行動モデルの基礎

M1　石河 万衣

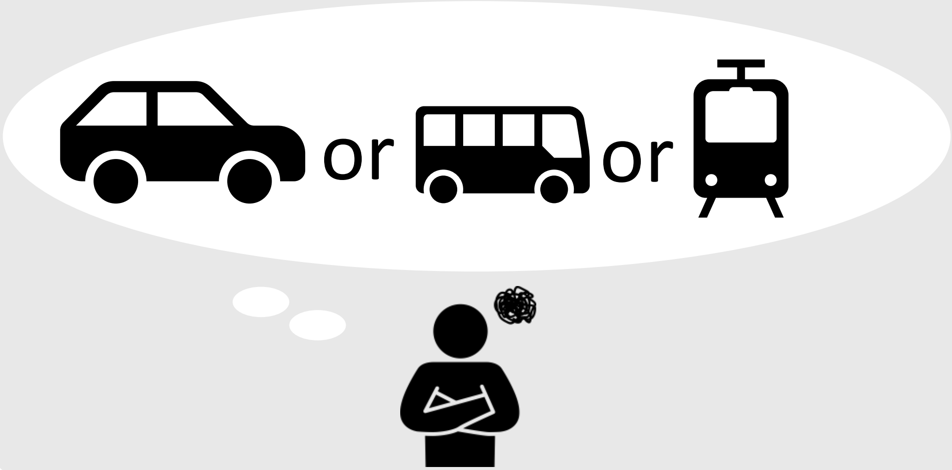
目次

行動モデルとは

人の意思決定やメカニズムを数式で記述するモデルの総称

例）

交通手段選択



出発時刻選択

何に使う？
都市や交通の政策評価やシミュレーションなど
・需要予測（運賃を変えると利用がどの程度変化するか）等
・施設配置（観光案内所をどこに置くか）
・道路要領配分（渋滞が起きない程度に車道を減らしたい）

効用の記述とランダム性の導入

意思決定のルール

意思決定者の効用が最大となる選択肢が選ばれる

効用：旅行時間・費用・快適さなどから決まる満足度

効用関数の立式

- パラメータβは説明変数が効用に与える影響の大きさを示す
- 分析者がすべての構成を把握することは原理的に困難→誤差項の導入

$$U_a = \underbrace{V_{car}}_{\text{確定項}} + \underbrace{\epsilon_{car}}_{\text{誤差項}} = \underbrace{\beta_{time}}_{\text{パラメータ}} x_{time,car} + \beta_{cost} x_{cost,car} + \epsilon_{car}$$

行動モデルの目的：効用関数を用いて個人の選択行動を記述する
→各パラメータβを推定する

誤差項の独立同一性の過程（i.i.d）

$$U_{car} = V_{car} + \epsilon_{car}$$

独立性

ある選択肢に関する未観測要因がほかの選択肢の未観測要因と相関を持たない

同一分布

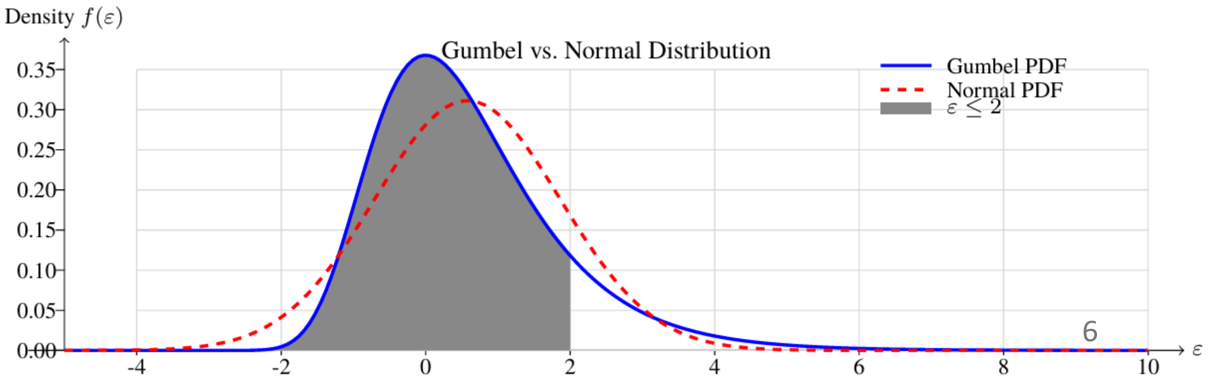
未観測要因のばらつき方がすべての選択肢・個人で同じである

選択確率の計算

選択確率は誤差項εの確率分布に依存する！

$$
\begin{aligned}
\Pr(y_k = a) &= \Pr(V_a + \epsilon_a > V_a' + \epsilon_a' \ \ \forall a' \neq a) \\
&= \Pr(V_a - V_a' > \epsilon_a' - \epsilon_a \ \ \forall a' \neq a) \\
&= Pr_\epsilon\left(V_a - V_a' > \epsilon \ \forall a' \neq a\right) \\
&= \int_{-\infty}^{V_a - V_a'} f(\epsilon) d\epsilon \\
&= F_\epsilon(V_a - V_a')
\end{aligned}
$$

※教科書より



選択確率の計算

誤差項の分布

- 正規分布の場合： $P(car) = \int_{-\infty}^{\infty} d\epsilon_{car} \int_{-\infty}^{V_{car}-V_{bus}} d\epsilon_{bus} \int_{-\infty}^{V_{car}-V_{train}} d\epsilon_{train} \varphi(\epsilon_{walk}, \epsilon_{car}, \epsilon_{train})$

  プロビットモデル

  - 多項選択の場合は閉形で簡潔に表せない（初等関数で書けない）→計算コストが大きい

- i.i.d　ガンベル分布の場合： $P(y_k = a) = \dfrac{\exp(V_a)}{\sum_a' \exp(V_a')}$

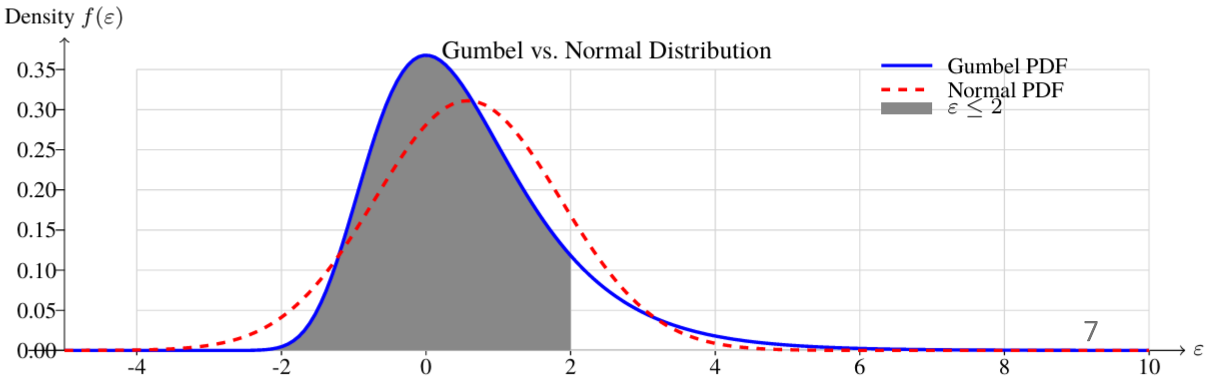
ガンベル分布の累積分布関数：

$$F(x) = \exp(-\exp(-\mu(x-\eta)))$$

※通常η=0, μ=1に正規化する

- 正規分布と形が似ている
- 閉形で表せる

※教科書より



パラメータの推定方法

行動モデルの目的：効用関数を用いて個人の選択行動を記述する
→各パラメータβを推定する→実際の行動データから推定する

$$U_a = V_{car} + \epsilon_{car} = \beta_{time} x_{time,car} + \beta_{cost} x_{cost,car} + \epsilon_{car}$$

条件

	鉄道	バス	車
時間	1h	2h	0.5h
料金	1500円	1000円	700円

実データ

選択肢	鉄道	バス	車
選択人数	40	20	40
選択割合	0.4	0.2	0.4

できるだけ近くなるようなβを見つける

$U_a = V_{car} + \epsilon_{car} = \beta_{time} x_{time,car} + \beta_{cost} x_{cost,car} + \epsilon_{car}$

モデルの出力（パラメータβを用いて算出した確率）

選択肢	鉄道	バス	車
選択人数			
選択割合			

最尤推定法

最尤推定：モデルが実データを再現する確率が最も高くなるようにパラメータを設定

モデルによる個人kの選択肢ykの選択確率を$P(y_k; x_k)$とおく（ x_k は説明変数）
全kについて観測と同じ結果を再現する確率は、

$$\prod_{k=1}^{M} P(y_k; x_k) = \prod_{k=1}^{M} \frac{\exp(V_{yk})}{\sum_{a=1}^{N} \exp(V_a)}$$

これをパラメータβの関数とみて（＝尤度関数）
この尤度関数L(β)を最大化するパラメータβを求める

$$L(\beta) \coloneqq \prod_{k=1}^{M} P(y_k; x_k) = \prod_{k=1}^{M} \frac{\exp(V_{yk})}{\sum_{a=1}^{N} \exp(V_a)}$$

L(β)は確率の積であるため、値が小さくなりすぎる→対数尤度を最大化する問題に

$$LL(\beta) \coloneqq \log L(\beta) = \sum_{k=1}^{M} \log P(y_k; x_k)$$

選択肢・個人間の相関・異質性の扱い

i.i.d過程

独立性
ある選択肢に関する未観測要因がほかの選択肢の未観測要因と相関を持たない

同一分布
未観測要因のばらつき方がすべての選択肢・個人で同じである

⬇ しかし実際は…

- 互いに類似した交通手段や重複区間などで未観測要因は似通う
- 時間価値や混雑回避傾向は個人によって異なる
  →個人間で同一の分布を仮定するのは微妙

選択肢・個人間の相関・異質性の扱い

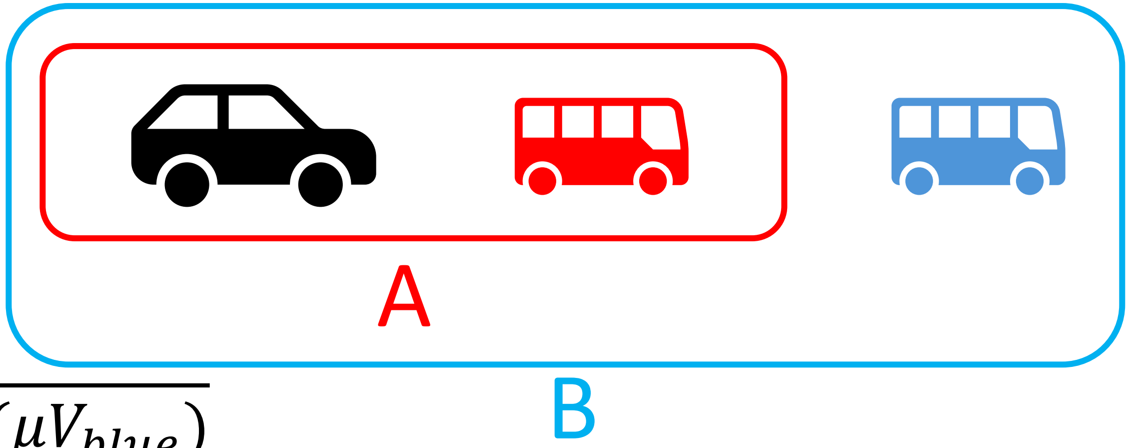
MNLモデルでは、選択確率の比はその他の選択肢の影響を受けない
（赤バス・青バス問題）

選択肢がAのとき

$$P(car|A) = \frac{\exp(\mu V_{car})}{\exp(\mu V_{car}) + \exp(\mu Vred)}$$

選択肢がBのとき

$$P(car|B) = \frac{\exp(\mu V_{car})}{\exp(\mu V_{car}) + \exp(\mu Vred) + \exp(\mu V_{blue})}$$



$$\frac{P(red|A)}{P(car|A)} = \frac{\exp(\mu V_{red})}{\exp(\mu V_{car})} = \frac{P(red|B)}{P(car|B)}$$

車と赤バスの選択確率の比が
青バスの有無によらず不変

赤バス・青バスは車と比べて公共交通という類似した選択肢
→両者の誤差項は相関している

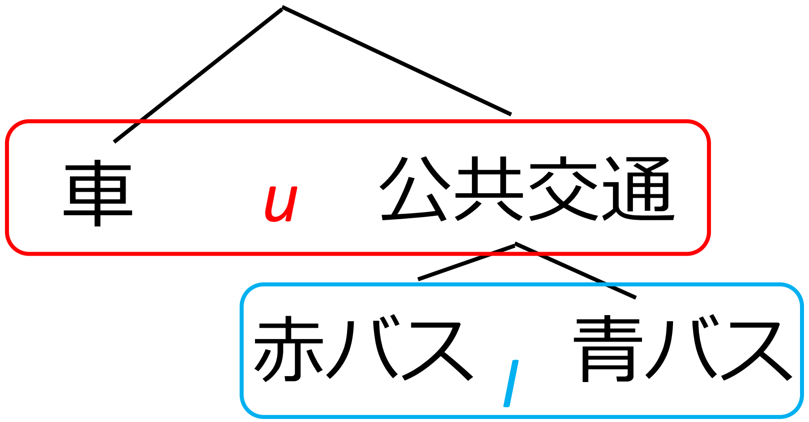
i.i.dを仮定したMNLモデルの限界！

NL（Nested Logit）Model

誤差項が互いに相関関係にある選択肢同士をネスト（グループ）にまとめる

⬇

一部の選択肢相関に相関がある場合でも多項選択を記述可能にしている



選択肢(u,l)の効用：$U_{ul} = V_u + V_l + V_{ul} + \epsilon_u + \epsilon_{ul}$

V_u, V_l :選択肢u,lに固有の確定項
V_{ul}:選択肢u,lの組み合わせで決まる確定項
ϵ_u:選択肢uの固有の誤差項
ϵ_{ul}:u,lの組み合わせで決まる効用の誤差項
（スケールパラメータμをもつガンベル分布に従う）

二つのバス間の誤差項の内相関する部分がϵ_uに吸収され、それぞれのバスに固有の誤差項がϵ_{ul}で表される

NL（Nested Logit）Model

選択確率の計算

$$P(u,l) = P(l \mid u) \cdot P(u)$$

$$P(l \mid u) = \frac{\exp\big(\mu(V_l + V_{ul})\big)}{\sum_{l'} \exp\big(\mu(V_{l'} + V_{ul'})\big)}$$

$$P(u) = \frac{\exp\big(\mu^u(V_u + V'_u)\big)}{\sum_{u'} \exp\big(\mu^u(V_{u'} + V'_{u'})\big)} \qquad V'_u \equiv \frac{1}{\mu} \ln \sum_l \exp\big(\mu(V_l + V_{ul})\big)$$

- どちらも閉形式で計算可能
- スケールパラメータは1つ固定する必要がある
  （最下層のスケールパラメータを1に固定することが多い）
- スケールパラメータは上位ネストのものが小さくなる
  μ＞μ^{μ}となる（このようにならなかった場合、ネスト構造の過程が成り立たなくなる）

$$\mu^u / \mu = 1$$

上位・下位の誤差相関に相関がなくなり、MNLに帰着

MXL（Mixed Logit）Model

MNLでは、個人間の選好の違い（異質性）を表現できない

性別や年齢層などのダミー変数を説明変数に導入
→ダミー変数で区分されたグループ単位での平均的な違いを
　とらえるにとどまる

⬇

Mixed Logit Model

MXL（Mixed Logit）Model

効用関数のパラメータを確率分布に従うランダムな係数として扱う

$$\beta_k \sim N(\bar{\beta}, \Omega)$$

$$P_{ka}(\beta_k) = \frac{\exp(V_{ka})}{\sum_j \exp(V_{kj})}$$

実際にβは観測されてないので

$$P_{ka} = \int \frac{\exp(V_{ka})}{\sum_j \exp(V_{kj})} f(\beta \mid \theta)\, d\beta$$

これは積分できないので、
f(β|Θ)からR回乱数を抽出して平均

$$P_{ka}^{(r)} = \frac{\exp\left(V_{ka}^{(r)}\right)}{\sum_j \exp\left(V_{kj}^{(r)}\right)} \qquad \hat{P}_{ka} = \frac{1}{R}\sum_{r=1}^{R} P_{ka}^{(r)}$$

シミュレーション対数尤度
f(β|Θ)の母数を推定できる

$$\ln SL = \sum_{k=1}^{K}\sum_{a=1}^{A} \delta_{ka} \ln\left(\frac{1}{R}\sum_{r=1}^{R} P_{ka}^{(r)}\right)$$

推定の信頼性

モデルが回った！
→本当にいいモデル？

t検定:
パラメータβが0であることを帰無仮説とし、t分布より帰無仮説が棄却できれば
パラメータが統計的に0とは言い切れない

尤度比検定:　　※2025スタートアップゼミ資料より

ゴール：モデルがどれだけデータにフィットしているかを調べる

McFaddenの擬似決定係数

尤度比 (likelihood ratio)

$$\rho^2 = 1 - \frac{LL(\hat{\beta})}{LL(0)}$$

- 推定パラメーターが尤度を全く改善していないとき

$L(\hat{\beta}) = L(0)$　$\rightarrow \rho = 0$［下限］

- 推定パラメーターがサンプルを完璧に予測しているとき

$L(\hat{\beta}) = 1$　$\rightarrow \rho = 1$［上限］

$L(\hat{\beta})$　最大尤度/最終尤度

$L(0)$　無情報モデル(Null Model)尤度/初期尤度 全てのパラメーターが0であるときの対数の値

欠点：説明変数を増やせば必ず増加する（← $L(\hat{\beta})$が大きくなる）

自由度調整済み尤度比 (adjusted-likelihood ratio)

$$\rho^2 = 1 - \frac{LL(\hat{\beta}) - h}{LL(0)}$$

h: パラメーター数

0.3程度あればいい，とかいわれたりますがcase by case

2025.04.17

課題：2023松山都市圏PT調査データを用いて構築した行動モデルの考察

１．説明変数の対数変換
２．政策シミュレーション
３．NLとの比較

目次

1．概要
2．使用したデータ
3．基礎集計
4．MNLの構築（説明変数の対数変換）
5．政策シミュレーション
6．NLの構築

用いたデータの概要

愛媛県松山市とその近隣自治体からなる
都市圏における令和5年度実施のPT調査の結果
を整形したもの。

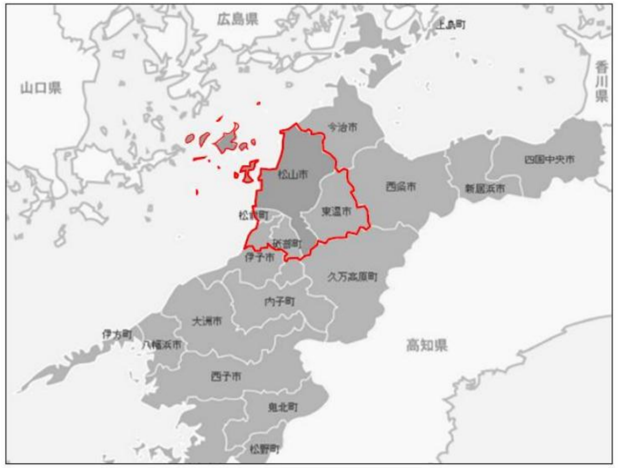
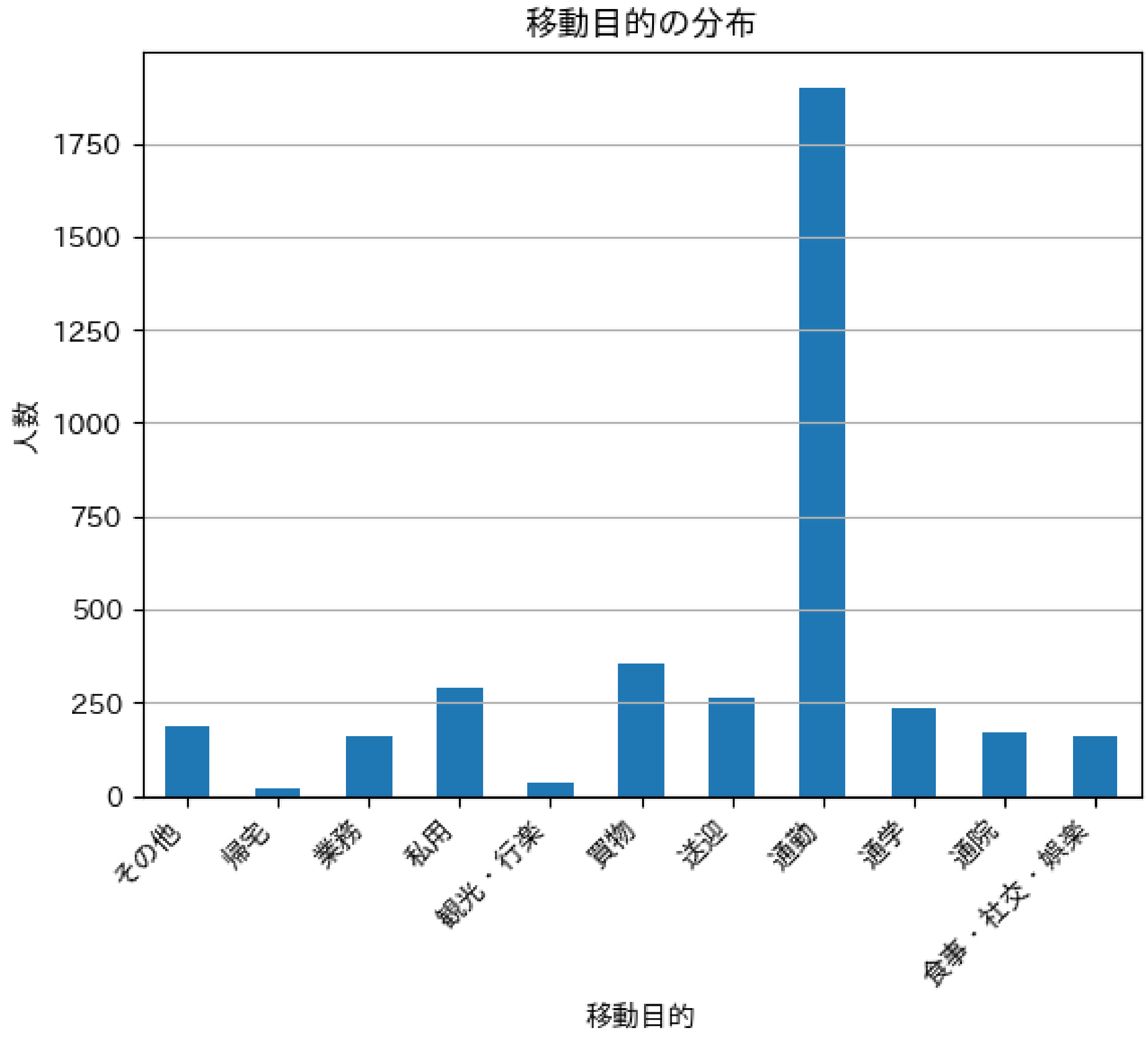


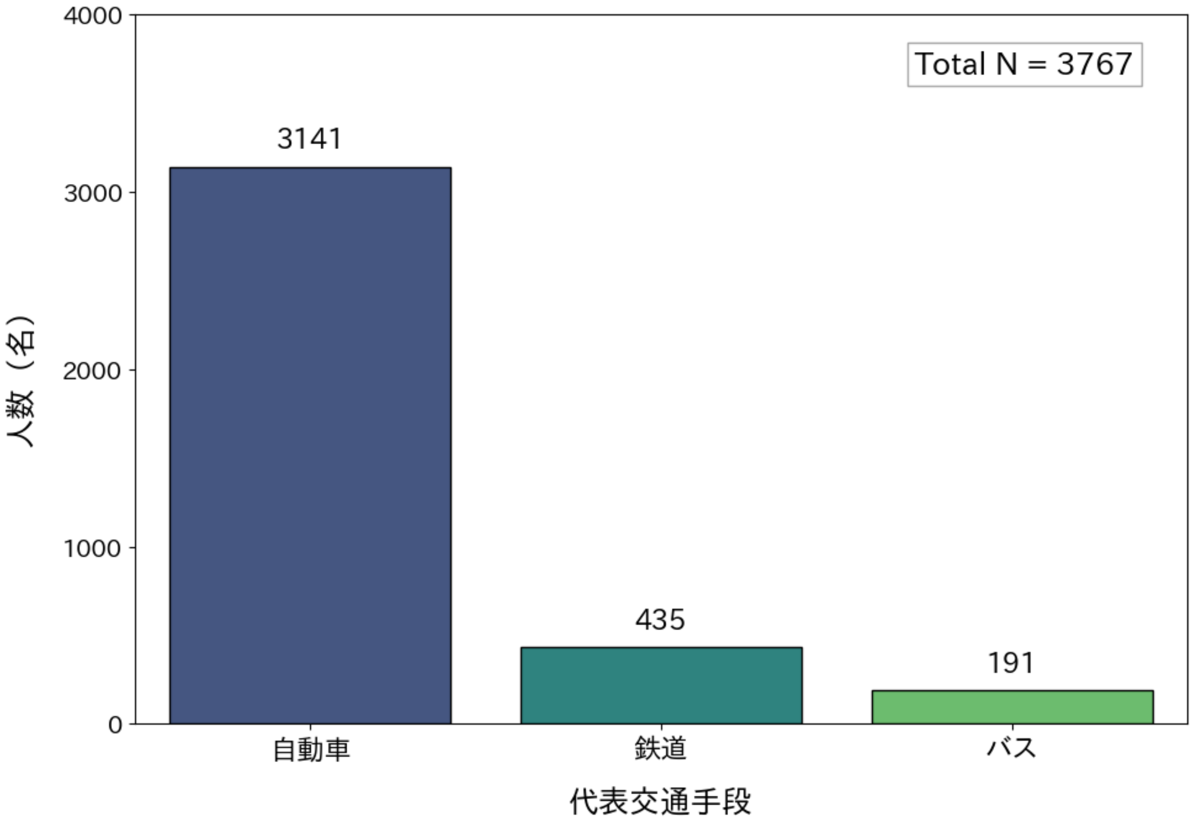
図　調査対象範囲

● 中心市街地と郊外を行き来するトリップのみを抽出している。
○ 中心市街地 ... 伊予鉄市内電車の外側500m圏内
○ 郊外 ... 中心市街地の外側で、バス停から500m圏内
● 代替交通手段生成可否train～総走行距離carはLOS(Level of Service)である。そのトリップにおいて鉄道/バス/車を利用するとしたときの所要時間やアクセスイグレス時間等を計算して記載している。
○
本当にそのトリップで鉄道/バス/車が使われたかは関係ない。
● 個人を特定できるような情報(正確な位置情報等)は取り除いている。

用いたデータの概要



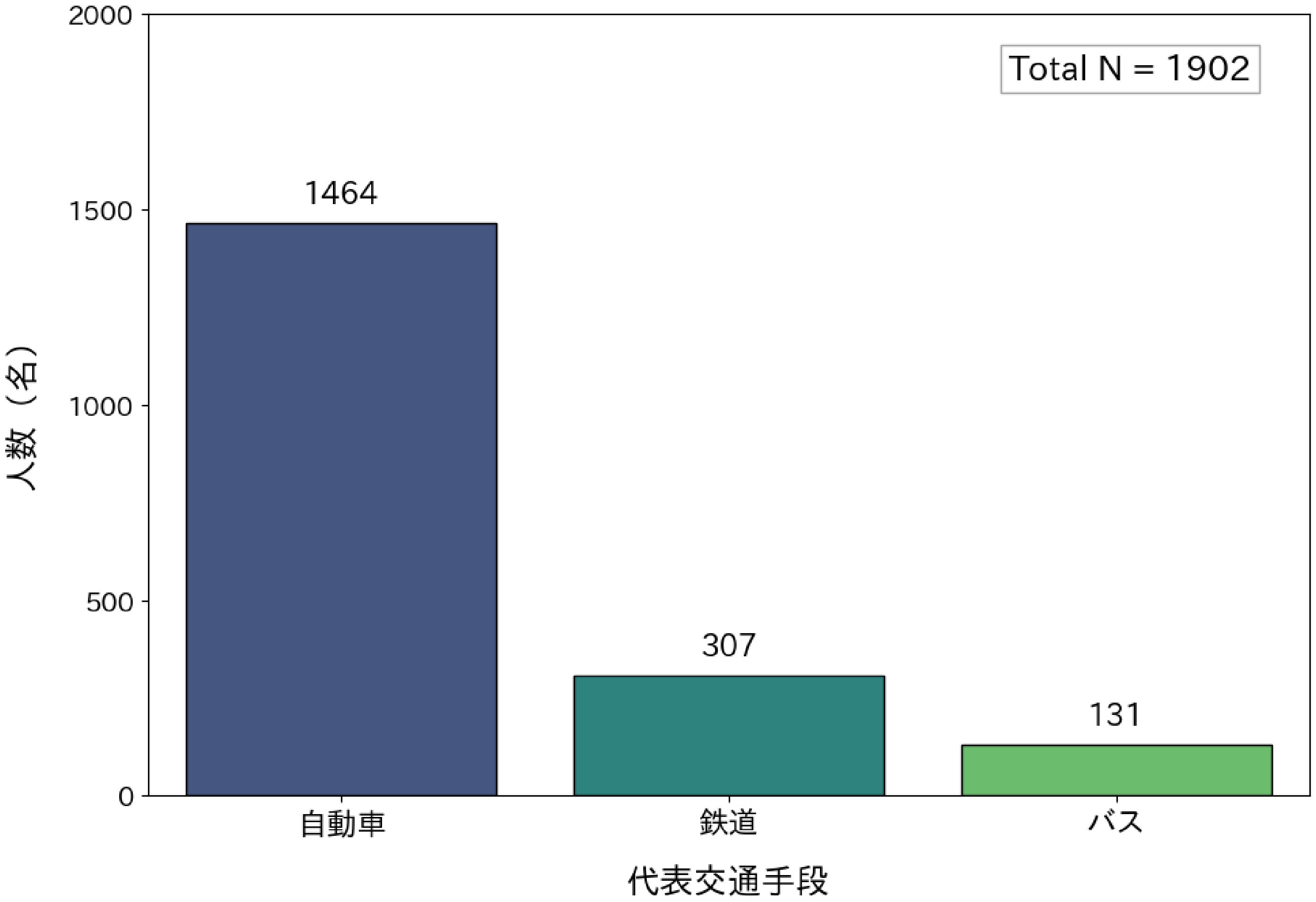
移動目的の分布
人数
1750
1500
1250
1000
750
500
250
0
その他
帰宅
業務
私用
観光・行楽
買物
送迎
通勤
通学
通院
食事・社交・娯楽
移動目的



Total N = 3767
人数（名）
4000
3000
2000
1000
0
3141
435
191
自動車
鉄道
バス
代表交通手段

用いたデータの概要

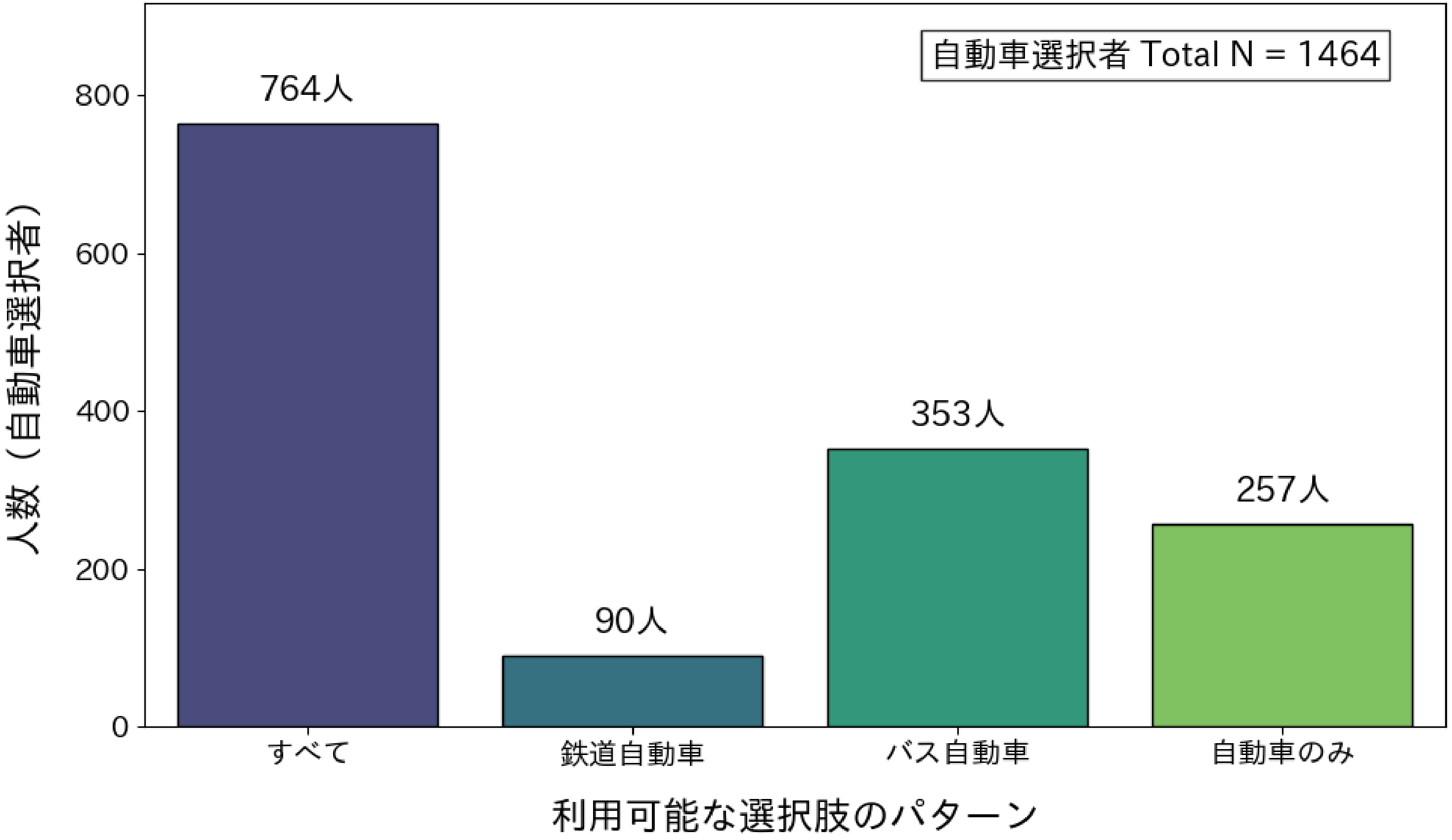
通勤・通学目的にしぼって分析



依然として自動車が多い

用いたデータの概要

通勤・通学目的にしぼって分析

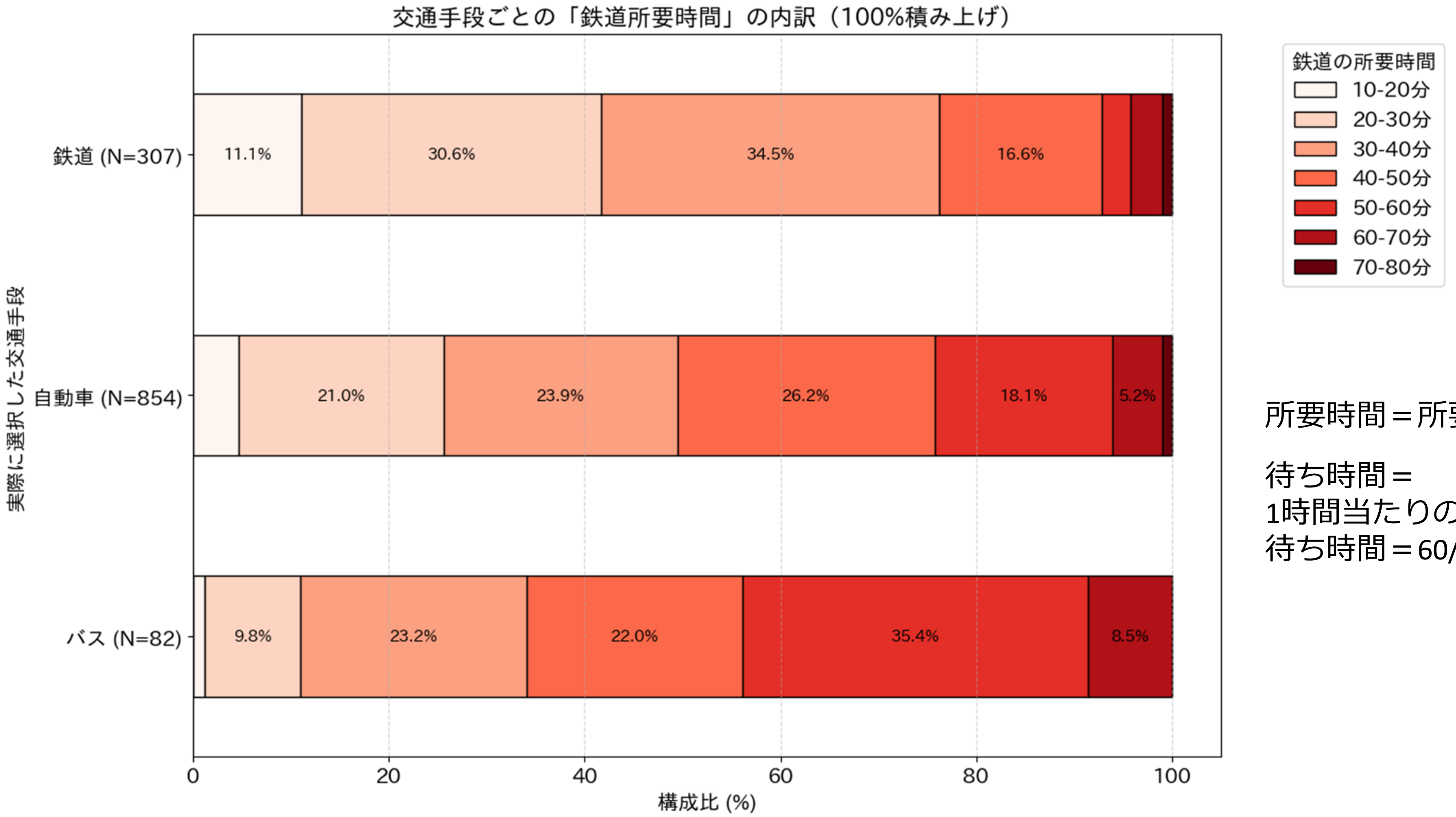


自動車利用者の利用可能な選択肢のパターンをみると、代替交通手段もあるのに自動車を選んでいる人が多い

⬇

そういう人たちの交通手段を変えよう

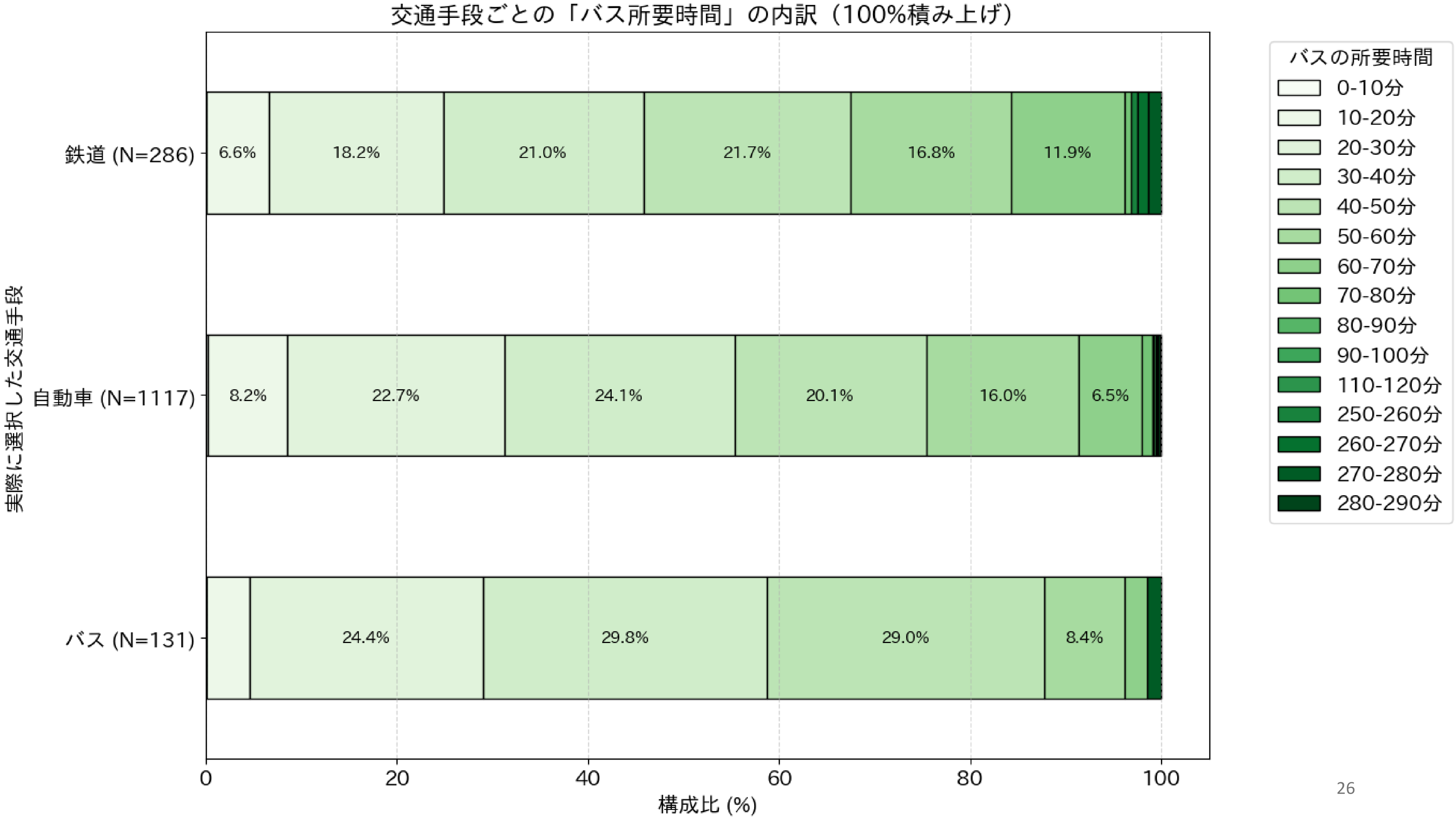
用いたデータの概要



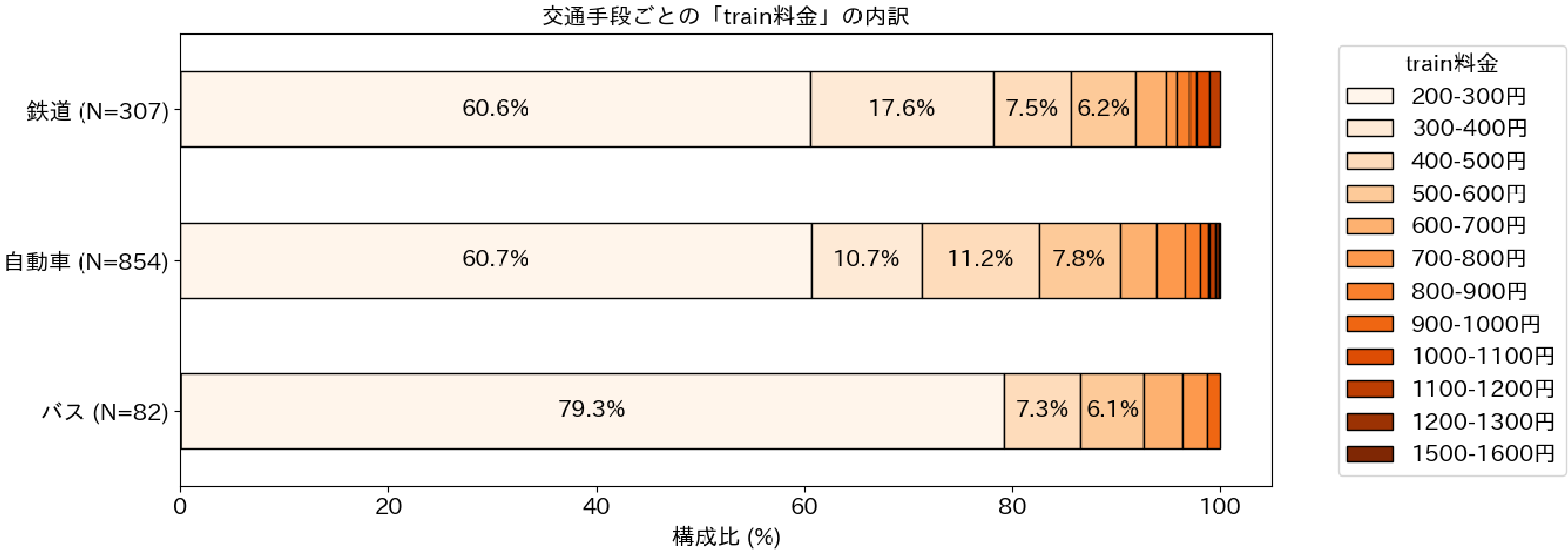
所要時間＝所要時間＋待ち時間

待ち時間＝
1時間当たりの本数＝1日の運行本数/(18-10)
待ち時間＝60/（一時間当たりの本数）/2

鉄道が選択肢に入っている人の中で鉄道を選んだ人の方が
ほかを選んだ人に比べて所要時間が短い傾向？

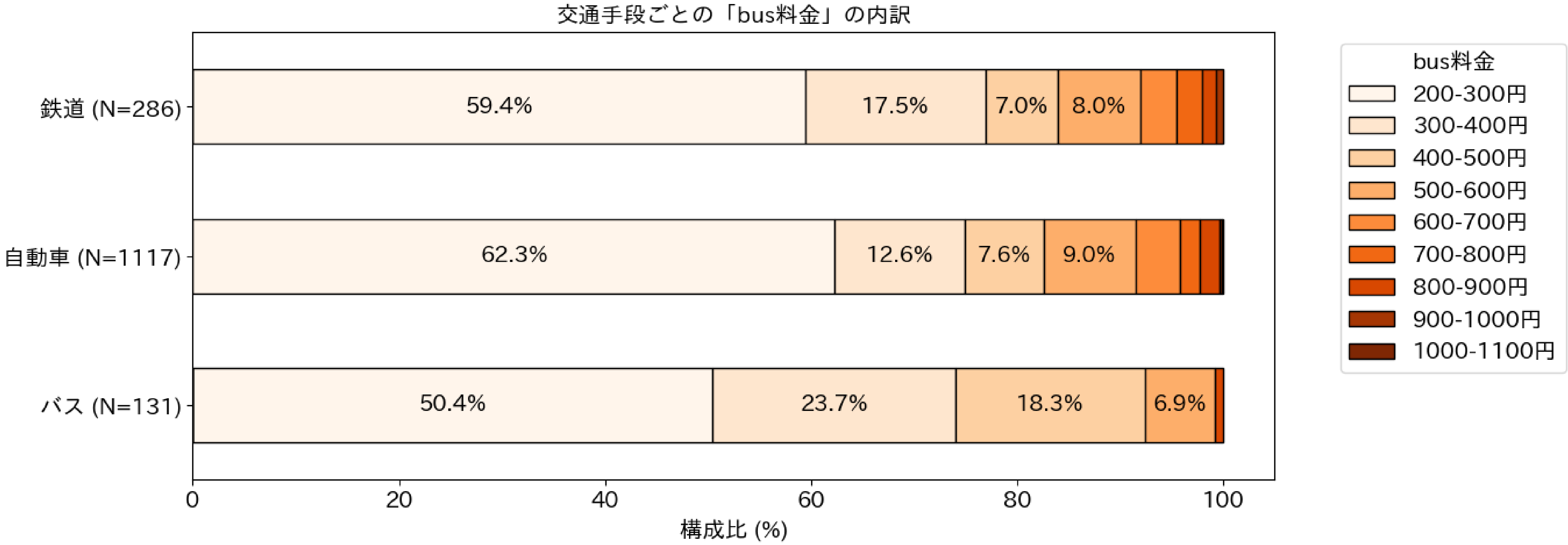


交通手段ごとの「バス所要時間」の内訳（100%積み上げ）
鉄道 (N=286)
自動車 (N=1117)
バス (N=131)
6.6%
18.2%
21.0%
21.7%
16.8%
11.9%
8.2%
22.7%
24.1%
20.1%
16.0%
6.5%
24.4%
29.8%
29.0%
8.4%
実際に選択した交通手段
構成比 (%)
0
20
40
60
80
100
バスの所要時間
0-10分
10-20分
20-30分
30-40分
40-50分
50-60分
60-70分
70-80分
80-90分
90-100分
110-120分
250-260分
260-270分
270-280分
280-290分



鉄道が選択肢に入っている人の中で鉄道を選んだ人の方が
ほかを選んだ人に比べて運賃が短い傾向？

基本運賃230円
3km超えで1km50円ずつ加算
乗り換え回数×230円



バスが選択肢に入っている人の中でバスを選んだ人の方が
ほかを選んだ人に比べて運賃が安い傾向？

基本運賃230円
3km超えで1km50円ずつ加算
乗り換え回数×230円

MNLの構築

$$U_{train} = d_1 x_{time}/10 + d_2 x_{fare}/100 + b_1$$

$$U_{bus} = d_1 x_{time}/10 + d_2 x_{fare}/100 + b_2$$

$$U_{car} = d_1 x_{time}/10 + d_2 x_{fare}/100$$

	推定値	t 値	
b1 (train)	0.161	0.958	
b2 (bus)	-0.980	-5.609	***
d1 (time)	-0.431	-7.546	***
d2 (fare)	0.071	1.530	
サンプル数 (N)		1902	
初期対数尤度 $L(0)$		-1599.214	
最終対数尤度 $L(\beta)$		-1080.426	
決定係数 ρ^2		0.324	
修正済み決定係数 $\bar{\rho}^2$		0.322	

***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$

運賃の符号が正：
料金が上がるほど効用が増す？
→感覚と合わない！

MNLの構築

$$U_{train} = d_1 x_{time}/10 + d_2 \log(x_{fare}) + b_1$$
$$U_{bus} = d_1 x_{time}/10 + d_2 \log(x_{fare}) + b_2$$
$$U_{car} = d_1 x_{time}/10 + d_2 log(x_{fare})$$

	推定値	t 値	
b1 (train)	0.809	2.882	**
b2 (bus)	-0.310	-1.059	
d1 (time)	-0.393	-7.509	***
d2 (log(fare))	-0.672	2.764	**
サンプル数 (N)		1902	
初期対数尤度 $L(0)$		-1599.214	
最終対数尤度 $L(\beta)$		-1077.551	
決定係数 ρ^2		0.326	
修正済み決定係数 $\bar{\rho}^2$		0.324	

***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$

運賃の対数をとってみる

⬇

0円-100円の違いと400-500円の違いは一緒だった（線形）
→0-100円の違いの方が
　大きく出る

符号ok

適合度ok

政策シミュレーション

施策：バスの運賃を変えてみる

施策前

交通手段	選択確率
自動車	0.770
鉄道	0.161
バス	0.069

基本運賃230円
3km超えで1km50円ずつ加算
乗り換え回数×230円

施策後

交通手段	選択確率
自動車	0.753
鉄道	0.158
バス	0.089

基本運賃150円
3km超えで1km30円ずつ加算
乗り換え回数×150円

バスの選択確率が2%向上

NLの構築

MNL(fare対数)

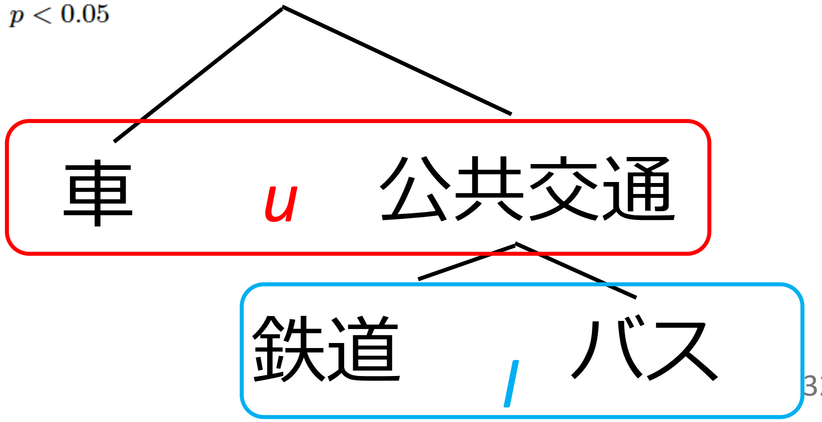
	推定値	t 値	
b1 (train)	0.809	2.882	**
b2 (bus)	-0.310	-1.059	
d1 (time)	-0.393	-7.509	***
d2 (log(fare))	-0.672	-2.764	**
サンプル数 (N)		1902	
初期対数尤度 $L(0)$		-1599.214	
最終対数尤度 $L(\beta)$		-1077.551	
決定係数 ρ^2		0.326	
修正済み決定係数 $\bar{\rho}^2$		0.324	

***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$

NL(fare対数)

	推定値	t 値	
b1 (train)	0.478	1.710	
b2 (bus)	-0.258	-1.130	
d1 (time)	-0.361	-7.543	***
d2 (log(fare))	-0.364	-1.657	
λ	0.551	4.687	***
サンプル数 (N)		1902	
初期対数尤度 $L(0)$		-1599.214	
最終対数尤度 $L(\beta)$		-1072.256	
決定係数 ρ^2		0.330	
修正済み決定係数 $\bar{\rho}^2$		0.326	

***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$



NLにすることで、尤度が大きくなり、
スケールパラメータも0.551鉄道とバスの誤差項の相関がありそう？

⬇

NLを使った方がモデルの説明力が上がった

ご清聴ありがとうございました

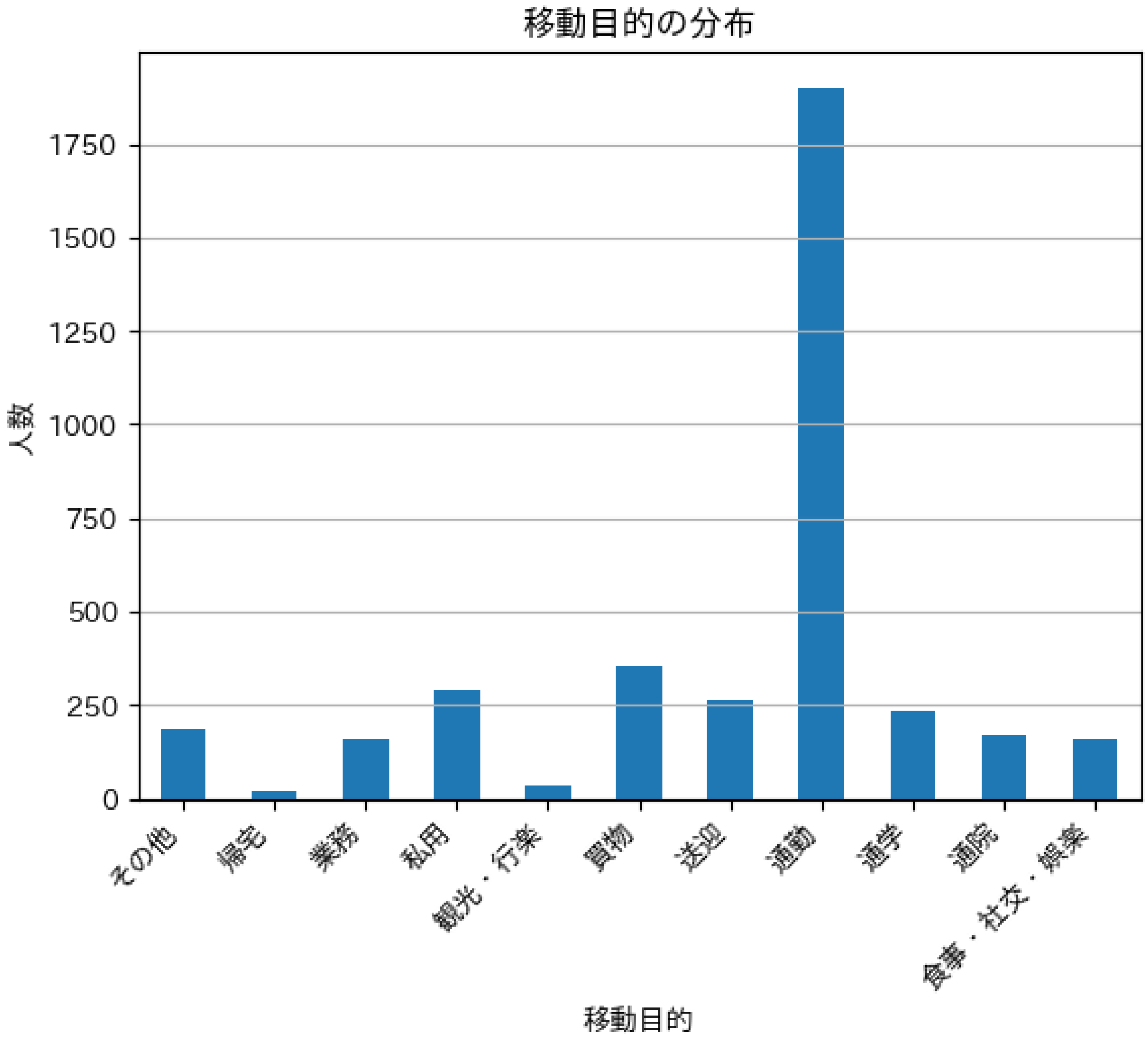
政策シミュレーション

```
=== Multinomial Logit Model Final Results ===
Parameter    |      Est. |      t-val |      p-val | Signif.
------------------------------------------------------------------
b1 (train)   |    0.1612 |     0.9575 |     0.3384 |
b2 (bus)     |   -0.9801 |    -5.6090 |     0.0000 |    ***
d1 (time)    |   -0.4313 |    -7.5462 |     0.0000 |    ***
d2 (fare)    |    0.0715 |     1.5296 |     0.1263 |
------------------------------------------------------------------
Signif. codes:  0.01 '***', 0.05 '**', 0.1 '*'

=== Model Fit Indices (MNL) ===
初期対数尤度 L(0)       : -1599.2136
最終対数尤度 L(beta)    : -1080.4262
決定係数 rho^2          :     0.3244
修正済み決定係数 adj.rho^2:     0.3219
```

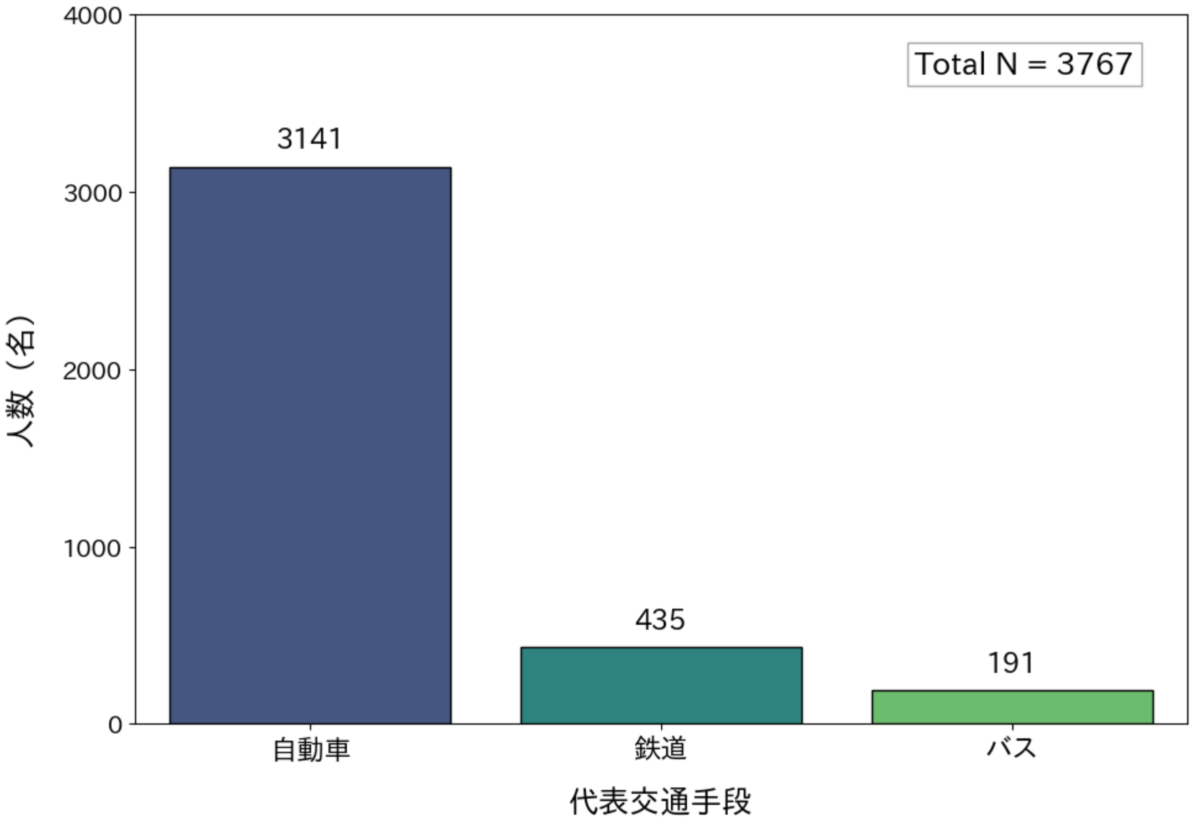
```
=== Mixed Logit Model Results ===
Parameter    |      Est. |   Std.Err |      t-val
--------------------------------------------------------
b_train      |    0.8138 |    1.2294 |     0.6620
b_bus        |   -0.4147 |    0.5956 |    -0.6963
mu_time      |   -0.7448 |    0.3100 |    -2.4028
sig_time     |    0.4026 |    0.2611 |     1.5422
beta_fare    |    0.1376 |    0.2095 |     0.6568
--------------------------------------------------------
初期対数尤度 L(0)       : -2089.56
最終対数尤度 L(beta)    : -1063.19
決定係数 rho^2          : 0.4912
修正済み決定係数 adj.rho^2: 0.4888
```

```
=== Nested Logit Model Final Results ===
Parameter    |      Est. |      t-val |      p-val | Signif.
------------------------------------------------------------------
b_train      |    0.0441 |     0.2769 |     0.7819 |
b_bus        |   -0.6162 |    -4.4205 |     0.0000 |    ***
beta_time    |   -0.3719 |    -7.0152 |     0.0000 |    ***
beta_fare    |    0.0608 |     1.5947 |     0.1109 |
lambda       |    0.4807 |     4.7903 |     0.0000 |    ***
------------------------------------------------------------------
Signif. codes:  0.01 '***', 0.05 '**', 0.1 '*'

=== Model Fit Indices ===
初期対数尤度 L(0)       : -1402.3374
最終対数尤度 L(beta)    : -1072.5790
決定係数 rho^2          :     0.2351
修正済み決定係数 adj.rho^2:     0.2316
```

課題

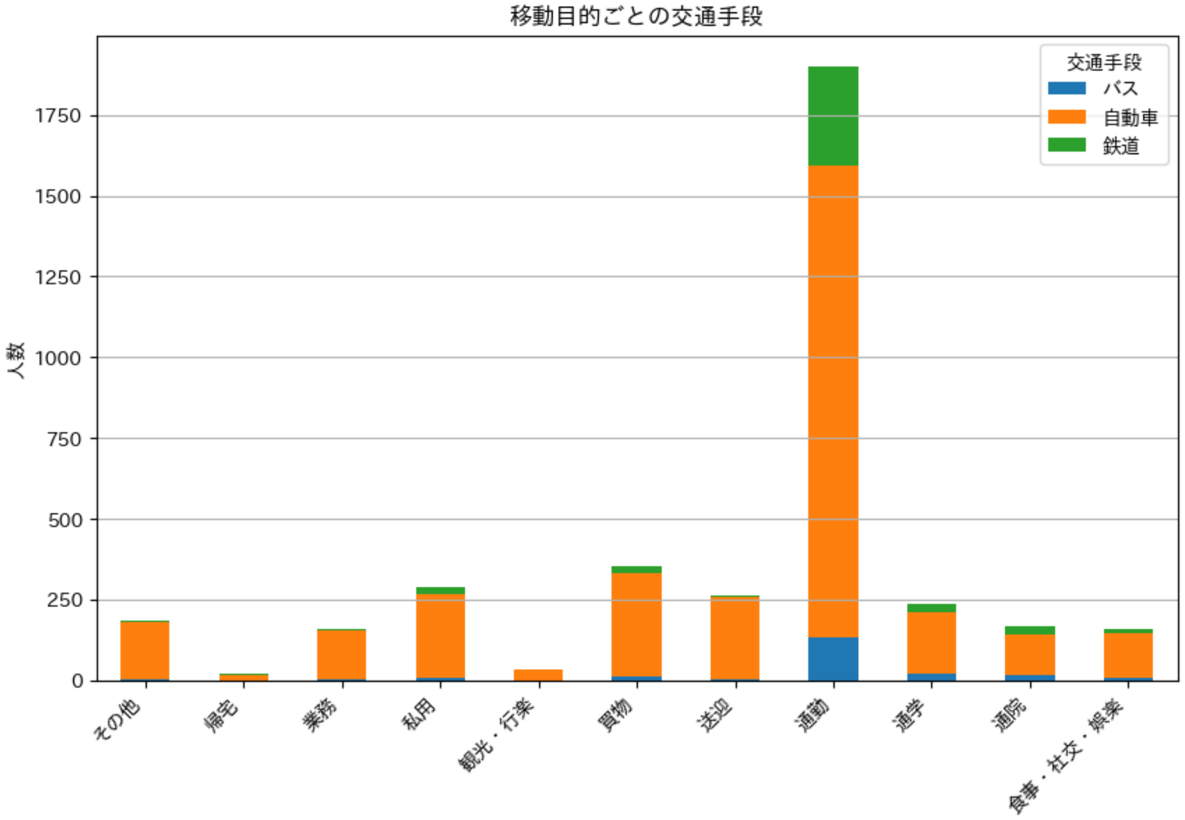


移動目的の分布
1750
1500
1250
1000
750
500
250
0
人数
その他
帰宅
業務
私用
観光・行楽
買物
送迎
通勤
通学
通院
食事・社交・娯楽
移動目的

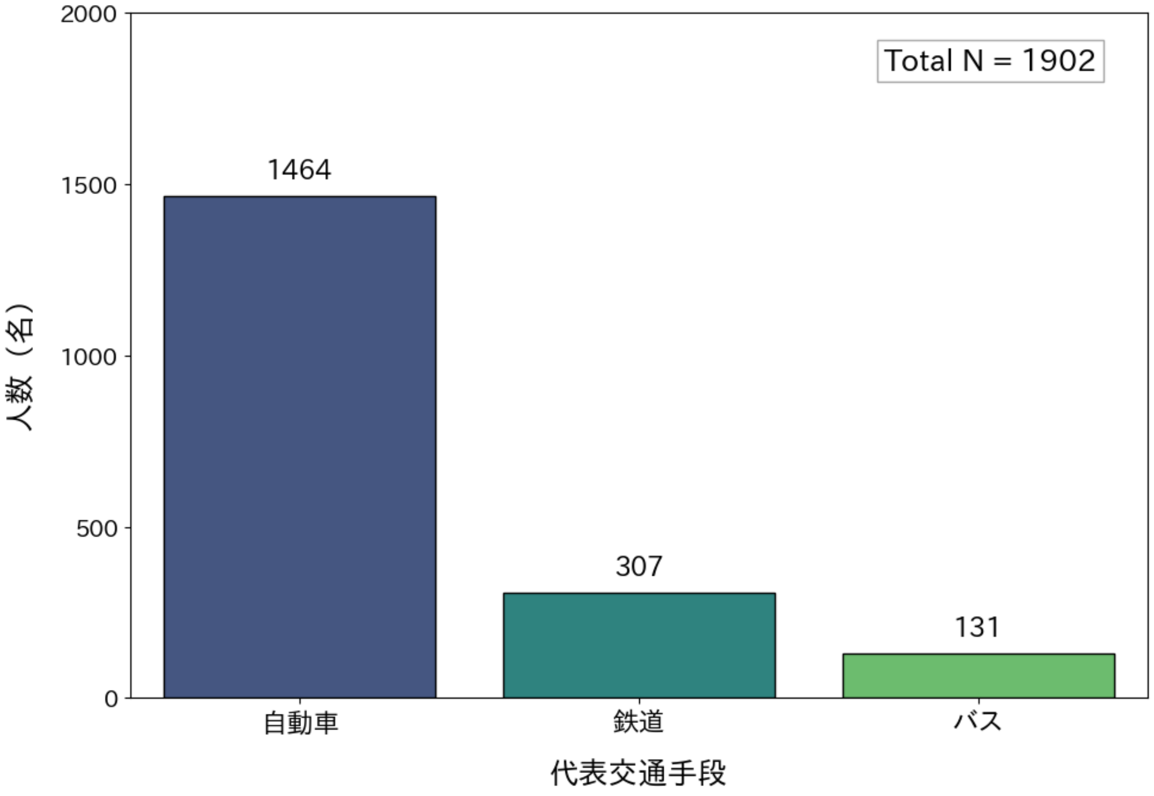


4000
3000
2000
1000
0
人数（名）
Total N = 3767
3141
435
191
自動車
鉄道
バス
代表交通手段

課題

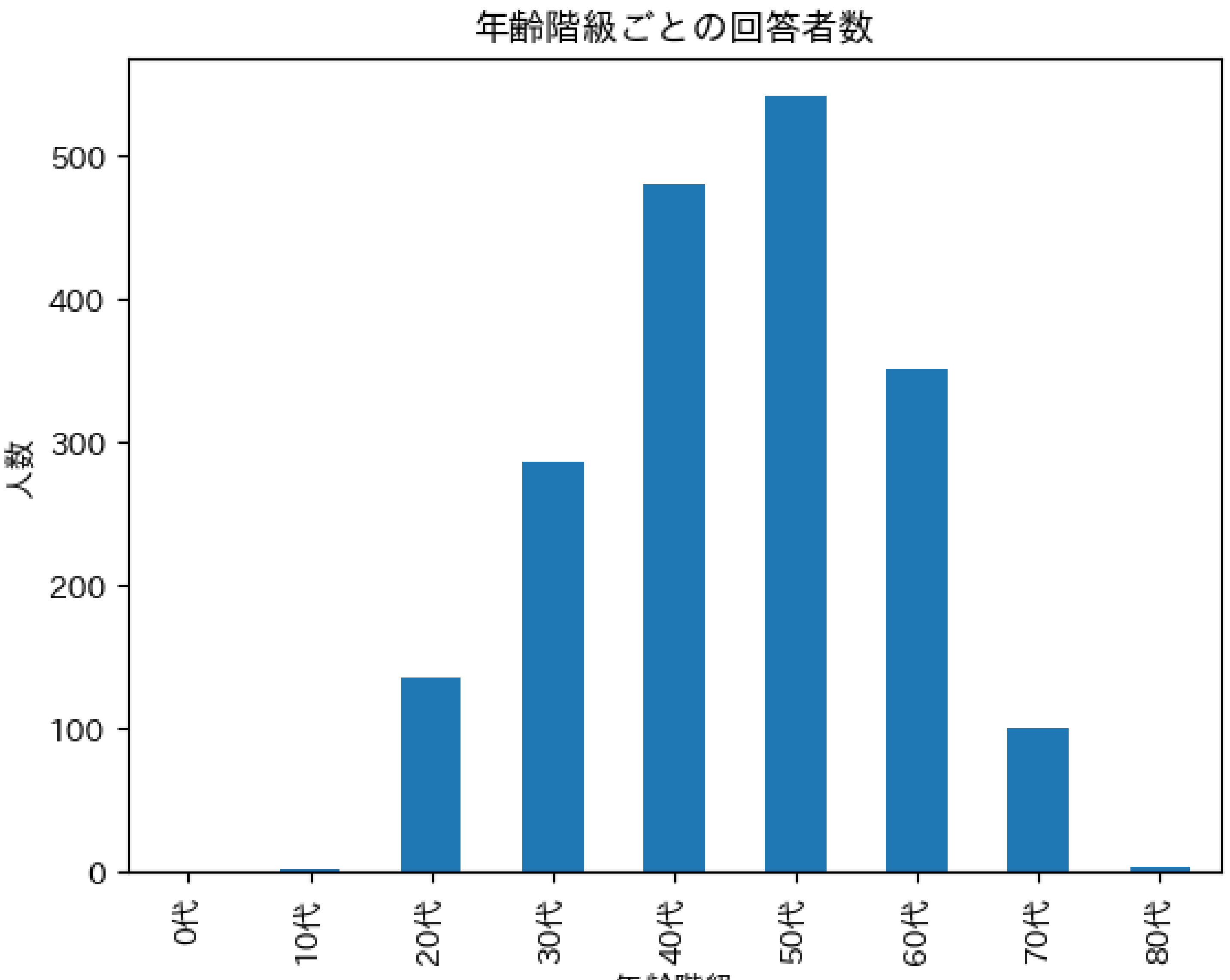


移動目的ごとの交通手段
交通手段
バス
自動車
鉄道
1750
1500
1250
1000
750
500
250
0
人数
その他
帰宅
業務
私用
観光・行楽
買物
送迎
通勤
通学
通院
食事・社交・娯楽
移動目的

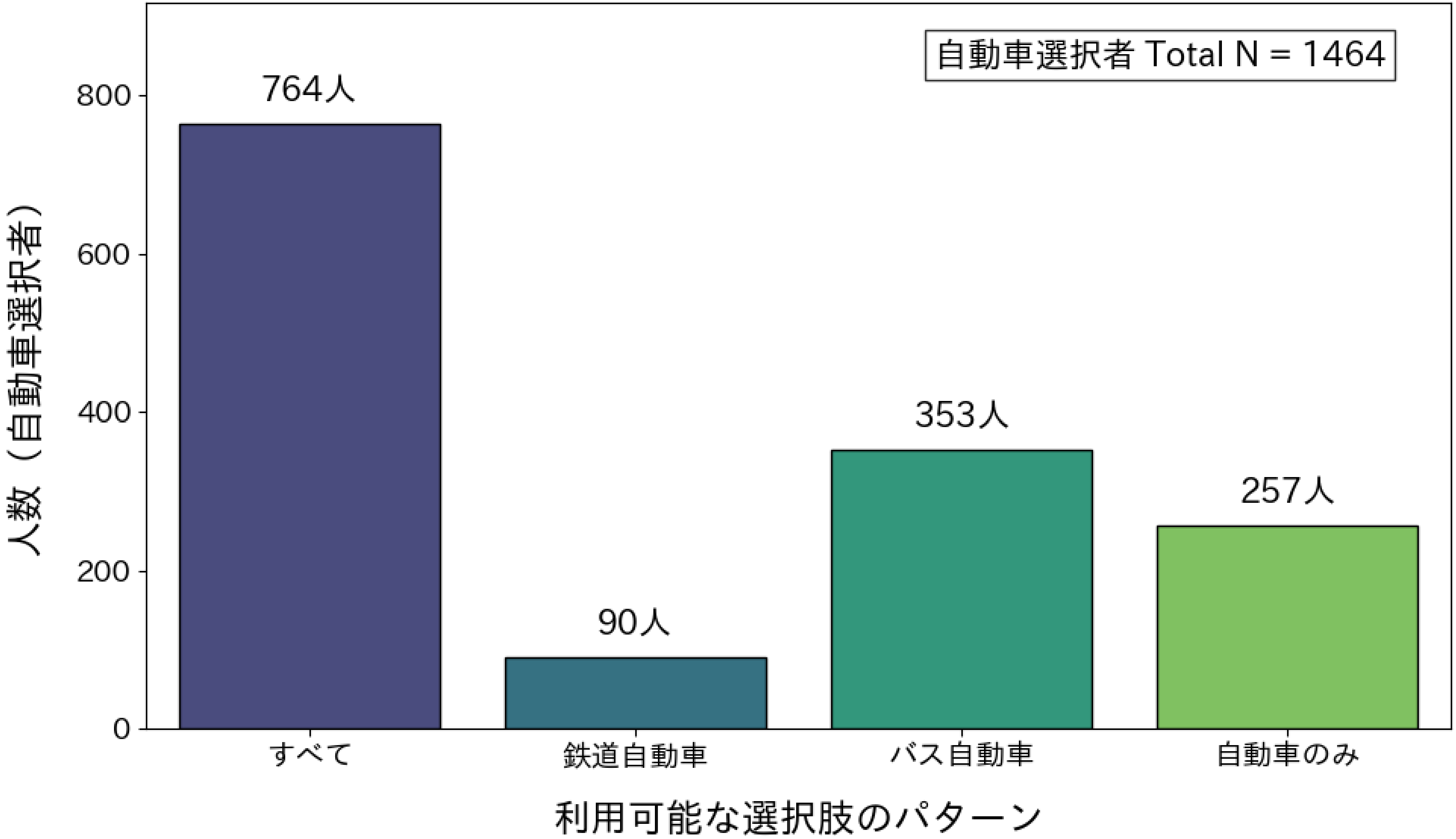


Total N = 1902
1464
307
131
2000
1500
1000
500
0
人数（名）
自動車
鉄道
バス
代表交通手段

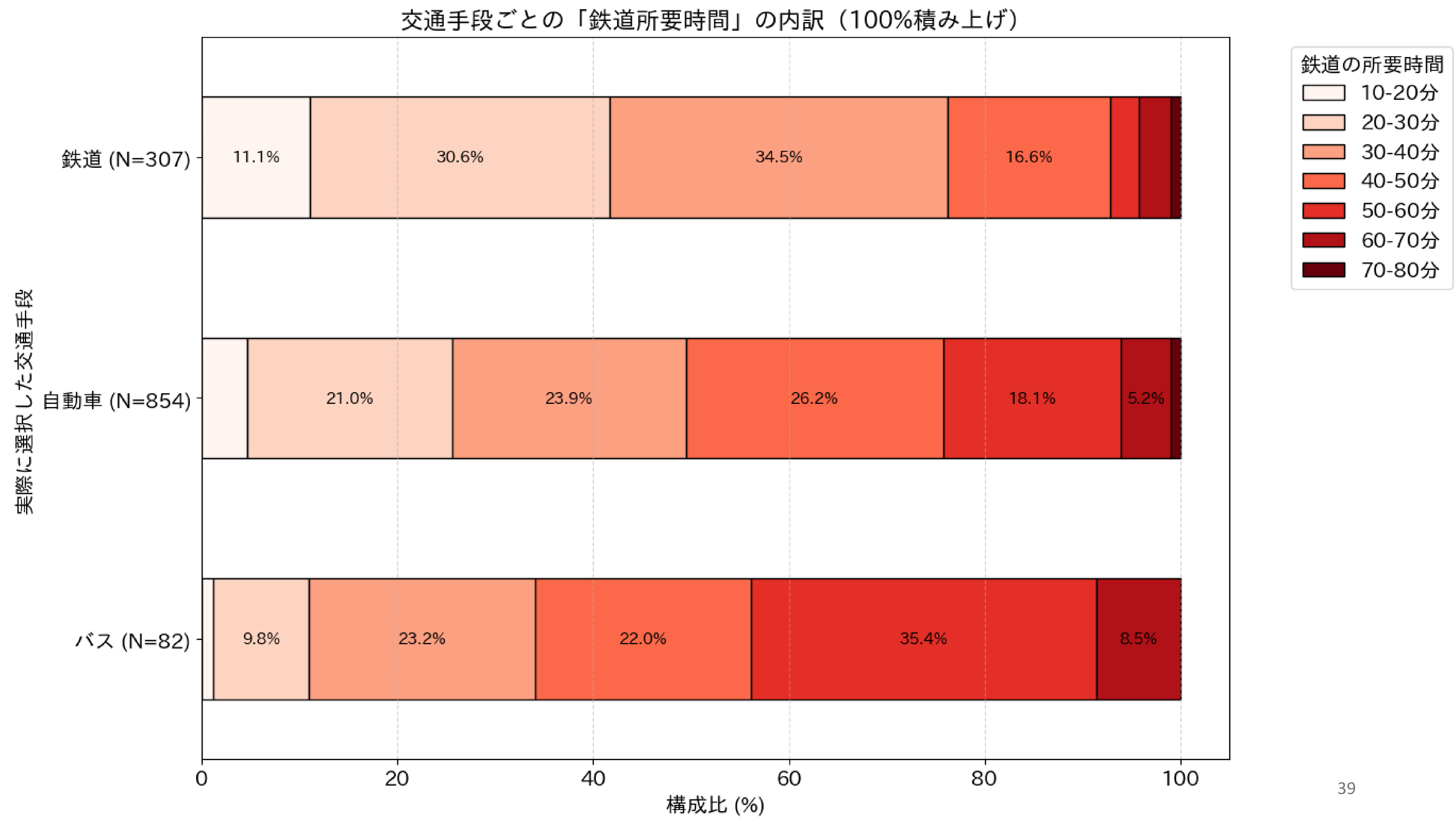
課題


年齢階級ごとの回答者数
人数
0
100
200
300
400
500
0代
10代
20代
30代
40代
50代
60代
70代
80代

課題

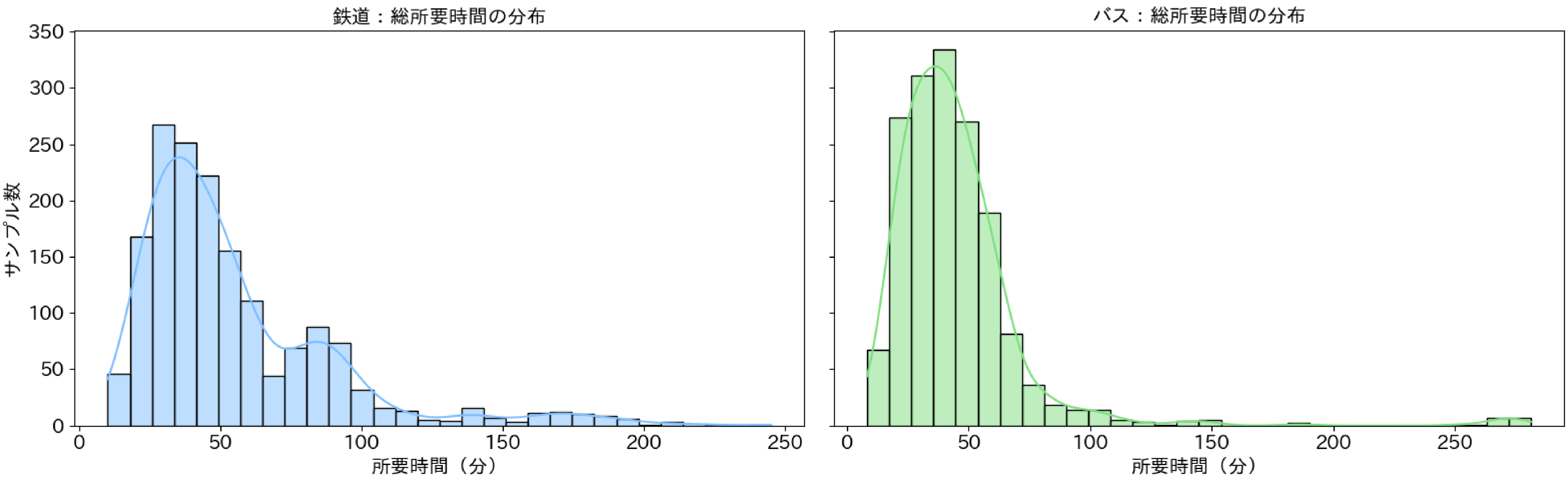


自動車選択者 Total N = 1464
764人
90人
353人
257人
800
600
400
200
0
人数（自動車選択者）
すべて
鉄道自動車
バス自動車
自動車のみ
利用可能な選択肢のパターン

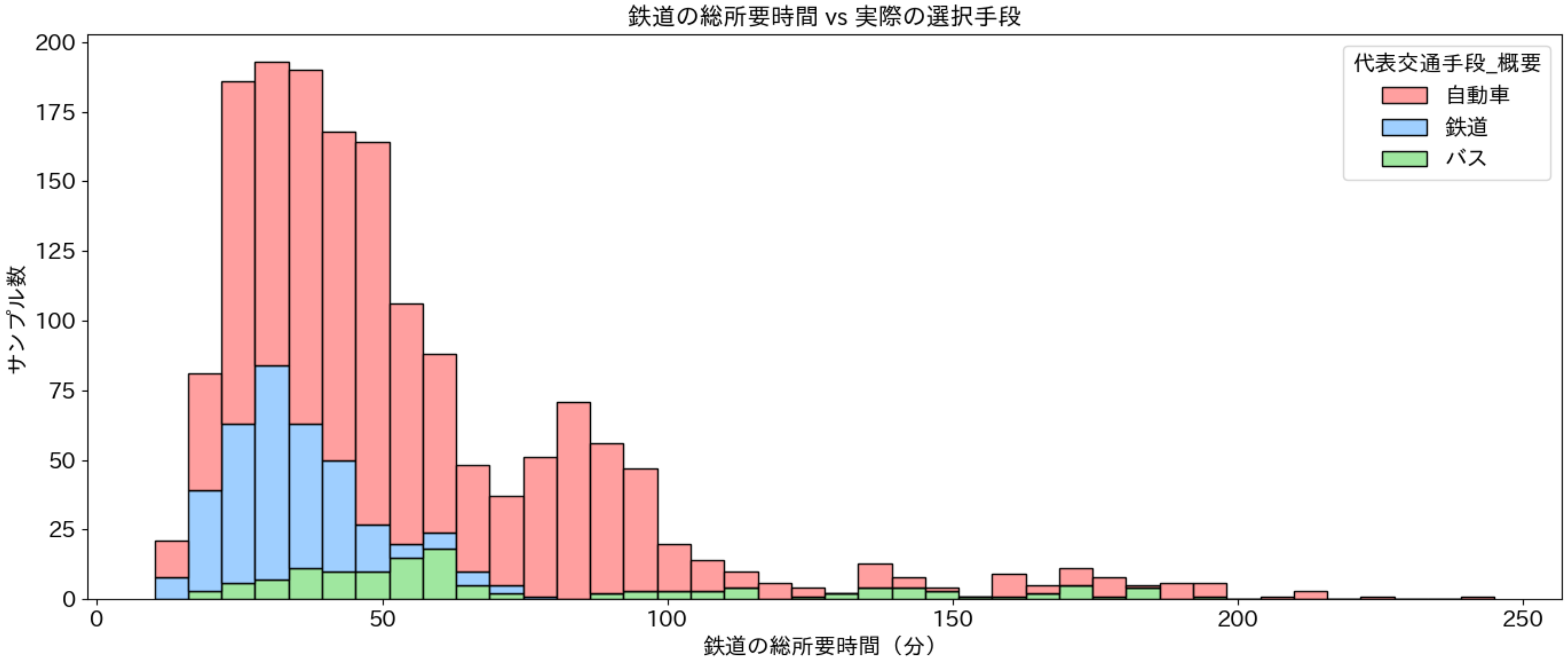


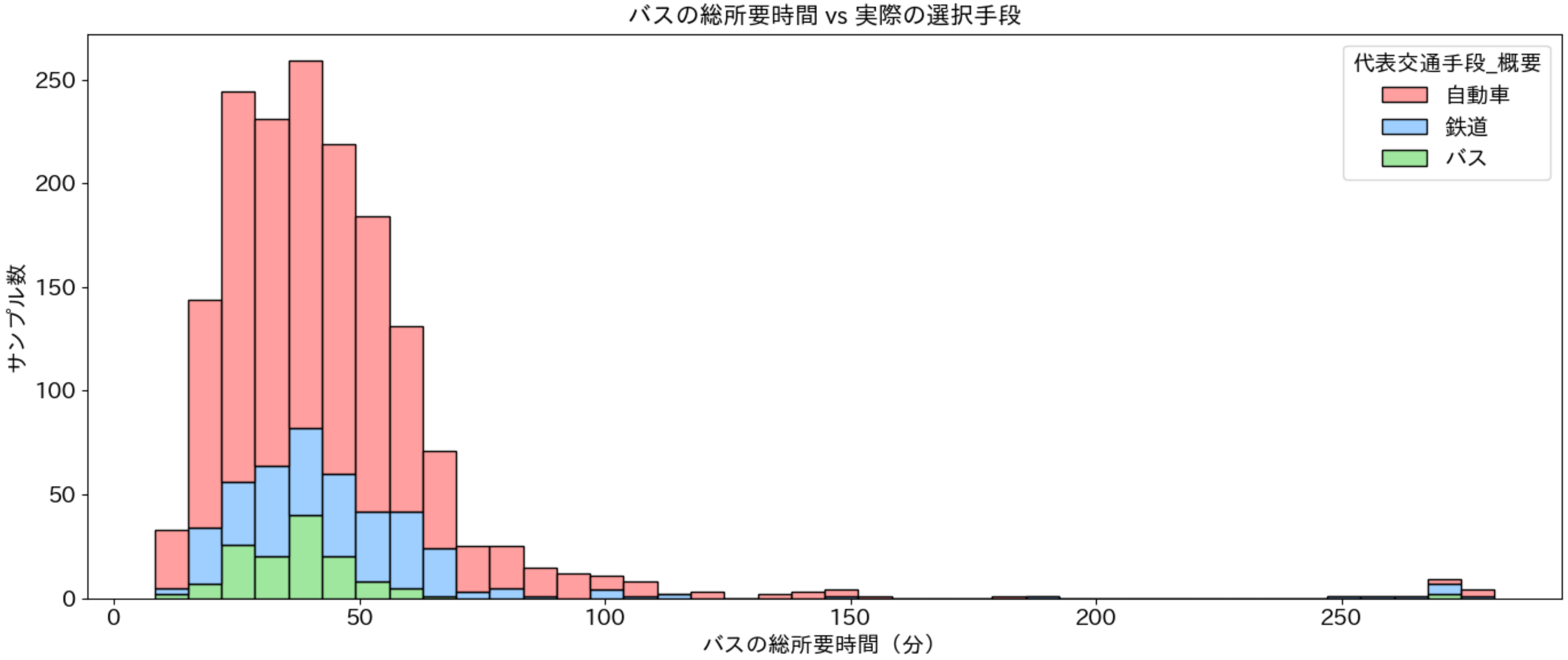
交通手段ごとの「鉄道所要時間」の内訳（100%積み上げ）
鉄道の所要時間
10-20分
20-30分
30-40分
40-50分
50-60分
60-70分
70-80分
実際に選択した交通手段
鉄道 (N=307)
自動車 (N=854)
バス (N=82)
11.1%
30.6%
34.5%
16.6%
21.0%
23.9%
26.2%
18.1%
5.2%
9.8%
23.2%
22.0%
35.4%
8.5%
0
20
40
60
80
100
構成比 (%)

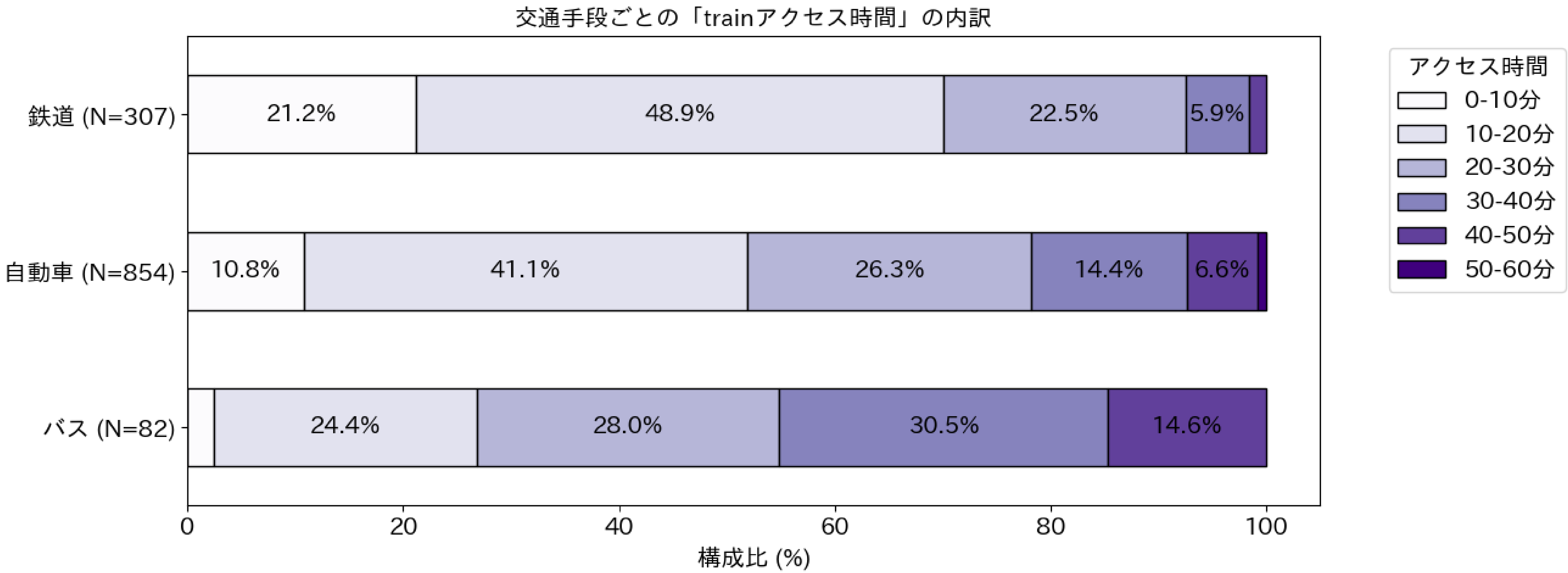
公共交通機関の所要時間分布（自動車以外の選択肢がある層）



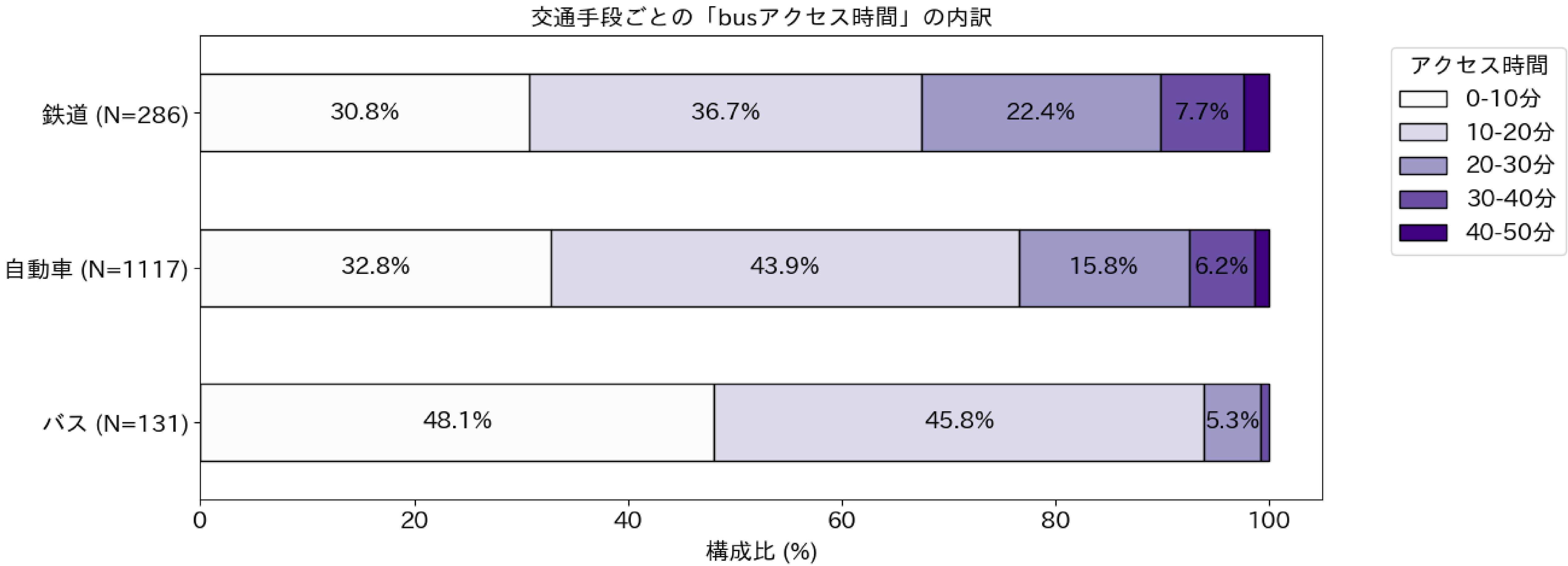
LOSの作り方的に運行本数が少ないほど所要時間が非常に大きくなる


鉄道の総所要時間 vs 実際の選択手段
サンプル数
鉄道の総所要時間（分）
代表交通手段_概要
自動車
鉄道
バス
0
25
50
75
100
125
150
175
200
250


バスの総所要時間 vs 実際の選択手段
サンプル数
バスの総所要時間（分）
代表交通手段_概要
自動車
鉄道
バス
0
50
100
150
200
250



交通手段ごとの「trainアクセス時間」の内訳
鉄道 (N=307)
21.2%
48.9%
22.5%
5.9%
自動車 (N=854)
10.8%
41.1%
26.3%
14.4%
6.6%
バス (N=82)
24.4%
28.0%
30.5%
14.6%
0
20
40
60
80
100
構成比 (%)
アクセス時間
0-10分
10-20分
20-30分
30-40分
40-50分
50-60分


交通手段ごとの「busアクセス時間」の内訳
鉄道 (N=286)
自動車 (N=1117)
バス (N=131)
30.8%
36.7%
22.4%
7.7%
32.8%
43.9%
15.8%
6.2%
48.1%
45.8%
5.3%
0
20
40
60
80
100
構成比 (%)
アクセス時間
0-10分
10-20分
20-30分
30-40分
40-50分