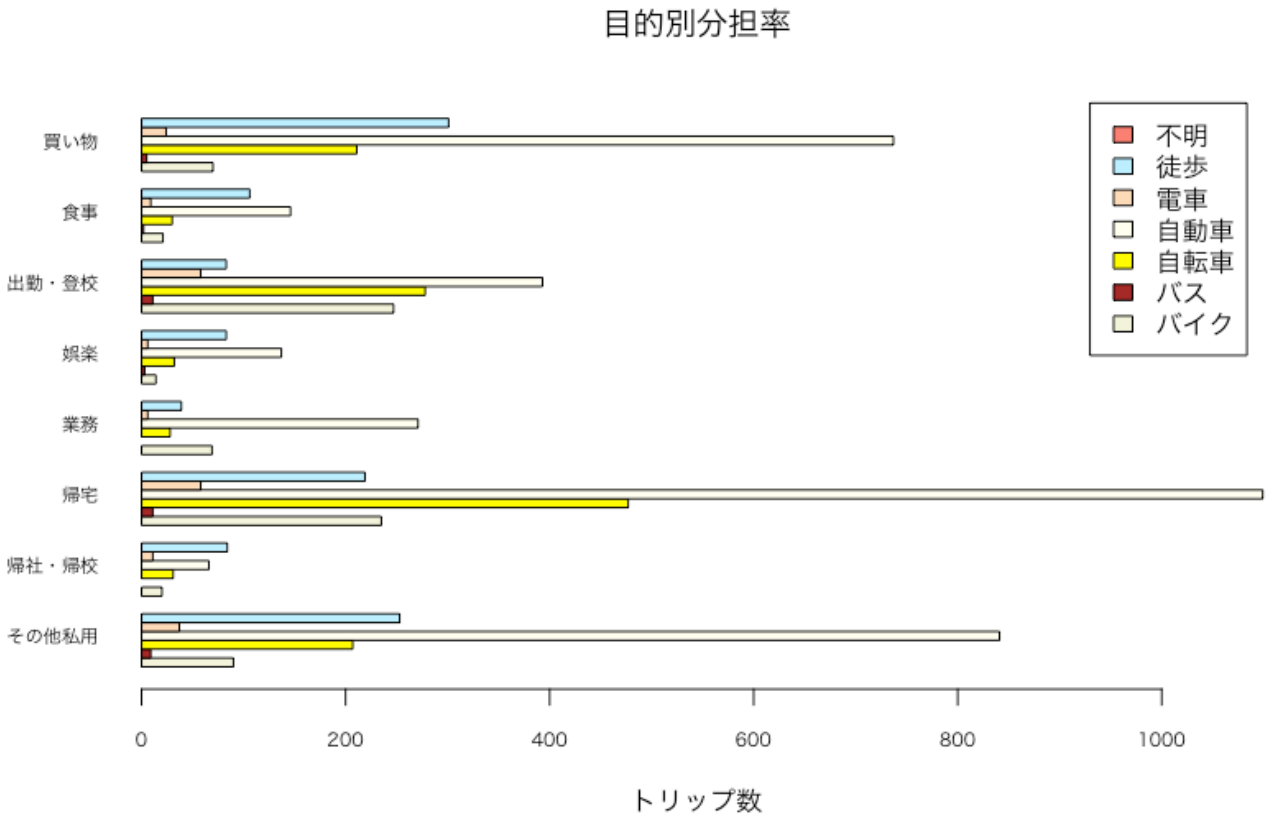
- 流れ
  - 基礎集計
  - 鉄道の傾向を探る
  - モデルの概要設計
  - 推定結果
  - 結果の分析
  - 確定項の一般化
  - 政策シミュレーション
  - 発展的考察、今後の課題

1.基礎集計

データの傾向を掴む

基礎集計①　目的と交通手段

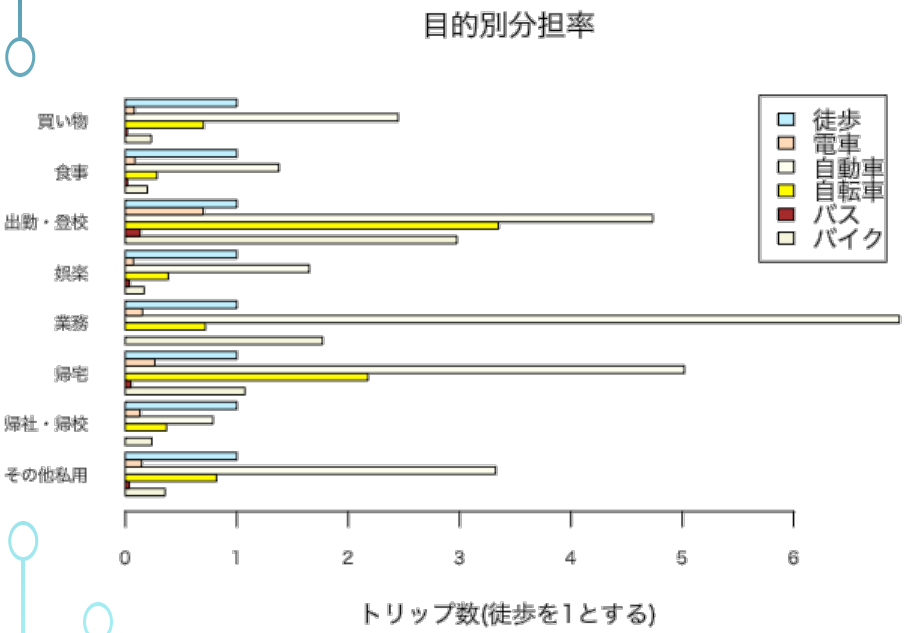
- 代表交通手段を抽出し、これと目的との関係を調べた。



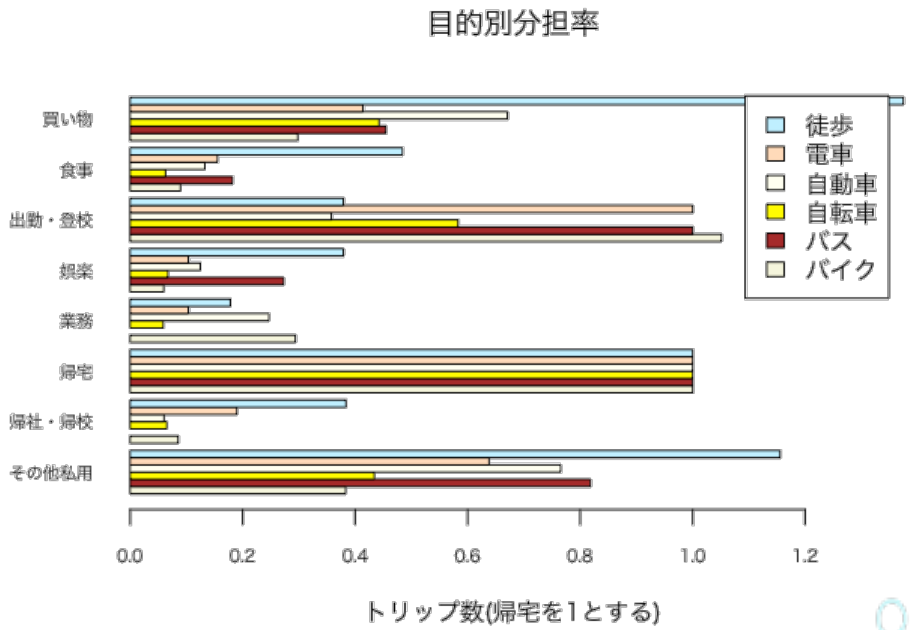
自動車が支配的で、次いで徒歩、自転車、バイクが多くなっている

基礎集計①　目的と交通手段

- 次に、徒歩のトリップ数、帰宅のトリップ数で正規化。



ほぼ全てにおいて自動車は多い
次に来るのは徒歩と自転車のところがある
出勤登校では自転車やバイクが比較的多い
帰宅でも、自動車に次いで自転車が多い
業務では自動車が圧倒的で、バイクが次ぐ

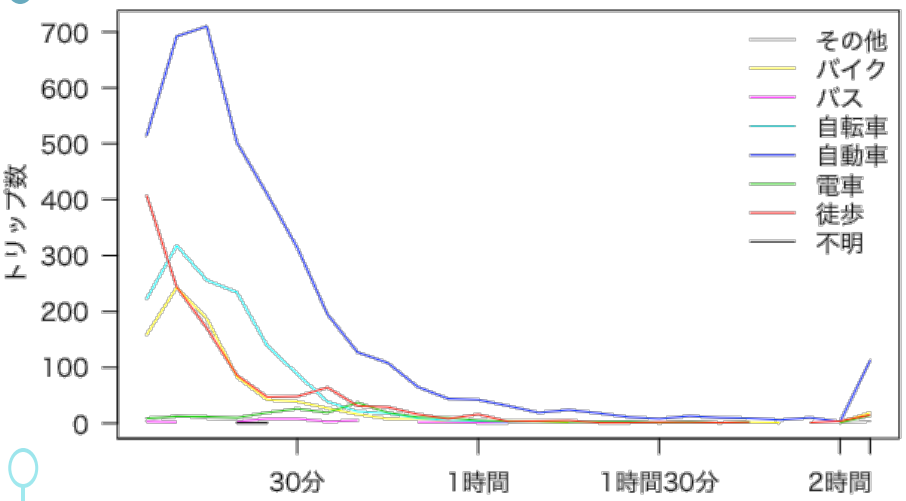


買い物と食事では徒歩が多い
出勤登校では電車バスが比較的多い
食事と娯楽では自転車は少ない
食事と娯楽では徒歩やバスが比較的多い

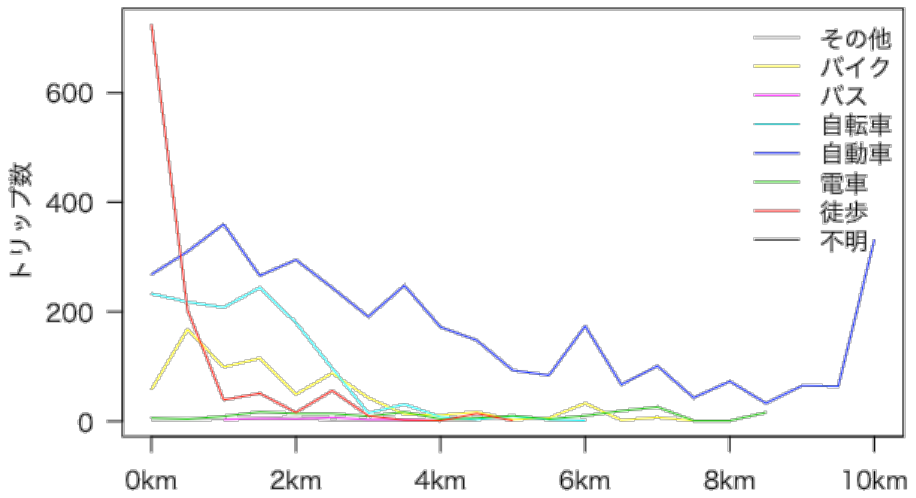
基礎集計②　時間・距離と交通手段

- 次に、時間や距離により手段の差を見る

手段別トリップ時間(5分ごとのヒストグラム)



手段別トリップ距離(500mごとのヒストグラム)



自動車・自転車・バイク・徒歩を見たときに、
左端(0~5分)を除くすべての時間において、
同じような比率→確定項は時間の関数？

左端(0~500m)において徒歩が顕著に高い
→徒歩の確定項は、0~500mで大きくなる

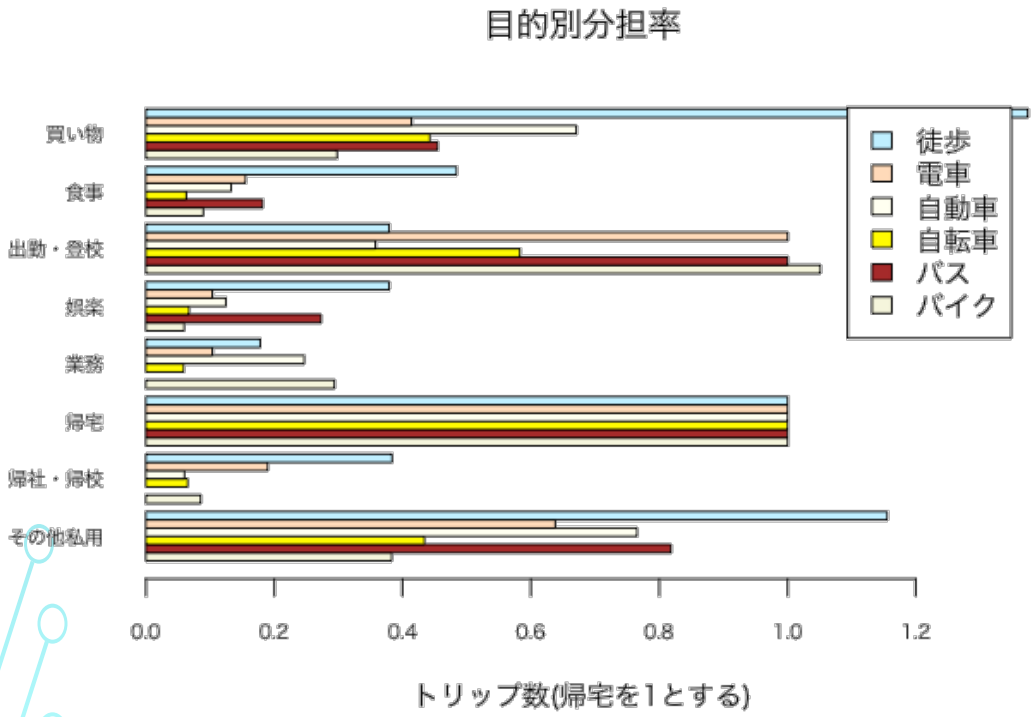
バスはトリップ数が少なくて傾向が掴めない
電車は、明らかに他と違う挙動をしている(ピークが違う)→確定項に時間・距離が入らないのでは？

2.鉄道の傾向を探る

時間や距離の関数ではなさそうな鉄道の確定項。では何の関数か。

鉄道の傾向①　目的を考える

- まず、目的別分担率から考える



・鉄道、バスは出勤・登校と帰宅で特に多い
→ルートが決まっている公共交通は、ルートが決まっている出勤・登校と相性が良い？

・鉄道は食事、娯楽、業務で特に少ない。ただし、食事については、徒歩が大きいだけと見ることもできる。

・娯楽においてバスが突出しているのも気になる。

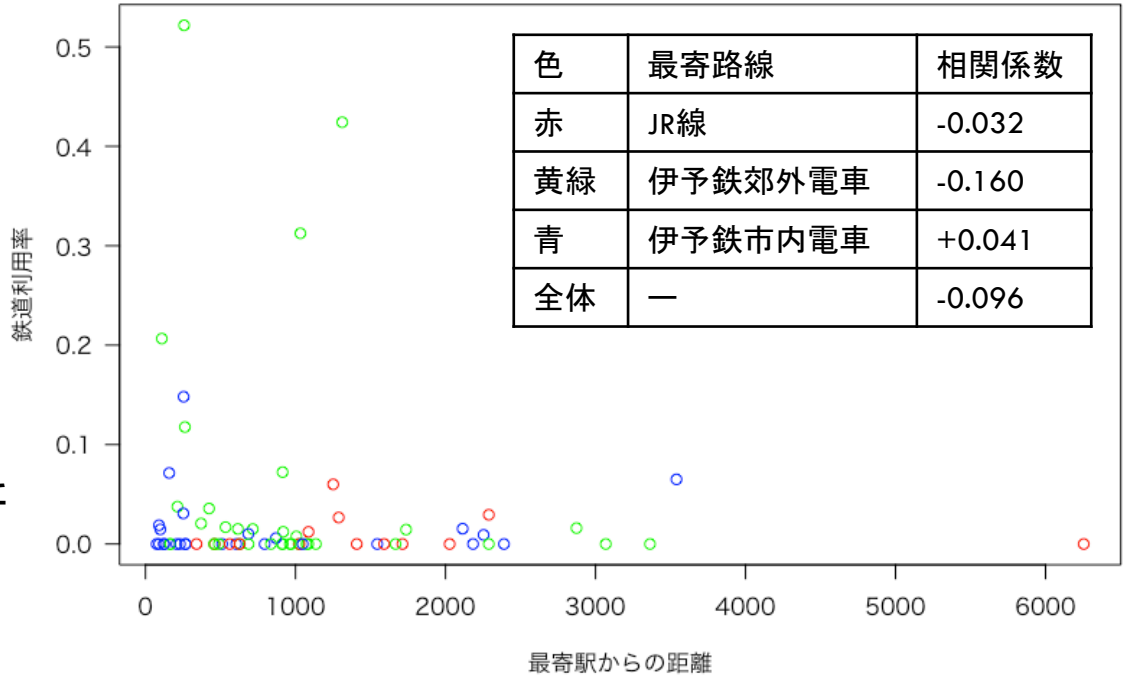
鉄道の傾向②　利用者を考える

- 次に、各手段を何人が利用しているか、
  また各人の利用率考える。

手段	人数
総数	86
徒歩	73
鉄道	37
自動車	80
自転車	62
バス	21
バイク	33

・徒歩、自動車、自転車はほとんどの人が利用するのに対し、鉄道やバスは利用する人数が限られる。
→**一部の人が何度も乗っている**のか。

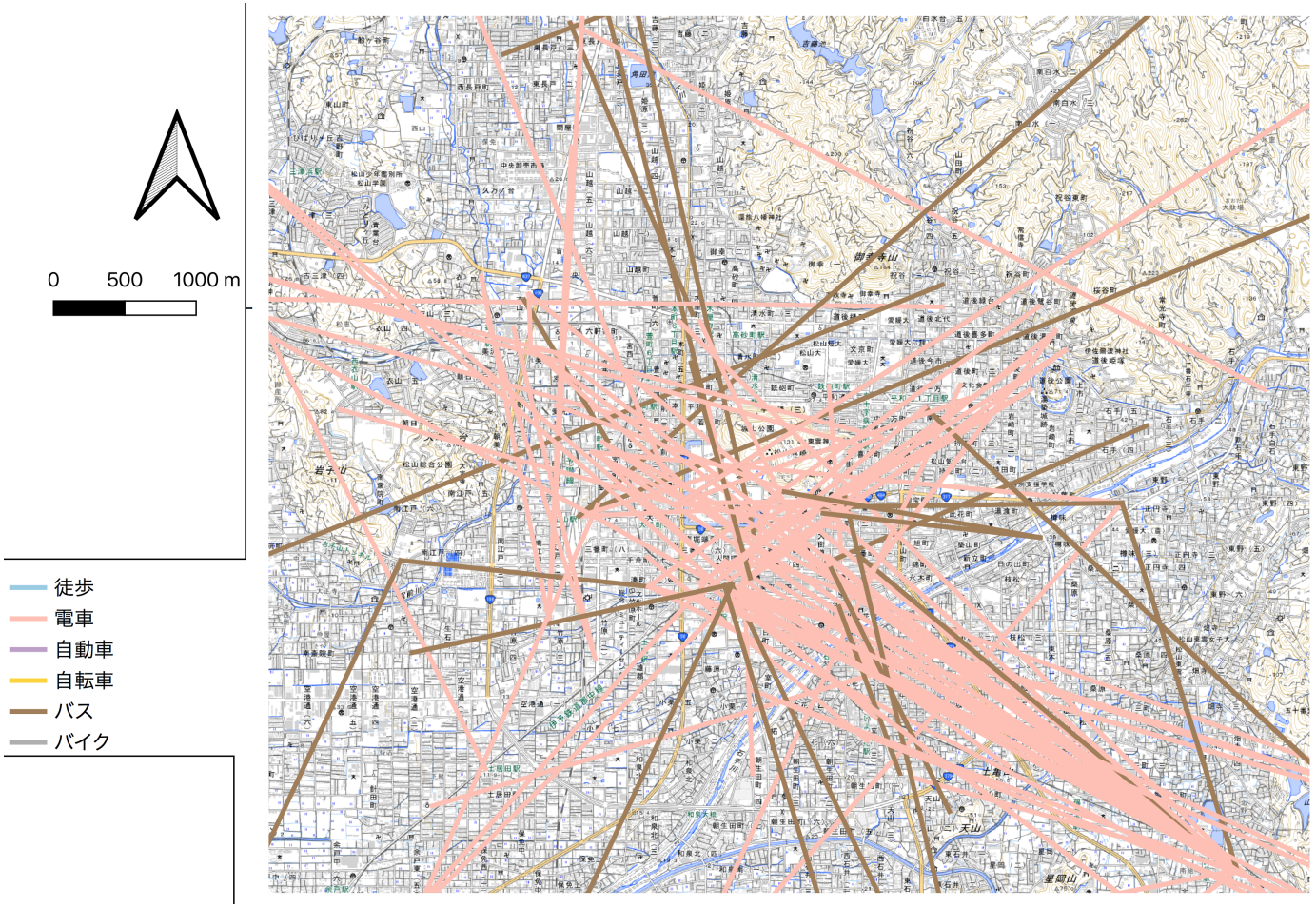
・それぞれの自宅と駅との距離が鉄道利用率に影響するかと考えた
→顕著な相関は見られなかった



色	最寄路線	相関係数
赤	JR線	-0.032
黄緑	伊予鉄郊外電車	-0.160
青	伊予鉄市内電車	+0.041
全体	―	-0.096

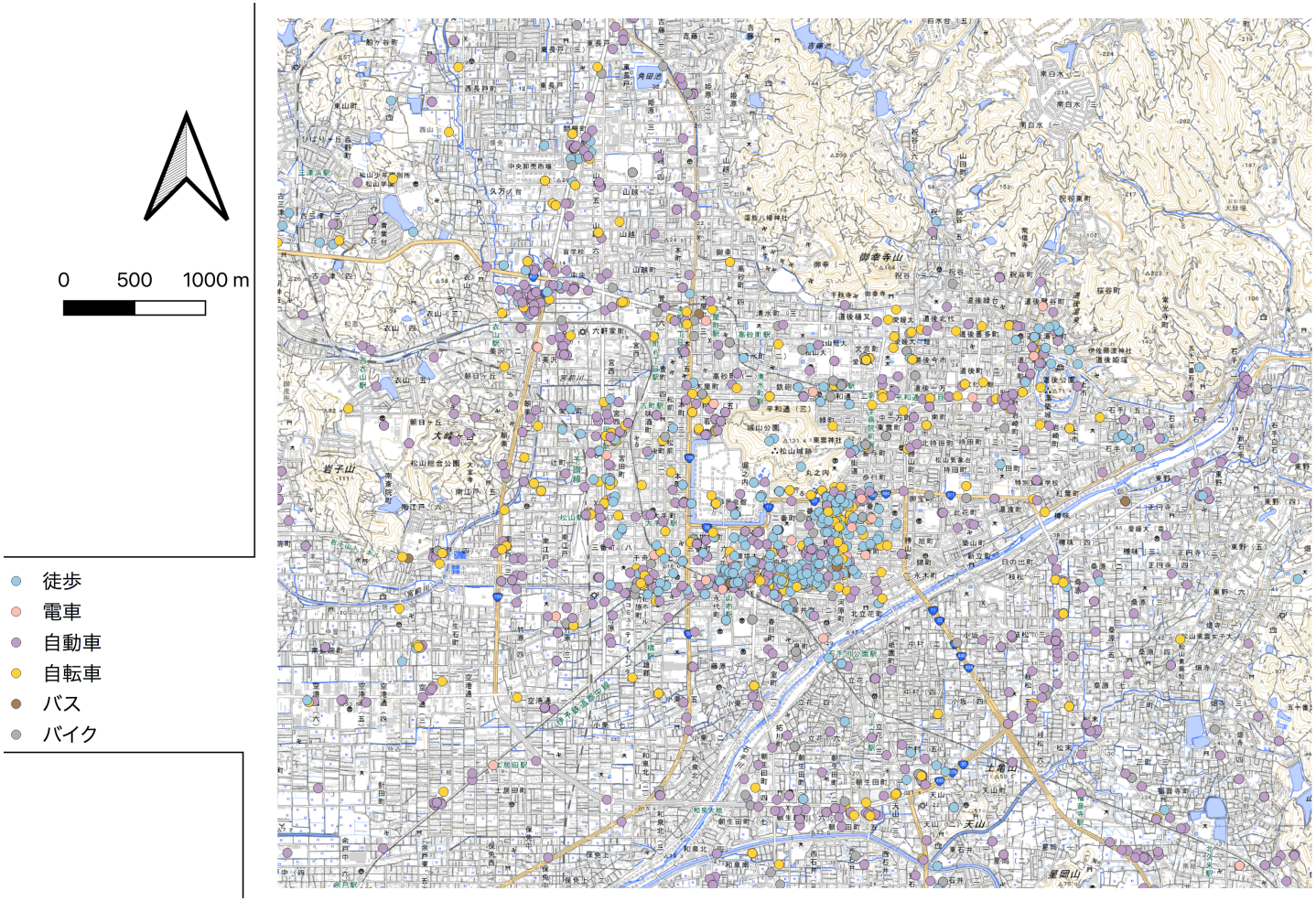
鉄道の傾向③　ODを考える

- 鉄道、バス利用のトリップのODを考える。



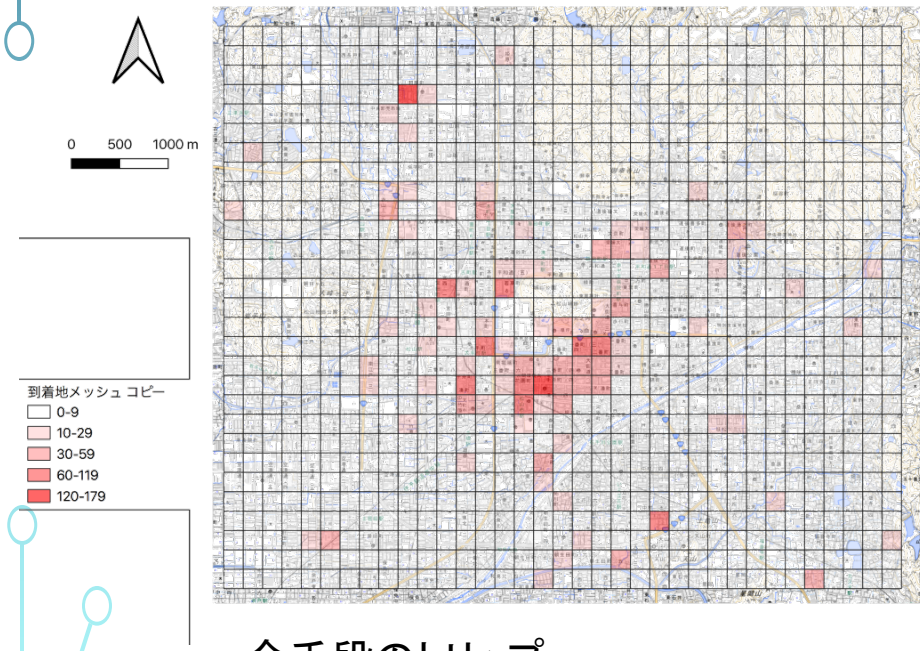
鉄道の傾向④　目的地を考える

- 「帰宅」目的以外の到着地の分布を見る

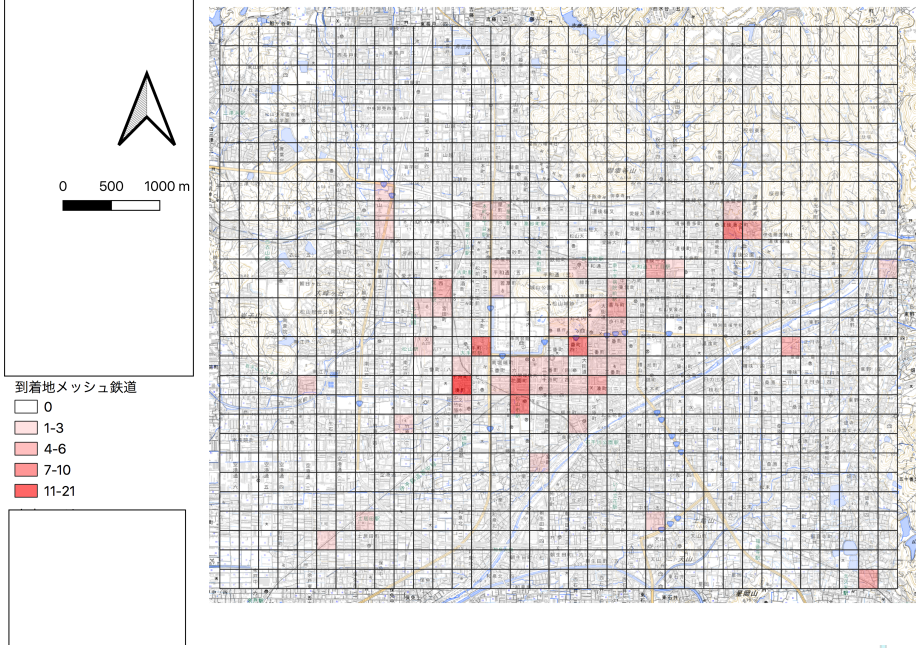


鉄道の傾向④　目的地を考える

- 「帰宅」目的以外の到着地の分布を見る(200mメッシュ)



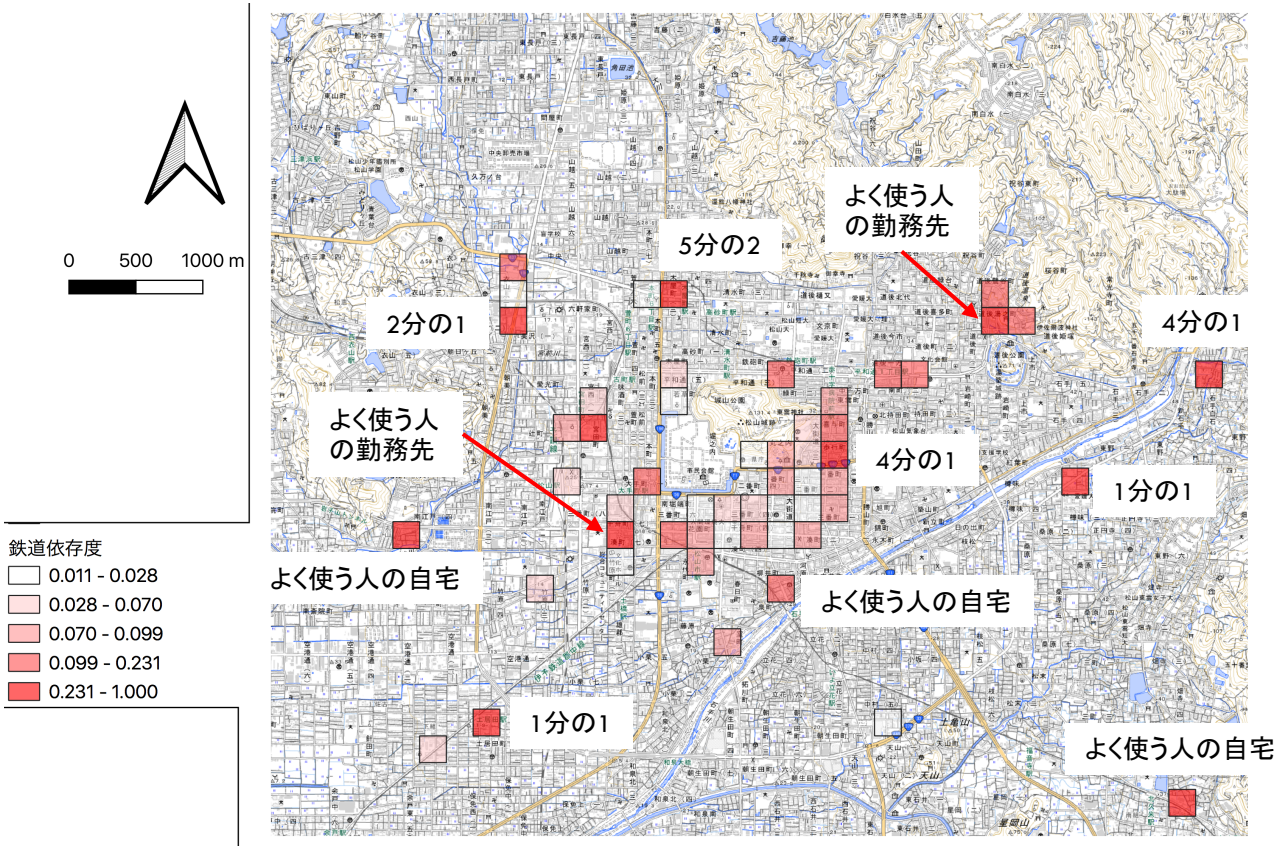
全手段のトリップ
・市駅周辺から大街道にかけてがやはり多い
・外にも、道後、合同庁舎、愛媛大フジグラン、パルティフジ、伊予銀など多い



鉄道・バスのトリップ
・全体的な傾向は同じ
・伊予銀に鉄道利用者はいない
・SPCで多くなっているのは出勤
・愛媛大は少ない

鉄道の傾向④　目的地を考える

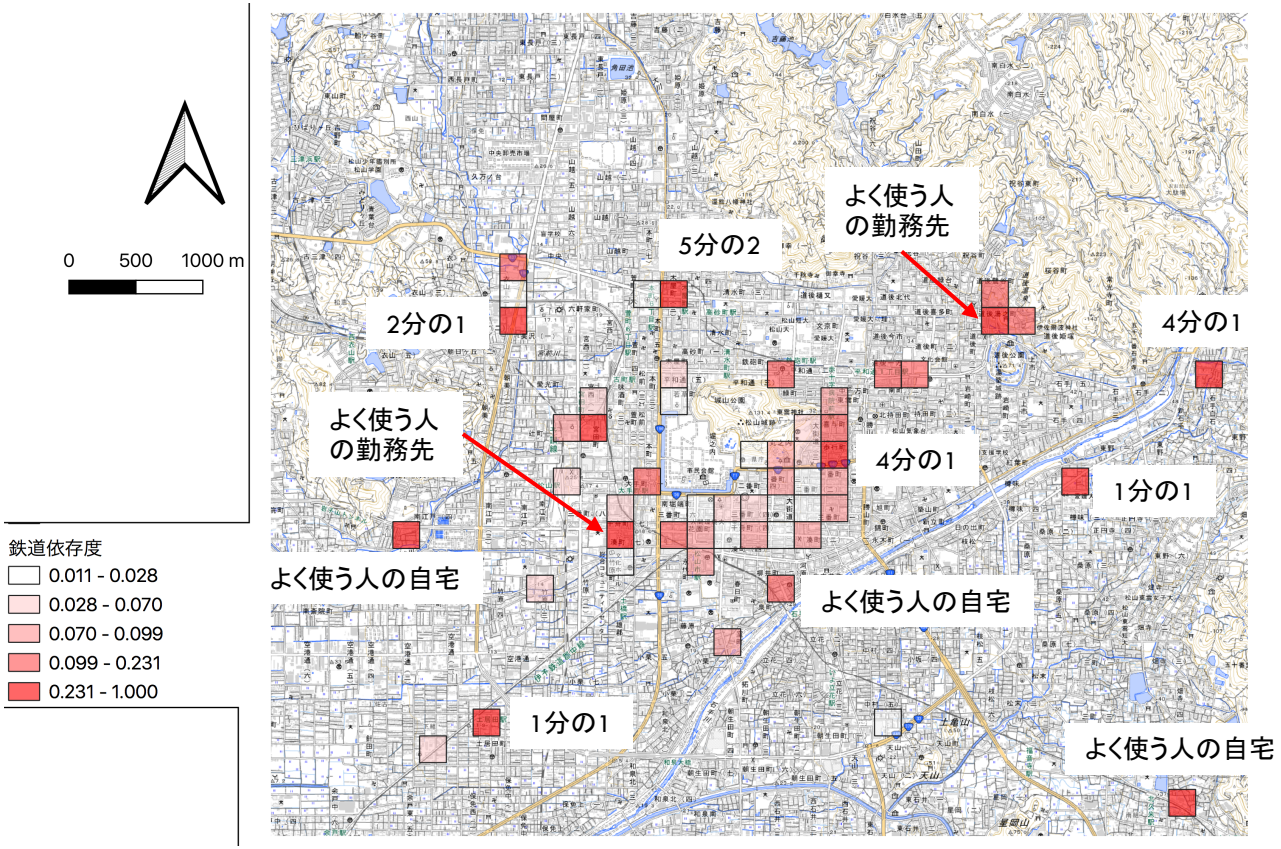
- 鉄道・バスの利用があった場所の依存率



図上に注釈したように、依存率が一見高く見えるところは、そもそも
データ数が少ない場所か、よく使う人がよく行く場所
→分析する上で信頼できるデータでは無い。

鉄道の傾向④　目的地を考える

- あえて傾向を挙げるなら…



松山駅前、松山市、大街道、道後温泉駅から近いところの割合が高い？

鉄道の傾向④　目的地を考える

- 松山駅前、松山市、大街道、道後温泉駅から近いところ
  ＝駅からあまり歩かない場所では鉄道の利用率が高い？

検証

出発地or到着地が上記4駅からXメートル以下の場合における、鉄道の利用率を考える

X=500のとき0.04635539
X=360のとき0.04964222
X=160のとき0.05988858(最大)
(参考)距離制限がない(X=∞)とき、0.02897244

(参考)徒歩の割合が最大となる距離X(メートル)
X=520のとき、割合0.409468

3.モデルの概要設計

MNLを用いたパラメータ推定

モデルの概要設計①　手順

- 交通手段に関する離散選択モデル
- MNLロジットモデルによる推定

MNLモデル(選択肢iを選択する確率P_i 誤差項にガンベル分布を仮定)

- $P_i = \exp(V_i) / \sum_{j=1}^{J} exp(V_j)$　　V_iは効用関数の確定項

手順

- 効用関数の確定項の形を決定
- 選択確率の式から対数尤度関数を計算
- これの最大化問題を解く(最尤推定法)
- 推定結果が信頼できるかどうかをチェック

モデルの概要設計②　確定項の設計

確定項

説明変数

- $V_i = \alpha_i + \beta' x_i = \alpha_i + (\beta_1 \quad \cdots \quad \beta_l) \begin{pmatrix} x_{i;1} \\ \vdots \\ x_{i;l} \end{pmatrix}$

固有定数項
個人や説明変数に依存せず、選択肢そのものが持つ特性

パラメータ
説明変数にかかる項
線形の関数を仮定し
全選択肢に共通
今回これを推定する

これまでの考察により、

- 自動車　$V_{自動車} = \alpha_{自動車} + \beta_{時間} x_{自動車時間}$
- バイク　$V_{バイク} = \alpha_{バイク} + \beta_{時間} x_{バイク時間}$
- 自転車　$V_{自転車} = \alpha_{自転車} + \beta_{時間} x_{自転車時間}$
- 徒歩　$V_{徒歩} = \beta_{時間} x_{徒歩時間} + \beta_{徒歩距離} (\text{if } x_{距離} < X_1)$
- 鉄道　$V_{鉄道} = \alpha_{鉄道} + \beta_{駅までの距離} (\text{if } x_{主要駅までの距離} < X_2)$

モデルの概要設計③　時間を求める

- 説明変数$x_{自動車時間}$, $x_{バイク時間}$, $x_{自転車時間}$, $x_{徒歩時間}$を求めたい。
- しかしPPデータには距離と実際にかかった時間しかない
  →各手段の平均速度を求め、距離をそれで割り時間とする。
- ただし、PPデータには欠損や測定誤差から明らかにおかしなデータも含まれており、それらを抜く必要がある。
- 徒歩　距離10m~4500m, 時間5s~7200sのものを平均
  秒速1.183m(時速4.258km)
- 自転車　距離20m~, 時間30s~7200s,時速~50kmのものを平均
  秒速1.931m(時速6.952km)
- バイク　距離20m~, 時間30s~7200s,時速~50kmのものを平均
  秒速3.226m(時速11.614km)

モデルの概要設計③　時間を求める

- 説明変数$x_{自動車時間}$, $x_{バイク時間}$, $x_{自転車時間}$,$x_{徒歩時間}$を求めたい。
- しかしPPデータには実際にかかった時間しかない
  →各手段の平均速度を求める。
- ただし、PPデータには欠損や測定誤差から明らかにおかしなデータも含まれており、それらを抜く必要がある。

- 自動車　距離100m~, 時間60s~10800sのもので、
  (10分で1km以下or 60分で4km以下or 10分で40km以上or 5分で8km以上のもの)を除外したものを平均
  秒速4.3208m(時速15.555km)

4.推定結果

3で作った確定項を用いてパラメータを推定する

推定結果①　$X_1=520, X_2=160$

説明変数	パラメータ	t値
$\alpha_{自動車}$	1.460	26.381**
$\alpha_{バイク}$	-0.052	-0.828
$\alpha_{自転車}$	0.614	11.061**
$\alpha_{鉄道}$	-2.571	-18.156**
$\beta_{時間[s/600]}$	-0.132	-14.699**
$\beta_{徒歩距離}$	2.256	29.814**
$\beta_{駅までの距離}$	2.247	13.052**
サンプル数	7067	
初期尤度	-11246.75	
最終尤度	-8080.418	
尤度比	0.2815	
修正尤度比	0.2809	

*5%有意　**1%有意

- バイクの固有定数項以外で有意な値が得られた
- ただし、値が収束していない(convergence=1)

推定結果②　Xの値を変えてみる

X_1	520		290		450		520	
X_2	160		290		450		520	
説明変数	パラメータ	t値	パラメータ	t値	パラメータ	t値	パラメータ	t値
$\alpha_{自動車}$	1.460	26.381**	0.787	17.352**	1.264	24.766**	1.434	26.879**
$\alpha_{バイク}$	-0.052	-0.828	-0.730	-13.338**	-0.238	-4.024**	-0.110	-1.795
$\alpha_{自転車}$	0.614	11.061**	0.049	1.069	0.454	8.782**	0.536	9.916**
$\alpha_{鉄道}$	-2.571	-18.156**	-3.502	-23.094**	-2.493	-18.304**	-2.458	-16.665**
$\beta_{時間[s/600]}$	-0.132	-14.699**	-0.202	-20.142**	-0.138	-15.774**	-0.111	-13.157**
$\beta_{徒歩距離}$	2.256	29.814**	1.936	21.858**	2.202	28.783**	2.385	31.818**
$\beta_{駅までの距離}$	2.247	13.052**	1.987	11.374**	0.892	5.417**	0.739	4.068**
サンプル数	7067		7067		7067		7067	
初期尤度	-11246.75		-11246.75		-11246.75		-11246.75	
最終尤度	-8080.418		-8311.373		-8157.956		-8080.418	
尤度比	0.2815		0.2610		0.2746		0.2784	
修正尤度比	0.2809		0.2604		0.2740		0.2778	

Nelder-Mead 法　　*5%有意　**1%有意

- Xの値を変えて試したが、X_1=520, X_2=160が一番良さそうな値となる
- 次に、推定法を変えて試す

推定結果③　推定法を変えてみる

推定方法	Nelder-Mead 法		SANN法		CG法		BFGS法	
説明変数	パラメータ	t値	パラメータ	t値	パラメータ	t値	パラメータ	t値
$\alpha_{自動車}$	1.460	26.381**	1.835	30.731**	1.836	30.675**	1.833	30.481**
$\alpha_{バイク}$	-0.052	-0.828	0.328	4.929**	0.310	4.638**	0.315	4.700**
$\alpha_{自転車}$	0.614	11.061**	0.922	15.245**	0.941	15.553**	0.950	15.664**
$\alpha_{鉄道}$	-2.571	-18.156**	-1.602	-13.683**	-1.594	-13.639**	-1.709	-14.073**
$\beta_{時間[s/600]}$	-0.132	-14.699**	-0.093	-11.613**	-0.095	-11.822**	-0.098	-11.972**
$\beta_{徒歩距離}$	2.256	29.814**	2.641	33.317**	2.620	33.045**	2.631	33.101**
$\beta_{駅までの距離}$	2.247	13.052**	1.649	10.557**	1.447	8.802**	1.752	11.035**
サンプル数	7067		7067		7067		7067	
初期尤度	-11246.75		-11246.75		-11246.75		-11246.75	
最終尤度	-8080.418		-8052.77		-8053.56		-8051.717	
尤度比	0.2815		0.2840		0.2839		0.2841	
修正尤度比	0.2809		0.2834		0.2833		0.2835	

X_1=520, X_2=160　　*5%有意　**1%有意

- SANN法とBFGS法で収束(convergence=0)した
- 尤度比の最も大きいBFGS法を採用

5.結果の分析

推定の結果を分析する

結果の分析

- ほとんど基礎集計で見たような
  鉄道<徒歩<バイク<自転車<自動車
  という効用関数固定値の順位
- 距離が520m以下であるとき、効用の固定値だけで自動車を超える。その超過分は時間にして(2.631-1.833)/0.098*10=81分もの効用に匹敵する。
- 主要駅から160mの距離に出発地または目的地がある場合、その分の効用がある。しかしこれは徒歩の効用に追いつく程度であって、自動車の効用には到底追いつけない。

推定方法	BFGS法	
説明変数	パラメータ	t値
$\alpha_{自動車}$	1.833	30.481**
$\alpha_{バイク}$	0.315	4.700**
$\alpha_{自転車}$	0.950	15.664**
$\alpha_{鉄道}$	-1.709	-14.073**
$\beta_{時間[s/600]}$	-0.098	-11.972**
$\beta_{徒歩距離}$	2.631	33.101**
$\beta_{駅までの距離}$	1.752	11.035**
サンプル数	7067	
初期尤度	-11246.75	
最終尤度	-8051.717	
尤度比	0.2841	
修正尤度比	0.2835	
X_1=520, X_2=160	*5%有意 **1%有意	

6.確定項の一般化

変数を増やし効用関数を一般化

確定項の一般化

今回のパラメータ推定は、実際に利用していない手段の所要時間が分からなかったため、実際には以下のように距離から所要時間を推定する方法をとった。

- 自動車　$V_{自動車} = \alpha_{自動車} + \beta_{時間} x_{距離} / 4.3208$
- バイク　$V_{バイク} = \alpha_{バイク} + \beta_{時間} x_{距離} / 3.226$
- 自転車　$V_{自転車} = \alpha_{自転車} + \beta_{時間} x_{距離} / 1.931$
- 徒歩　$V_{徒歩} = \beta_{時間} x_{距離} / 1.183 + \beta_{徒歩距離} (\text{if } x_{距離} < X_1)$
- 鉄道　$V_{鉄道} = \alpha_{鉄道} + \beta_{駅までの距離} (\text{if } x_{主要駅までの距離} < X_2)$

確定項の一般化

それぞれの手段の速さもパラメータであると考えると、次のように一般化することができる。

ただし、パラメータの数は3個増え10個になっている。

- 自動車　$V_{自動車} = \alpha_{自動車} + \beta_{自動車} x_{距離}$
- バイク　$V_{バイク} = \alpha_{バイク} + \beta_{バイク} x_{距離}$
- 自転車　$V_{自転車} = \alpha_{自転車} + \beta_{自転車} x_{距離}$
- 徒歩　$V_{徒歩} = \beta_{徒歩} x_{距離} + \beta_{徒歩距離} (\text{if } x_{距離} < X_1)$
- 鉄道　$V_{鉄道} = \alpha_{鉄道} + \beta_{駅までの距離} (\text{if } x_{主要駅までの距離} < X_2)$

確定項の一般化

前ページの効用関数でパラメータ推定を行なった(右表)

パラメータを増やしただけあって、尤度比、最終尤度は上がった。

もともと

$\beta_{自動車[km]}=\beta_{時間価値}/速さ$

だったこともあり、時間価値が固定だとすると、各交通手段の平均的な速さを出すことができる？

推定方法	BFGS法	
説明変数	パラメータ	t値
$\alpha_{自動車}$	1.832	22.444**
$\alpha_{バイク}$	0.674	7.147**
$\alpha_{自転車}$	1.903	20.791**
$\alpha_{鉄道}$	-1.455	-12.007**
$\beta_{自動車[km]}$	-0.0113	-4.278**
$\beta_{バイク[km]}$	-0.127	-8.726**
$\beta_{自転車[km]}$	-0.469	-20.599**
$\beta_{徒歩[km]}$	-0.105	-5.758**
$\beta_{徒歩距離}$	2.950	31.003**
$\beta_{駅までの距離}$	1.658	10.710**
サンプル数	7067	
初期尤度	-11246.75	
最終尤度	-7721.228	
尤度比	0.3135	
修正尤度比	0.3126	
X_1=520, X_2=160	*5%有意 **1%有意	

確定項の一般化

徒歩	4.2000km/h
バイク	3.4961km/h
自動車	39.1023km/h
自転車	0.9431km/h

今度は、自動車は良さそうな速さをしているが、バイクや自転車が遅すぎる。

→ $\beta_{バイク}$, $\beta_{自転車}$が小さすぎる(絶対値が大きすぎる)

つまり…バイクや自転車には「時間がかかること」以上の効用の低減がある。(疲れるなど)

推定方法	BFGS法	
説明変数	パラメータ	t値
$\alpha_{自動車}$	1.832	22.444**
$\alpha_{バイク}$	0.674	7.147**
$\alpha_{自転車}$	1.903	20.791**
$\alpha_{鉄道}$	-1.455	-12.007**
$\beta_{自動車}$[km]	-0.0113	-4.278**
$\beta_{バイク}$[km]	-0.127	-8.726**
$\beta_{自転車}$[km]	-0.469	-20.599**
$\beta_{徒歩}$[km]	-0.105	-5.758**
$\beta_{徒歩距離}$	2.950	31.003**
$\beta_{駅までの距離}$	1.658	10.710**
サンプル数	7067	
初期尤度	-11246.75	
最終尤度	-7721.228	
尤度比	0.3135	
修正尤度比	0.3126	
X_1=520, X_2=160	*5%有意 **1%有意	

7.政策シミュレーション

6で作成したパラメータを使い、政策のシミュレーションを行う

政策シミュレーション

自動車、自転車、徒歩は政策で弄りにくい。

また、政策的にも公共交通機関の利用を増やしたい。

→鉄道の利用を増やしたい。

鉄道駅から160m以内の場所では鉄道の効用が多少上がる

よって

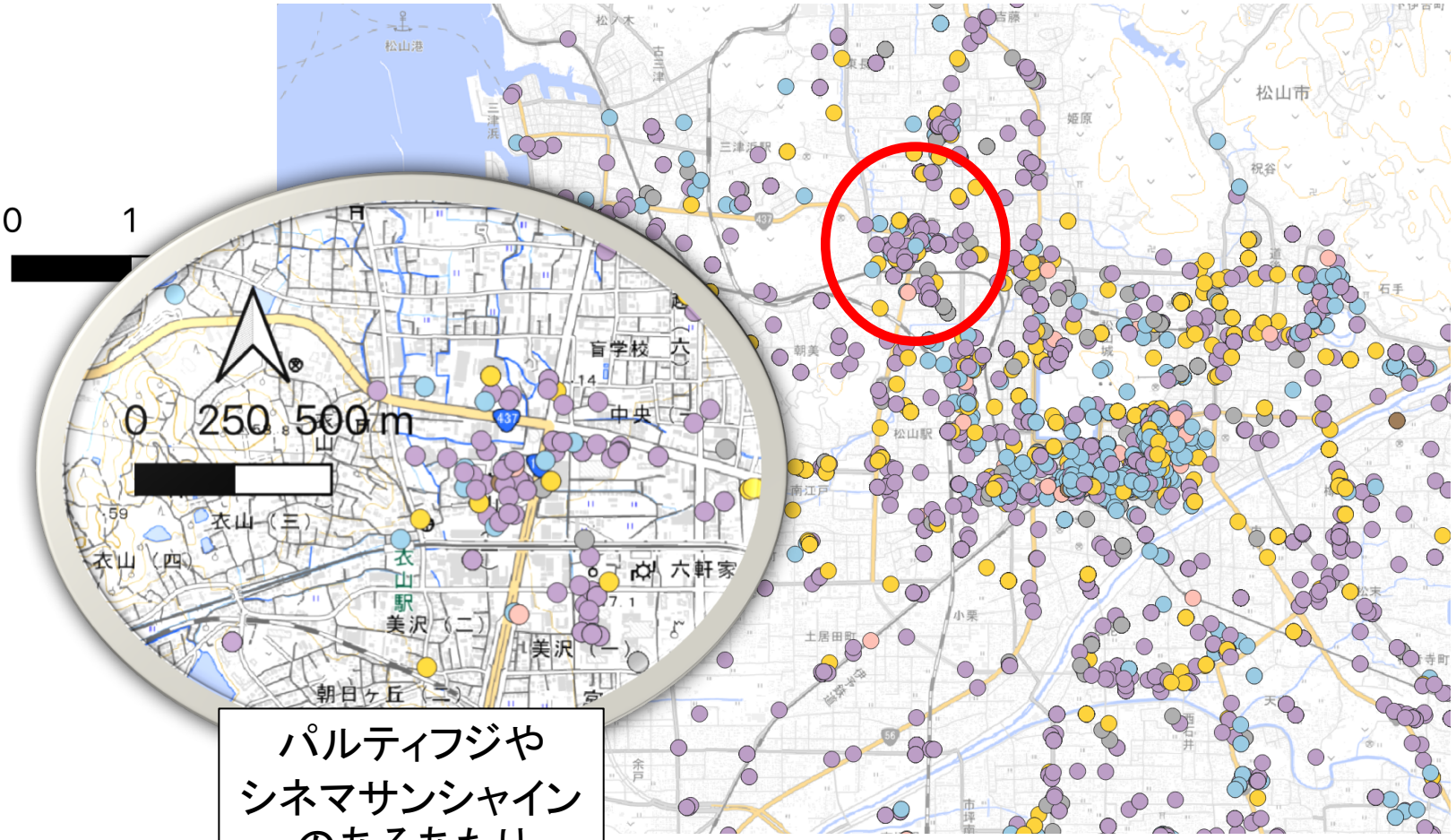
人が密集する場所の160m以内に駅を作ると、利用者が増える

説明変数	パラメータ
$\alpha_{自動車}$	1.832
$\alpha_{バイク}$	0.674
$\alpha_{自転車}$	1.903
$\alpha_{鉄道}$	-1.455
$\beta_{自動車}$[km]	-0.0113
$\beta_{バイク}$[km]	-0.127
$\beta_{自転車}$[km]	-0.469
$\beta_{徒歩}$[km]	-0.105
$\beta_{徒歩距離}$	2.950
$\beta_{駅までの距離}$	1.658

X_1=520, X_2=160

政策シミュレーション

人が密集する場所の160m以内に駅を作り、利用者を増やしたい

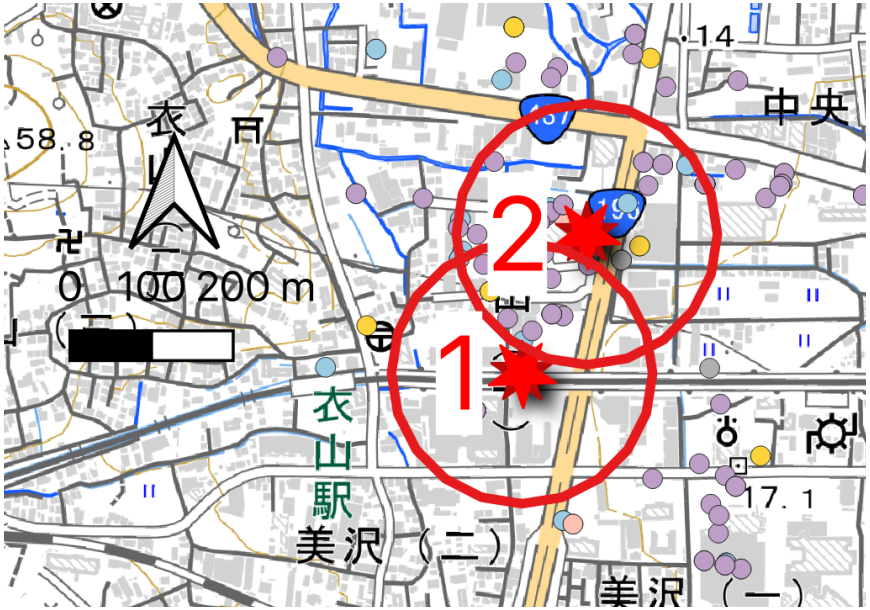


政策シミュレーション

次の2パターンを考える。

①衣山駅を200m東に移動させ、シネマサンシャインの近くに置く。 起or終点が半径160m円に入るトリップ：92トリップ

②市内電車を延伸させ、2の位置に新たな駅を作る
起or終点が半径160m円に入るトリップ：132トリップ



政策シミュレーション

先ほどのパラメータを用いて、確率的にいくつのトリップが鉄道利用になるのかを期待値で考える。

(トリップ数)=Σ(各トリップの選択確率)

新駅の位置	1	2
半径160m圏内発or着のトリップ数	92	132
実際の鉄道トリップ数	1	2
右のパラメータを置いた時、現在のトリップ数期待値	1.875	2.735
新駅を設置した時のトリップ数期待値	9.006	12.875
域内利用者の鉄道利用者数	4.80倍	4.71倍

説明変数	パラメータ
$\alpha_{自動車}$	1.832
$\alpha_{バイク}$	0.674
$\alpha_{自転車}$	1.903
$\alpha_{鉄道}$	-1.455
$\beta_{自動車}$[km]	-0.0113
$\beta_{バイク}$[km]	-0.127
$\beta_{自転車}$[km]	-0.469
$\beta_{徒歩}$[km]	-0.105
$\beta_{徒歩距離}$	2.950
$\beta_{駅までの距離}$	1.658

X_1=520, X_2=160

8.発展的考察、今後の課題

発展的考察、今後の課題

- 目的と代表交通手段にある程度の相関が感じられたが、確定項中に入れることができなかった。娯楽における鉄道の利用は少ないため、前ページの結果ほど利用者が増えない可能性はある。

- 実際には鉄道にも運賃や時間、運行間隔などにより、確定項の説明変数は変わりうる。

- 鉄道やバスのデータを分析するのにデータが少ないと感じた。あまり人数が少ないと、たまたま調査した人の自宅が重要な位置だと認識される可能性がある。分析に必要なデータ数はいくつだろうか。

ご静聴ありがとうございました