

2016/05/22

# 生存時間モデルの推定 with R

まず，生存時間分析用ライブラリを  
下記コマンドでインストールして下さい  
`install.packages("survival")`

# 基礎概念

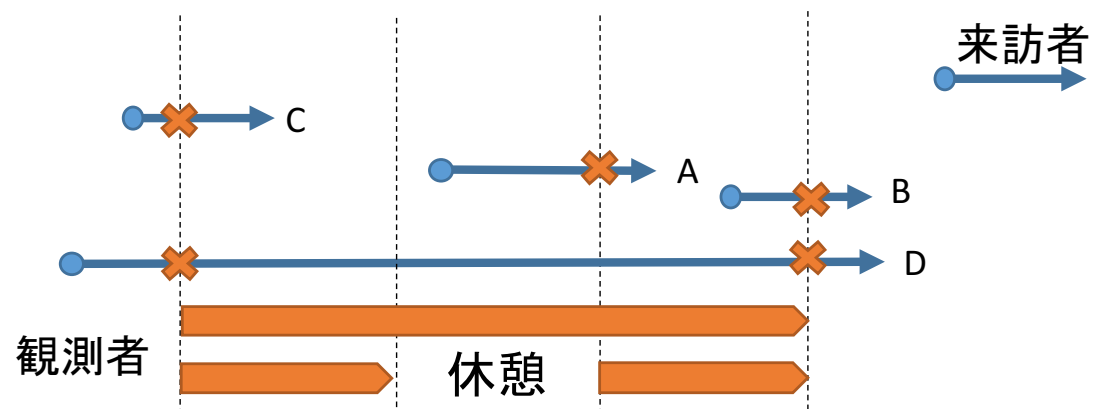
- 生存時間モデルとは ある基準の時刻からある事象が生起、あるいは終了するまでの時間の解析を対象とするモデル[期間モデル]
- 打ち切り
- 生存関数とハザード関数

## 特徴

解析対象となる事象が生起するまでの時間は必ず正の値

時間の分布の裾が右に長くなることが多い

分布の正規性を仮定することが適切でない場合が多い



# 基礎概念

- 生存時間モデルとは
- 打ち切り
- 生存関数とハザード関数

対象とする事象の生起するまでの時間  $T$  の分布を以下のように表すことが多い

生存関数  $S(t)$

— 対象とする事象がある時点  $t$  においてまだ生起していない確率

ハザード関数  $h(t)$

— 対象とする事象がある時点  $t$  においてまだ生起していないという条件の下で、次の瞬間に事象が生起するという条件付き確率

生存関数  $S(t)$

$$S(t) = P_r(T \geq t) = 1 - P_r(T \leq t) = 1 - F(t)$$

ハザード関数  $h(t)$

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P_r(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

$$= \lim_{\Delta t \rightarrow 0} \frac{S(t) - S(t + \Delta t)}{\Delta t * S(t)} = - \frac{d(\log S(t))}{dt} = \frac{f(t)}{S(t)}$$

$f(t)$  は  $T$  の確率密度関数

生存関数とハザード関数の関係

$$S(t) = \exp\left(-\int_0^t h(u) du\right)$$

# 生存時間の解析方法

- 生存時間のモデル化

- Non-parametric model

- Semi-parametric model

- Parametric model

	共変量がパラメータとしてモデルに導入されている	生存時間の分布に特定の確率分布が仮定されている	交通行動分析の分野での使用
Non-parametric model	×	×	△
Semi-parametric model	○	×	○
Parametric model	○	○	○

## Kaplan-Meier法

$d_i$  = 時点  $t_i$  で事象が生じたケース数、事象が生じた時点を  $t_1, t_2, \dots$

$n_i$  = 時点  $t_i$  の直前のリスク集合の大きさ (= その直前に観測されていたケース数)

生存時間が  $t_{i-1}$  より長いという条件の下で生存時間が  $t_i$  以上である条件付き確率の推定量

$$\frac{\hat{S}(t)}{\hat{S}(t_{i-1})} = \frac{n_i - d_i}{n_i}$$



生存関数  $S(t)$  の Kaplan-Meier 推定量  $\hat{S}(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i}$

# 生存時間の解析方法

- 生存時間のモデル化

- Non-parametric model
- Semi-parametric model
- Parametric model

	共変量がパラメーターとしてモデルに導入されている	生存時間の分布に特定の確率分布が仮定されている	交通行動分析の分野での使用
Non-parametric model	×	×	△
Semi-parametric model	○	×	○
Parametric model	○	○	○

## 比例ハザードモデルでのハザード関数

$$h(t|x_i) = h_0(t) \exp(\beta x_i)$$

$h(t|x_i)$ : 共変量ベクトル  $x_i$  を持つケース  $i$  のハザード関数、 $h_0(t)$ : 基準ハザード、

$\beta$ : 未知パラメーターベクトル、 $x_i$ : ケース  $i$  の共変量ベクトル

(ケース間でのハザード関数の比は時点によらず一定と仮定)

パラメーターの推定: 部分尤度法、Han and Hausman法

# 生存時間の解析方法

## • 生存時間のモデル化

- Non-parametric model
- Semi-parametric model
- Parametric model

共変量がパラメータとしてモデルに導入されている	生存時間の分布に特定の確率分布が仮定されている	交通行動分析の分野での使用
×	×	△
○	×	○
○	○	○

### ①比例ハザードモデル

### ②加速故障モデル

$$h(t|x_i) = h_0(t \exp(-\beta x_i)) \exp(-\beta x_i)$$

生存時間への共変量の

影響の仕方が異なる

→2つのモデルのうち対象とする

問題に適したモデルを選択する

確率分布	ハザード関数 $h_0(t)$	生存関数 $S_0(t)$
指数分布	$\lambda$	$\exp(-\lambda t)$
ワイブル分布	$\gamma \lambda t^{\gamma-1}$	$\exp(-\lambda t^\gamma)$
対数ロジスティック分布	$\frac{\gamma \lambda t^{\gamma-1}}{1 + \lambda t^\gamma}$	$\frac{1}{1 + \lambda t^\gamma}$
対数正規分布	$\frac{1}{t} \varphi\left(\frac{\log(t) - \mu}{\sigma}\right)$	$1 - \varphi\left(\frac{\log(t) - \mu}{\sigma}\right)$
ゴンベルツ分布	$\gamma \exp(-\lambda t)$	$\exp((\gamma/\lambda)(\exp(-\lambda t) - 1))$

## 適用事例

### • 活動時間・滞在時間の分析

- 個々人の日常的な自由活動の活動時間 / 観光地における滞在時間
- 詳しく知りたい人は下記二編の論文を参照:  
Bhat, C.R. and Pinjari, A.R. (2008). Duration Modeling. In David A. Hensher , Kenneth J. Button (ed.) Handbook of Transport Modelling (Handbooks in Transport, Volume 1) , pp.105-131.  
[http://www.caee.utexas.edu/prof/bhat/ABSTRACTS/Duration\\_Modeling\\_Nov1.pdf](http://www.caee.utexas.edu/prof/bhat/ABSTRACTS/Duration_Modeling_Nov1.pdf)  
Hensher, D.A. and Mannering, F.L. (1994). Hazard-based duration models and their application to transport analysis, Transport Reviews, Vol. 14, No. 1, pp. 63-82.  
<http://courses.washington.edu/cee500/Transport%20Reviews.pdf>
- 災害発生時 (e.g. ハリケーン・豪雨) の避難行動開始時間の分析等にも応用

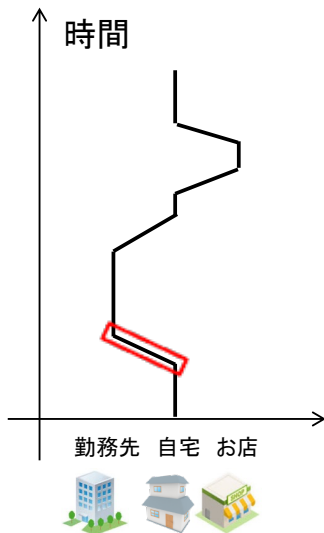
### • 自動車保有行動の分析

- 自動車保有期間・自動車取り替え行動の分析
- 詳しく知りたい人は山本先生 (名大) の博士論文のレビュー章を見て下さい  
山本俊行 (2000) 連続時間軸上における世帯の自動車保有更新行動及び世帯内での配分・利用行動に関する研究. 京都大学博士論文.  
[http://repository.kulib.kyoto-u.ac.jp/dspace/bitstream/2433/151521/3/D\\_Yamamoto\\_Toshiyuki.pdf](http://repository.kulib.kyoto-u.ac.jp/dspace/bitstream/2433/151521/3/D_Yamamoto_Toshiyuki.pdf)

# Activity-Based Travel Choice Modeling

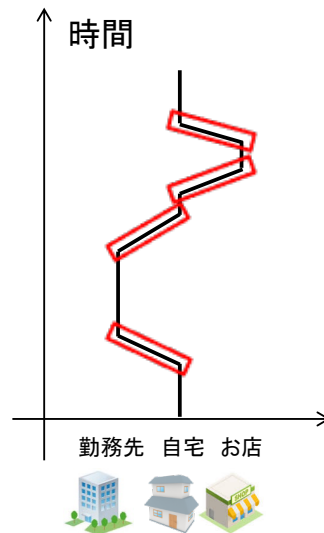
- トリップ

ある地点から  
ある地点への移動



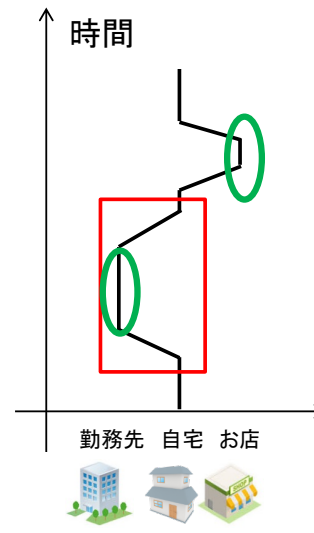
- トリップチェーン

1日のトリップの連続



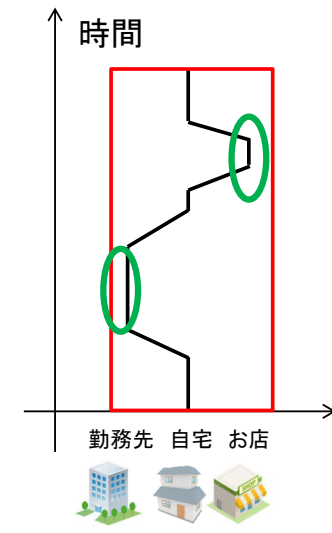
- ツアー

ある地点を出発して、  
ある地点に戻るまでの行動



- アクティビティパターン

1日の活動すべて (下図の場合、  
自宅→通勤→自宅→私事→自宅)



- 個人の一日の生活行動すべてを包括した分析
- トリップだけでなく活動 (Activity) とそのつながりを明示的に考慮
- 個人一人一人の一日の活動のマイクロな動的シミュレーション



Samiul, H., Rodrigo, M.R. and Satish, U. (2013). A random-parameter hazard-based model to understand household evacuation timing behavior, Transportation research part C: Emerging technologies, Vol. 27, pp. 108-116.

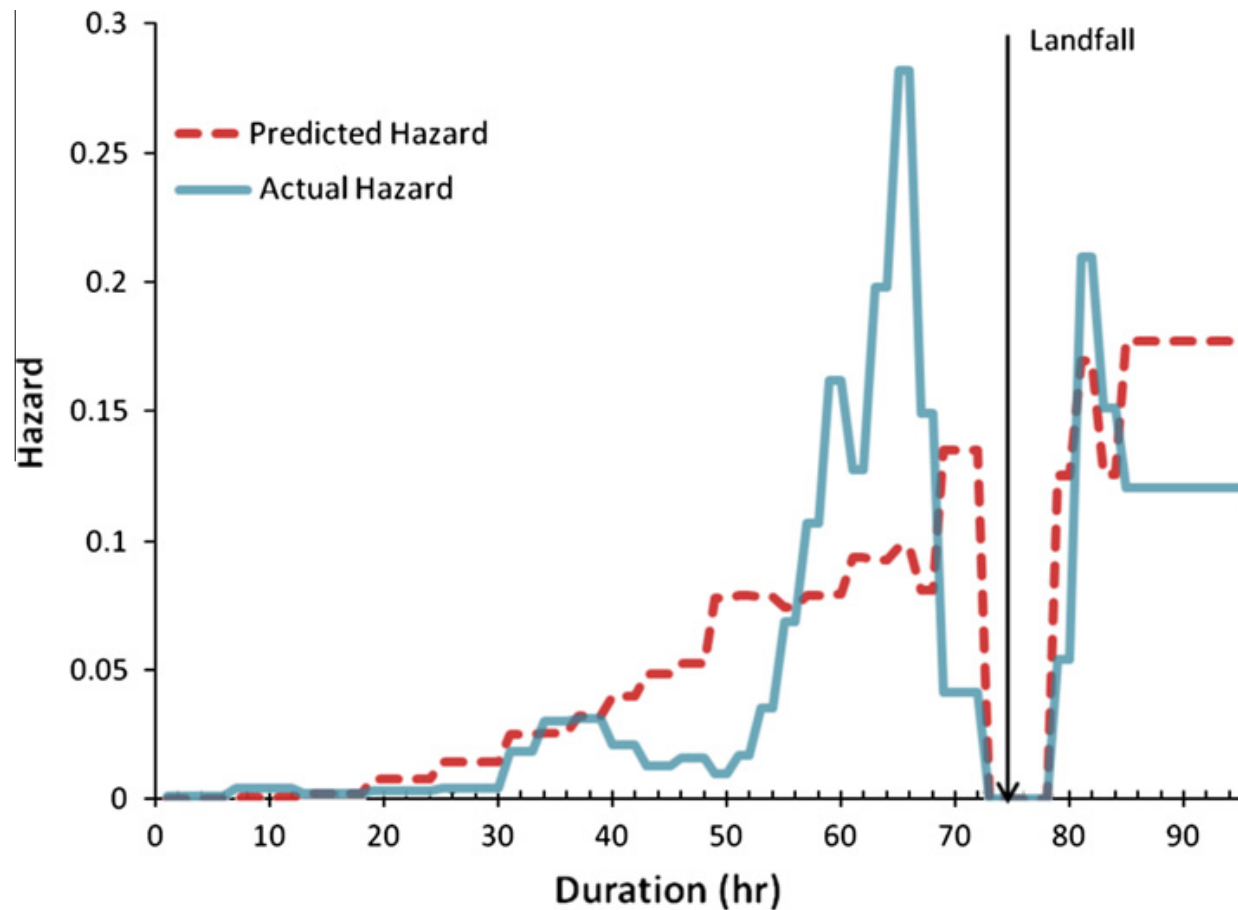


Fig. 2. Actual and predicted hazard at different departure times.

Hurricane Ivan (2004年) 襲来時のアラバマ州の避難行動調査を使った分析

# 本日の演習で用いるデータ

- 平成22年全国都市交通特性調査(全国PT調査)
- 各都市圏で行われる通常Person Trip調査と異なり、規模や構造の異なる各都市で満遍なく調査を行い、都市構造が交通行動に及ぼす影響等を横断的に分析するのに、主に用いられる。

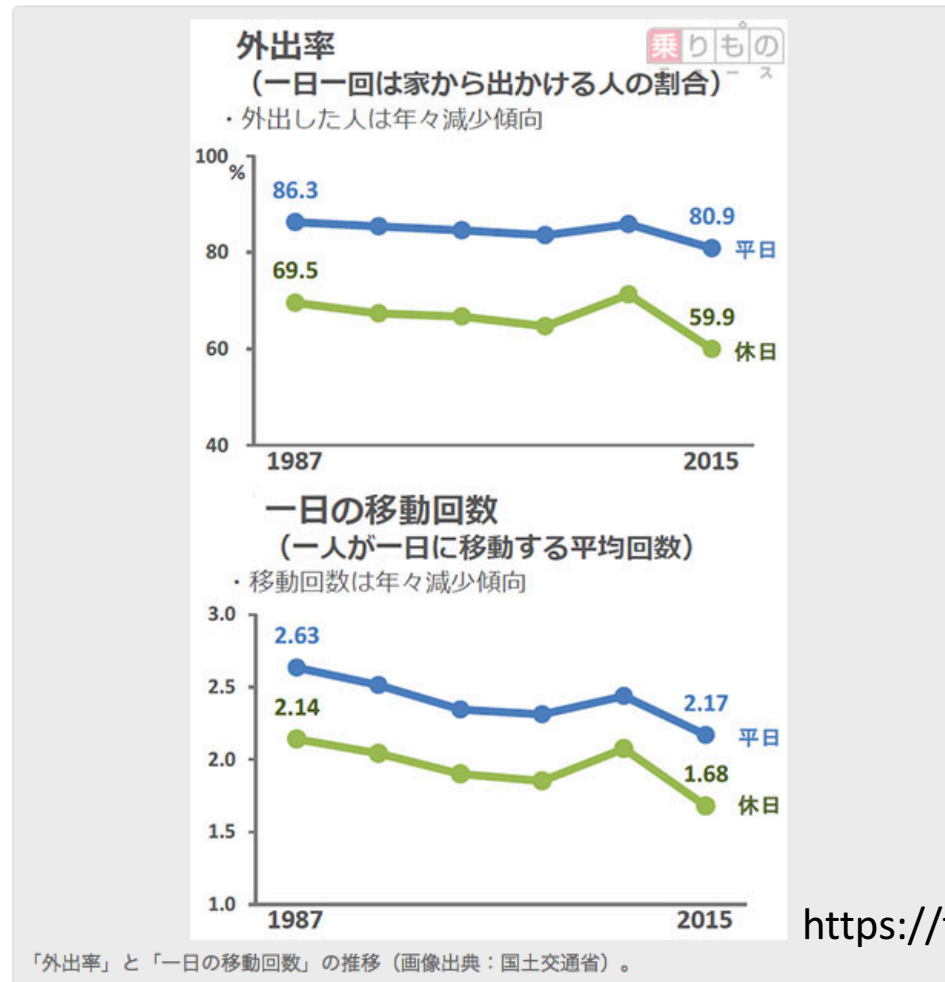
表 2.1: 調査対象都市

都市類型		調査対象都市
a	三大都市圏	中心都市
b		さいたま、千葉、東京区部、横浜、川崎、名古屋、京都、大阪、神戸
c		周辺都市
d	地方中枢都市圏	中心都市
e		周辺都市
f	地方中核都市圏	中心都市
g		周辺都市
h	地方中核都市圏	中心都市
i		周辺都市
j	地方中心都市圏	-

## 若者の「移動回数」、高齢者を下回る

国土交通省は2016年12月26日（月）、2015年度の「全国都市交通特性調査」（速報版）を公表しました。

人がどのような目的で、どのような交通手段を利用して移動しているかなど、人の動きをおおむね5年おきに調査するものです。今回は、調査日に外出した人の割合が平日で80.9%、休日で59.9%、ひとりが1日に移動する平均回数（移動回数）が平日で2.17回、休日で1.68回と、いずれも1987（昭和62）年の調査開始以来、最低の値でした。

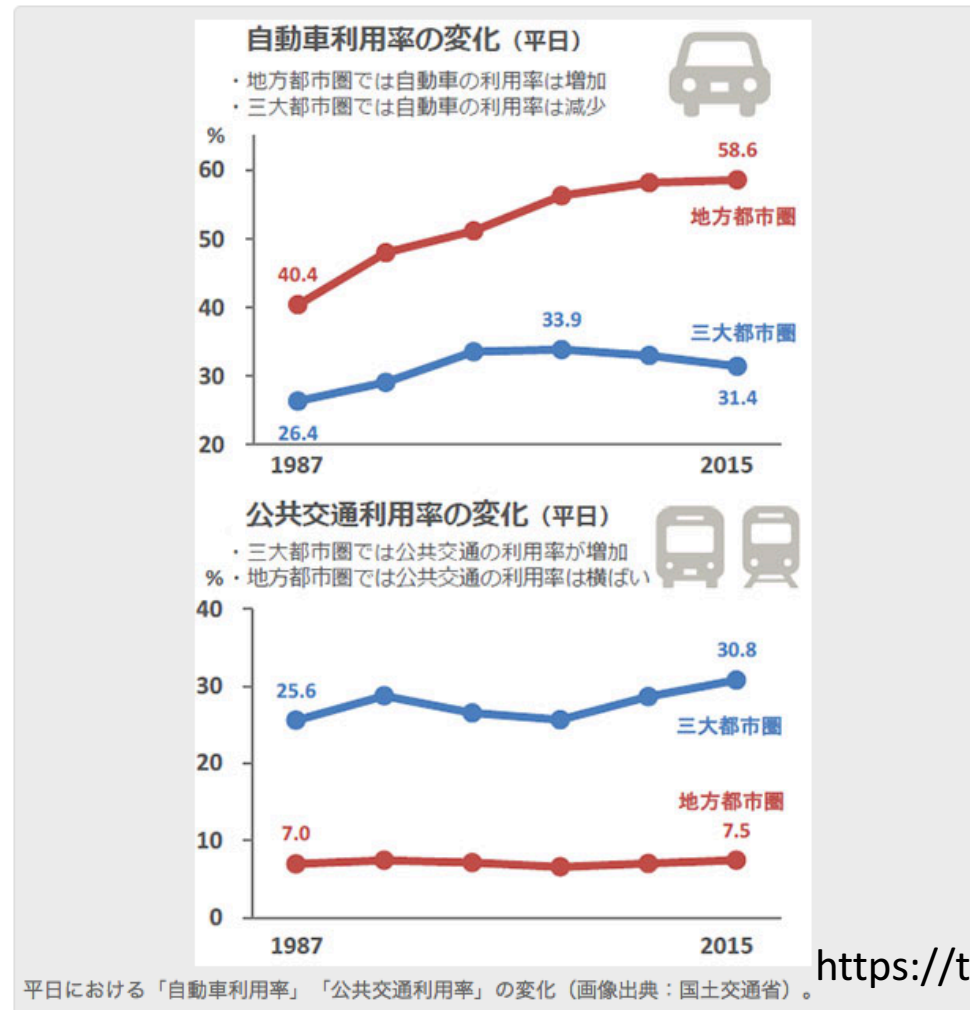


<https://trafficnews.jp/post/62155/>

若者の移動回数が減少し、高齢者の移動回数が増加しています。休日における20代の移動回数1.43回に対し、70代の移動回数はそれを上回る1.60回でした。特に20代の非就業者における外出率が大きく低下している一方で、人口が増加している60歳以上では就業者、非就業者ともに移動回数が増加していることがわかりました。

## 都市と地方で差が広がったもの、縮まったもの

交通手段の利用率については、三大都市圏では公共交通が増加し、自動車が増加。一方、地方都市圏では自動車が増加しているのに対し、公共交通はほぼ横ばいの傾向を示しました。ただ、全国平均で見ると自動車は平日、休日ともに5年前と大差なく推移しています。



ひとりが1日に動く回数の平均を目的別に見ると、通勤目的ではほぼ横ばい、通学や業務では年々減少しています。そして、休日における買い物、食事、娯楽といった私事目的では、1992（平成4）年と比べて約20%と、大きく減少しました。また、かつて三大都市圏よりも地方のほうが多かった移動回数は、いずれも減少傾向にあり、現在では地方都市圏2.18回、三大都市圏2.16回と差がほとんどなくなっています。

今回の調査は全国70都市の世帯が対象。郵送またはウェブで4万3700世帯が回答しました。国土交通省は調査結果の具体的な要因などについてさらなる分析を進める予定です。

坂東徹 (2013). 都市構造が個人の交通行動に及ぼす影響の分析.  
東京工業大学工学部土木・環境工学科卒業論文.

表 2.2: 質問項目

質問票	カテゴリ	質問項目
世帯票	世帯の情報	世帯人数・保有台数(自動車、貨物車、自動二輪、原付、自動車)
	個人の情報	年齢・性別・職業・就業形態・運転免許の有無・自由に使える自動車の有無
個人票	トリップ基本情報	出発地と目的地の場所・出発時刻と到着時刻・目的・使用交通手段・所要時間・距離
	自動車を使用した場合の追加情報	運転者・乗車人員・駐車した場所・有料道路の利用有無
都市交通に関する意識調査	交通の利用傾向	交通手段の利用頻度・交通手段を選ぶ際に重視する事・交通の満足度・最近1年間の交通手段の利用変化・交通が改善されたならば外出が増加するか
	日常生活について	目的ごとの外出についての情報・最寄のバス停や駅について・生活スタイル・転居の経験

表 2.3: 説明変数および被説明変数として用いるデータ

被説明変数	公共交通利用と自動車運転の年間比率	買い物トリップにおける片道距離(km) (対数値)	遠くても大きい店やいい店に行きたいかどうかの意識 (5段階指標)
データソース	交通手段の利用頻度	目的ごとの外出についての情報	生活スタイル
分析モデル	ロジスティック回帰	<del>線形回帰モデル</del>	<b>生存時間分析モデル</b>
説明変数	年齢・性別・労働者または学生の区分・家族人員		

# モデル構築の目的

- 都市構造, 個人のトリップ属性や社会経済属性が私事行動における活動範囲に及ぼす影響を分析する.
- より具体的には, 「どれぐらいの距離まで, 人々は買い物トリップを行うのか?」について, 生存時間分析を適用し, その規定要因について実証的に検討する.
- 「都市交通に関する意識調査票」の回答結果を利用する



秘

都市交通に関する意識調査票

この調査は、交通計画や、まちづくりを検討するための基礎資料を得ることを目的としています。調査票に記入された内容は、この目的以外には使用しません。

■ 記入についてお願い

- 黒の筆記用具で、18歳以上(平成22年10月1日現在)の方一人ひとりがご記入ください。
回答は、[ ]の部分(回答欄)に記入してください。
回答欄に1などの色がついた番号がある場合は、該当する番号に○をつけてください。

整理番号

問1 世帯票の2のあなたの個人番号を記入してください。 人目

問2 ふだんの外出で利用される交通手段についてお聞きします。

【記入例】週に5日、通勤で鉄道を利用する場合(往復は2回と数えて下さい) 1年に2月に3週に10回程度

(1) 以下の①～⑥の交通手段を、それぞれどのくらい利用していますか。

① 鉄道を利用することは ② 路線バスを利用することは ③ 自動車を運転することは
④ 自動車に同乗することは ⑤ 自転車を利用することは ⑥ 徒歩だけの外出は

(2) 以下の①～④の目的ごとに、外出する頻度・主な交通手段・片道の所要時間と距離についてお答えください。

Table with columns: 移動目的, 外出の有無, 外出する頻度, 最もよく行く目的地までの主な交通手段, 最もよく行く目的地までの片道の所要時間と距離

(3) ふだん移動するときの交通手段(鉄道・バス・自動車・自転車・徒歩など)を選ぶ際に、以下の①～⑧のそれぞれについてどの程度重視していますか。当てはまる番号に○をつけてください。

Table with columns: 全く重視していない, あまり重視していない, どちらでもない, やや重視している, とても重視している

(4) 最近1年間でふだん移動するときの交通手段の利用頻度は変化しましたか。当てはまる番号に○をつけてください。また、利用頻度が変わった場合は、その理由について表Bから当てはまるものを2つ以内選んで番号をご記入ください。

利用頻度の変化 (減った, 少し減った, 変わらない, 少し増えた, 増えた)
変化の理由 (表Bから2つ以内選択)

表B 変化の理由
1. 就職・転勤・退職・入学・卒業などで通勤や通学が変わった
2. 結婚・誕生・独立などで世帯人数が増加、または減少した
...

問3 お住まいの最寄りのバス停や鉄道駅についてお聞きします。当てはまる番号に○をつけてください。

最寄りのバス停について
① バス停までの徒歩での所要時間は?
② 1時間あたりのバスの運行本数は?
最寄りの鉄道駅について
③ 鉄道駅までの主な交通手段は?
④ ③で回答した交通手段での鉄道駅までの所要時間は?
⑤ 1時間あたりの鉄道の運行本数は?

問4 徒歩や車いす等で無理なく休まずに移動できる距離はどの程度ですか。当てはまる番号に○をつけてください。

1 100m まで 2 300m まで 3 500m まで 4 1km まで 5 1.5km まで 6 1.5km 以上

問5 あなたの生活スタイルについてお聞きします。当てはまる番号に○をつけてください。

Table with columns: 全く当てはまらない, あまり当てはまらない, どちらでもない, やや当てはまる, とても当てはまる

問6 交通手段についてのあなたの満足度をお聞きします。当てはまる番号に○をつけてください。

Table with columns: 不満, やや不満, どちらでもない, やや満足, 満足

問7 交通が改善されたなら今よりも外出を増やすと思うことはありますか。当てはまる番号に○をつけてください。

選択肢 1, 2 を選択した方にお聞きします。
どのようにすれば外出を増やすと思いますか。
下記 1 ~ 14 の中から当てはまるものを3つ以内選んで○をつけてください。
鉄道やバスを利用する際...
自動車を利用する際...
徒歩や自転車を利用する際...

裏面にもご記入下さい

**問 8 自動車運転免許の保有と自動車運転の状況についてお聞きします。当てはまる番号に○をつけてください。**



**問 8-1 現在自動車運転免許をお持ちの方にお聞きします。**

- (1) あなたご自身が最近1年間で自動車を運転した距離は何kmですか。 約  km
- (2) 自動車での移動をできるだけ控えてみようという気持ちをお持ちですか。

- 1 全くそう思わない  
2 少しそう思う  
3 そう思う  
4 とてもそう思う

選択肢 2, 3, 4 を選択した方にお聞きします。

自動車での移動をできるだけ控えてみようという理由は何ですか。  
(当てはまるもの全てに○をつけてください。)

- 1 自動車は、環境によく影響を及ぼすから  
2 自動車をたくさん使うと、お金がかかるから  
3 自動車は、健康・ダイエットにあまりよくないから  
4 自動車では、交通事故が心配だから

- (3) あなたが自動車を運転する理由について、以下の①～⑩のそれぞれについて当てはまる番号に○をつけてください。

	全く 当てはまらない	あまり 当てはまらない	どちら でもない	やや 当てはまる	とても 当てはまる
① 運転することが好き	← 1	2	3	4	5 →
② 好きなお店へ行ける	← 1	2	3	4	5 →
③ 公共交通より安上がりである	← 1	2	3	4	5 →
④ 多くの人や荷物をのせることができる	← 1	2	3	4	5 →
⑤ 自動車なら所要時間が短い	← 1	2	3	4	5 →
⑥ 業務が必要だから	← 1	2	3	4	5 →
⑦ 他に交通手段がない	← 1	2	3	4	5 →
⑧ 自動車に乗ることは自己表現の一つである	← 1	2	3	4	5 →
⑨ 好きなきに使える	← 1	2	3	4	5 →
⑩ プライベートな空間を確保できる	← 1	2	3	4	5 →

- (4) 自動車運転免許を、次回更新時期に更新しますか。当てはまる番号に○をつけてください。

1 更新する      2 どちらともいえない      3 更新しない

- (5) ご自身で自動車を保有していない方にお聞きします。保有していない理由は何ですか。  
右下の **表 C** から当てはまるものを2つ以内選んで番号をご記入ください。

保有していない理由

(保有している方は無記入で結構です)

**問 8-2 自動車運転免許をもっているが、ほとんど運転しない方にお聞きします。**

- (1) 運転しなくなったのはいつ頃ですか。(1年未満の方は「0年前」とご記入ください。)

1 約  年前から運転しなくなった      2 免許取得時から運転していない

- (2) 運転しない理由は何ですか。

右下の **表 C** から当てはまるものを2つ以内選んで番号をご記入ください。

運転しない理由

- (3) 運転しなくなってから移動はどのように変化しましたか。

右下の **表 D** から当てはまるものを2つ以内選んで番号をご記入ください。

移動の変化

**問 8-3 過去に自動車運転免許を持っていて、現在持っていない方にお聞きします。**

- (1) 免許を持たなくなったのはいつ頃ですか。 約  年前
- (2) 免許を持っていない理由は何ですか。  
下の **表 C** から当てはまるものを2つ以内選んで番号をご記入ください。
- (3) 免許を持たなくなってから移動はどのように変化しましたか。  
下の **表 D** から当てはまるものを2つ以内選んで番号をご記入ください。

免許を持っていない理由

移動の変化

**問 9 転居の経験についてお聞きします。**

- (1) 転居の経験はありますか。

1 ある → 現在の場所いつからお住まいですか。 平成・昭和  年  月から

2 ない → 質問は終了です。ご協力ありがとうございました。

- (2) 現在のご自宅に転居する前にお住まいの場所はどこでしたか。

都道府県  市区町村  丁目

- (3) (2) の場所では、ふだんの外出で以下の交通手段をどのくらい利用していましたか。

【記入例】週に5日、通勤で鉄道を利用する場合  
(往復は2回と数えて下さい) 1年に2月に③週に 10 回程度

① 鉄道を利用することは 1年に2月に3週に <input type="text"/> 回程度	② 路線バスを利用することは 1年に2月に3週に <input type="text"/> 回程度	③ 自動車を運転することは 1年に2月に3週に <input type="text"/> 回程度
④ 自動車に同乗することは 1年に2月に3週に <input type="text"/> 回程度	⑤ 自転車を利用することは 1年に2月に3週に <input type="text"/> 回程度	⑥ 徒歩だけの外出は 1年に2月に3週に <input type="text"/> 回程度

- (4) 転居する際に、複数の候補を比べて選びましたか。

または決まった転居先(社宅への入居・家族の家への同居・継続した家への居住など)がありましたか。

1 複数の候補を比べて選んだ → (5)へお進み下さい。

2 転居先は決まっていた → 質問は終了です。ご協力ありがとうございました。

- (5) 現在のお住まいを選ぶ際に、住みたい地域として①～⑧をどの程度考慮しましたか。

当てはまる番号に○をつけてください。

	全く 考慮していない	あまり 考慮していない	どちら でもない	少し 考慮した	よく 考慮した
① 都心・副都心やそれに近接する地域	← 1	2	3	4	5 →
② 商店街やそれに近接する地域	← 1	2	3	4	5 →
③ 郊外のショッピングセンターへ行きやすい地域	← 1	2	3	4	5 →
④ 緑や自然が豊かな地域	← 1	2	3	4	5 →
⑤ 駐車場の確保がしやすい地域	← 1	2	3	4	5 →
⑥ 鉄道・バスが便利な地域	← 1	2	3	4	5 →
⑦ 徒歩や自転車の移動範囲で日常生活の用事が済む地域	← 1	2	3	4	5 →
⑧ 自動車で移動しやすい地域	← 1	2	3	4	5 →

(ご協力ありがとうございました)

**表 C 自動車を保有していない理由・運転しない理由・免許を持っていない理由**

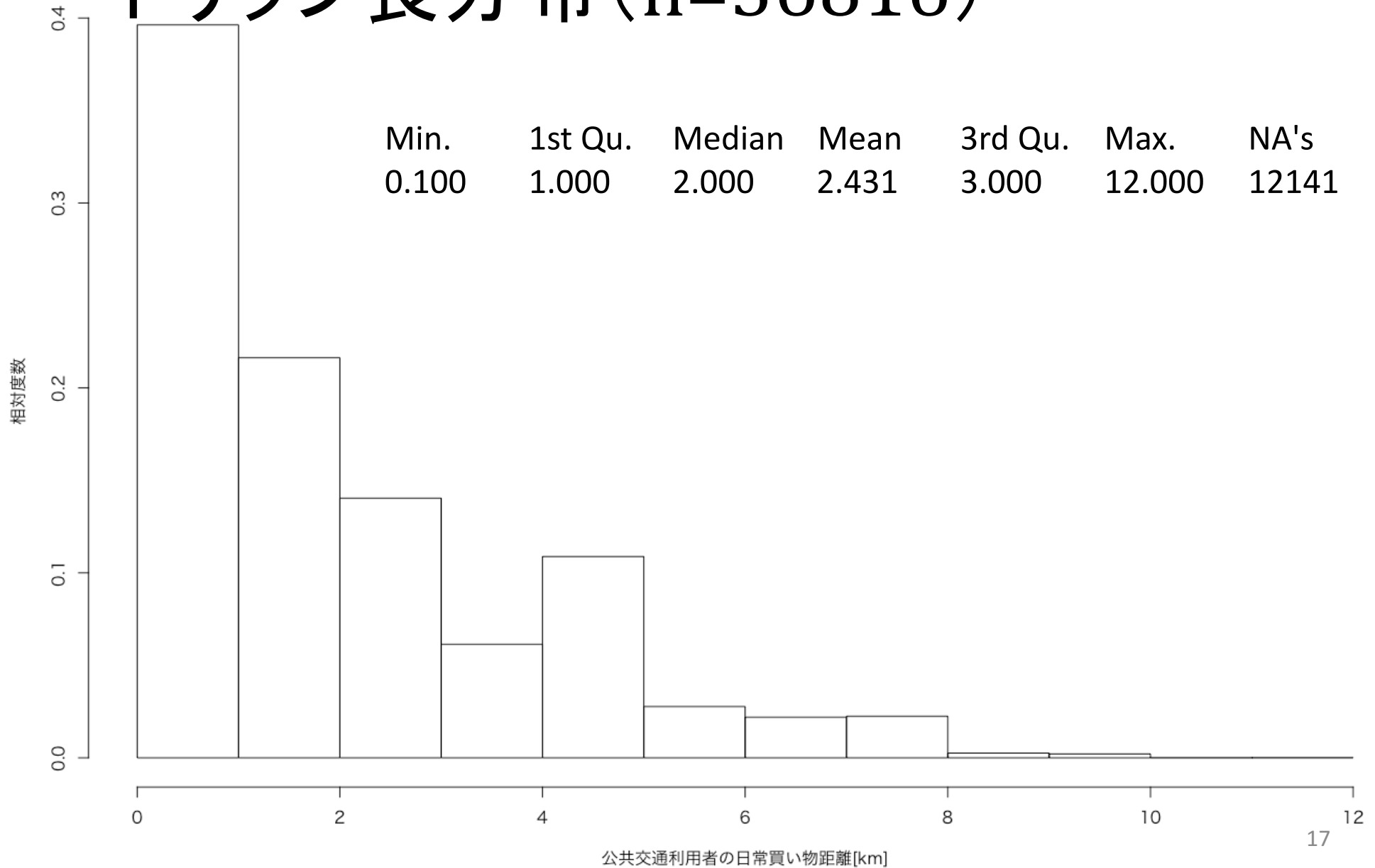
1. 自動車を利用する必要がないから  
2. 公共交通が便利だから  
3. 自分の運転に自信がないから  
4. 自動車の維持費用が高いから  
5. 自動車の利用は環境によく影響を及ぼすと思うから  
6. 歩いた方が健康にいいから  
7. 交通事故が心配だから  
8. 自動車に乗るのが好きではないから  
9. その他

**表 D 移動の変化**

1. 鉄道やバスなどの公共交通機関を利用している  
2. 徒歩や自転車で出かけられる範囲で移動している  
3. 他の人が運転する自動車に乗車するようになった  
4. 外出することが減った

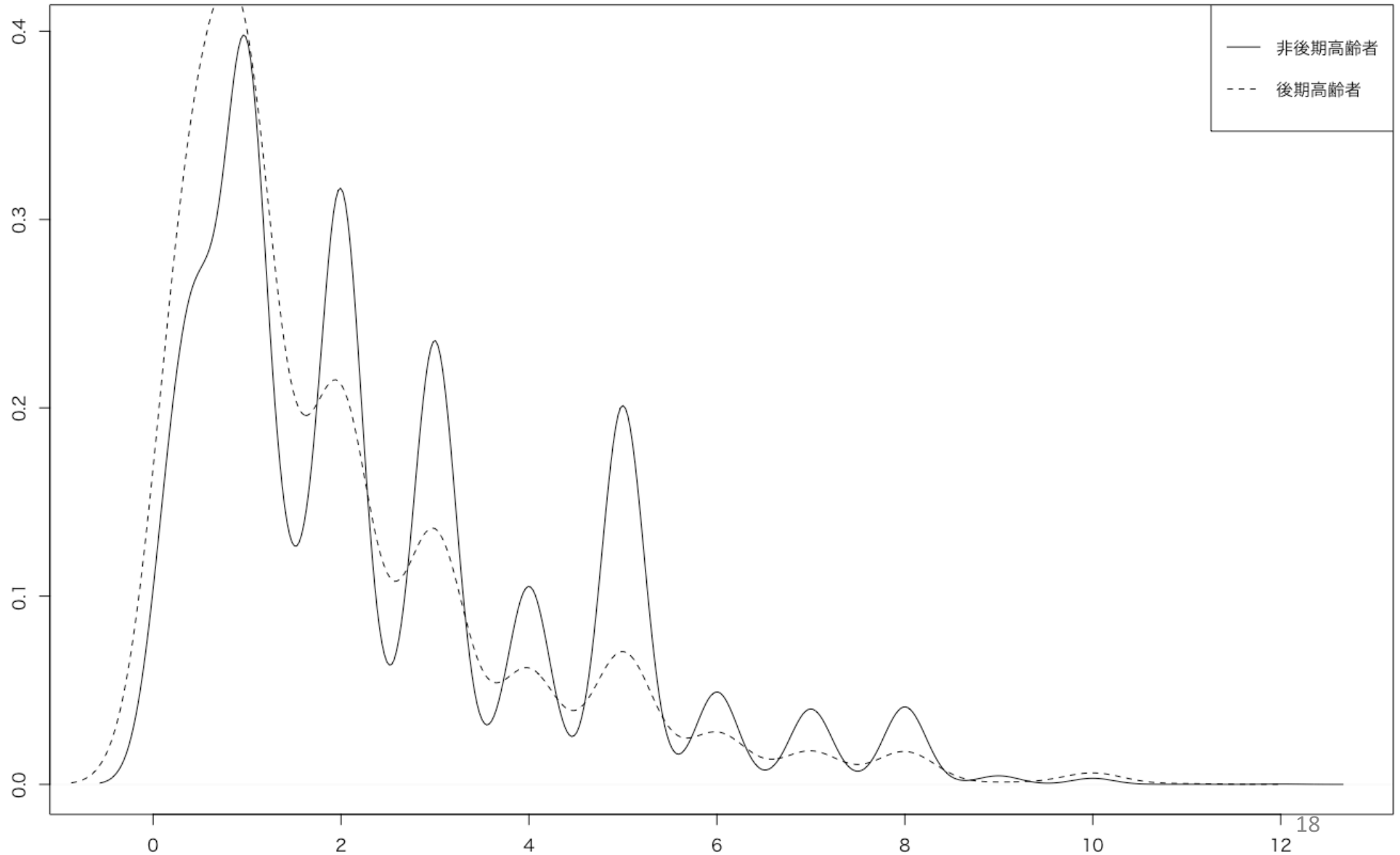


# 公共交通利用者の日常買い物目的 トリップ長分布 (n=36816)

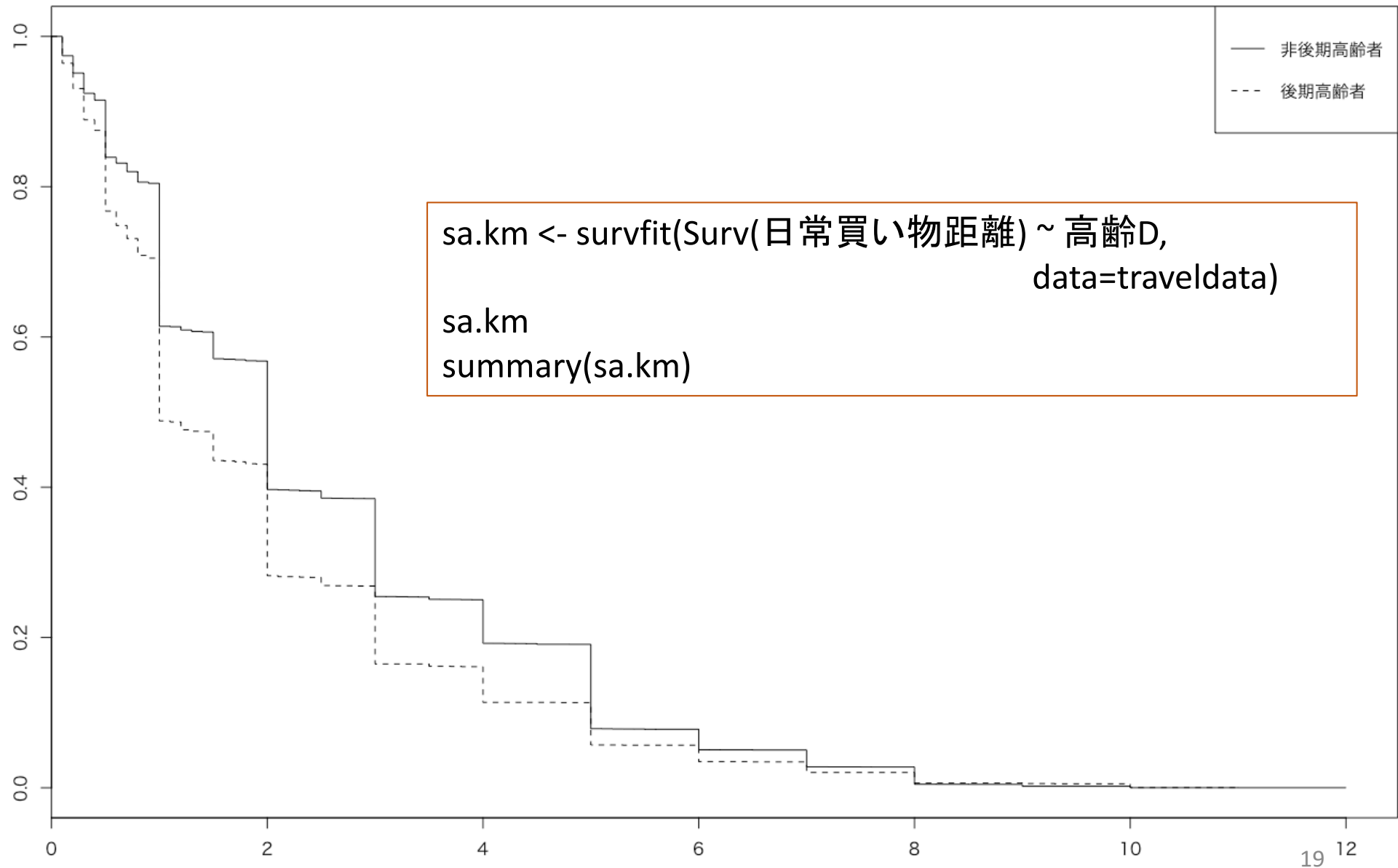


# 後期高齢者と非後期高齢者のトリップ長分布 (経験分布)

日常買い物距離の経験分布



# Kaplan-Meier法により後期高齢者と非後期高齢者のトリップ長分布を求める



# トリップ長を決める要因は多数ある →Semi-parametric/Parametric models

(Cox) 比例ハザードモデル [セミパラメトリックモデル]

$$h(t|\mathbf{x}) = h_0(t) \exp(\beta^t \mathbf{x}) = h_0(t) \exp \left( \sum_{i=1}^M \beta_i x_i \right)$$

“ベースラインハザード”  
ここはノンパラメトリック

“共変量”の値に依存して  
ハザードが変化する  
ここはパラメトリック

ハザードの比率が時間  $t$  に依存せず常に一定となる, 強い仮定を置いたモデル.

どのような目的関数を最大化して最適パラメータ  $\beta$  を求めているかについては各自教科書等で確認すること.

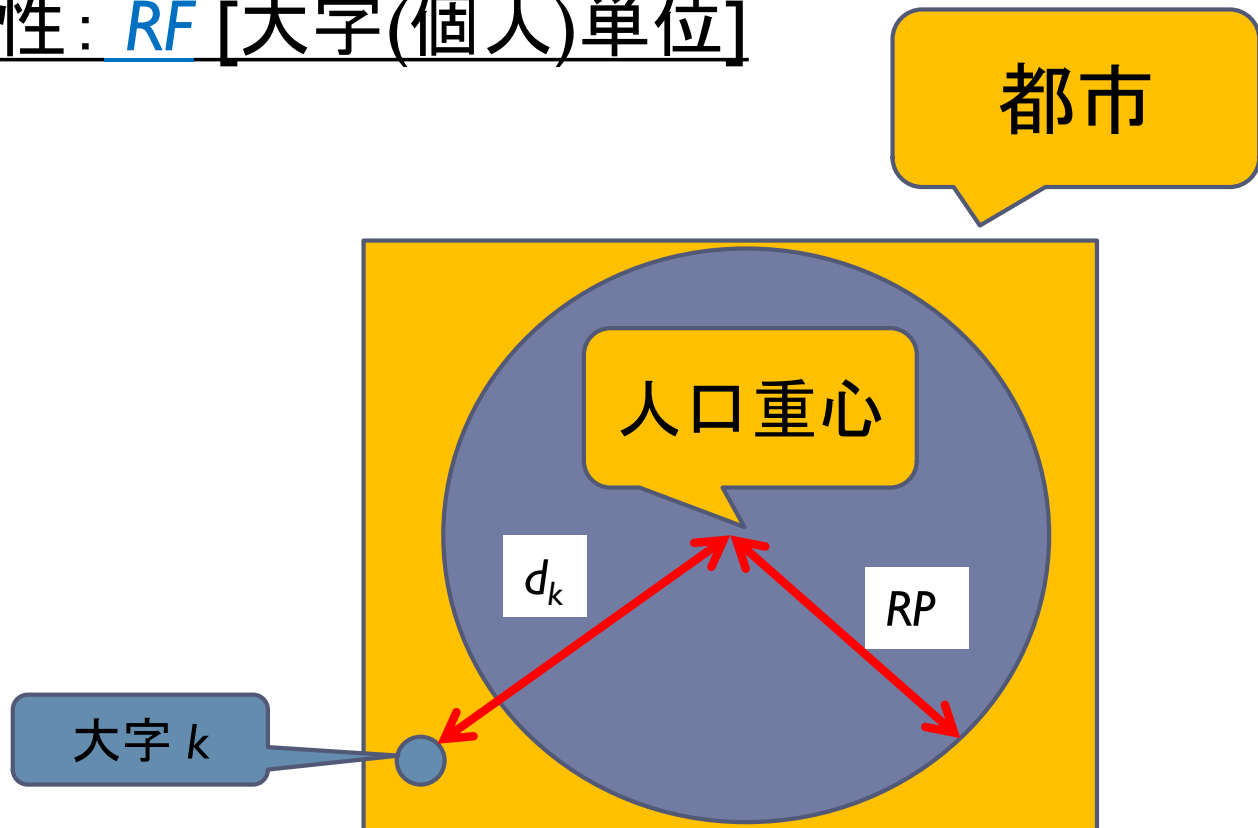
	coef	exp(coef)	se(coef)	z	Pr(> z )	
年齢	1.596e-03	1.002e+00	5.667e-04	2.816	0.004870	**
男性	-2.738e-01	7.605e-01	1.514e-02	-18.082	< 2e-16	***
職業学生	-9.109e-02	9.129e-01	1.698e-02	-5.365	8.11e-08	***
バス時間	-1.812e-02	9.820e-01	8.612e-03	-2.104	0.035388	*
バス本数	2.783e-02	1.028e+00	8.820e-03	3.155	0.001605	**
駅時間	-4.766e-02	9.535e-01	6.199e-03	-7.690	1.48e-14	***
駅本数	5.369e-02	1.055e+00	1.031e-02	5.208	1.91e-07	***
家族人員	-4.161e-02	9.592e-01	5.740e-03	-7.249	4.20e-13	***
市街化区域	-3.419e-01	7.104e-01	2.418e-02	-14.142	< 2e-16	***
政令指定都市	-4.953e-02	9.517e-01	2.115e-02	-2.342	0.019198	*
首都圏	-5.338e-02	9.480e-01	2.949e-02	-1.810	0.070265	.
中京圏	-2.307e-01	7.940e-01	2.778e-02	-8.306	< 2e-16	***
関西圏	-1.590e-01	8.530e-01	2.385e-02	-6.669	2.58e-11	***
DID人口密度	-6.667e-06	1.000e+00	7.555e-06	-0.882	0.377510	
都市化度	1.994e-03	1.002e+00	5.919e-04	3.368	0.000756	***
市街化度	-3.489e-03	9.965e-01	7.075e-04	-4.932	8.13e-07	***
小売業密度	7.641e-04	1.001e+00	8.052e-04	0.949	0.342639	
CP	1.579e-01	1.171e+00	3.384e-02	4.666	3.08e-06	***
RF	-1.417e-01	8.679e-01	1.862e-02	-7.610	2.74e-14	***
CI2	4.707e-05	1.000e+00	4.812e-06	9.783	< 2e-16	***
---						
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						

# コンパクトシティ指標の作成

1. 人口二次半径:  $RP$  (km) [都市単位]

2. 相対的遠距離性:  $RF$  [大字(個人)単位]

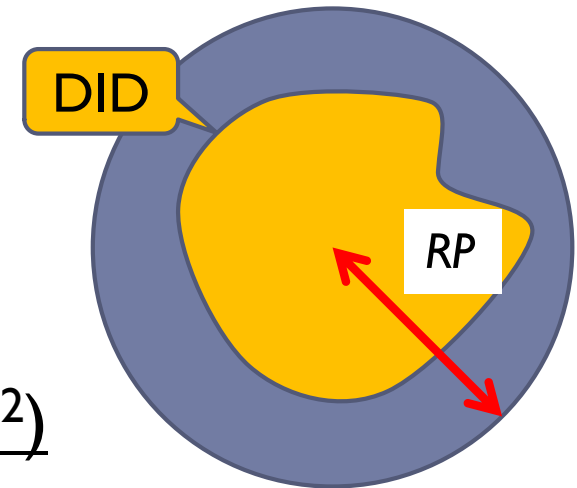
$$RF = \frac{d_k}{RP}$$



# コンパクトシティ指標の作成

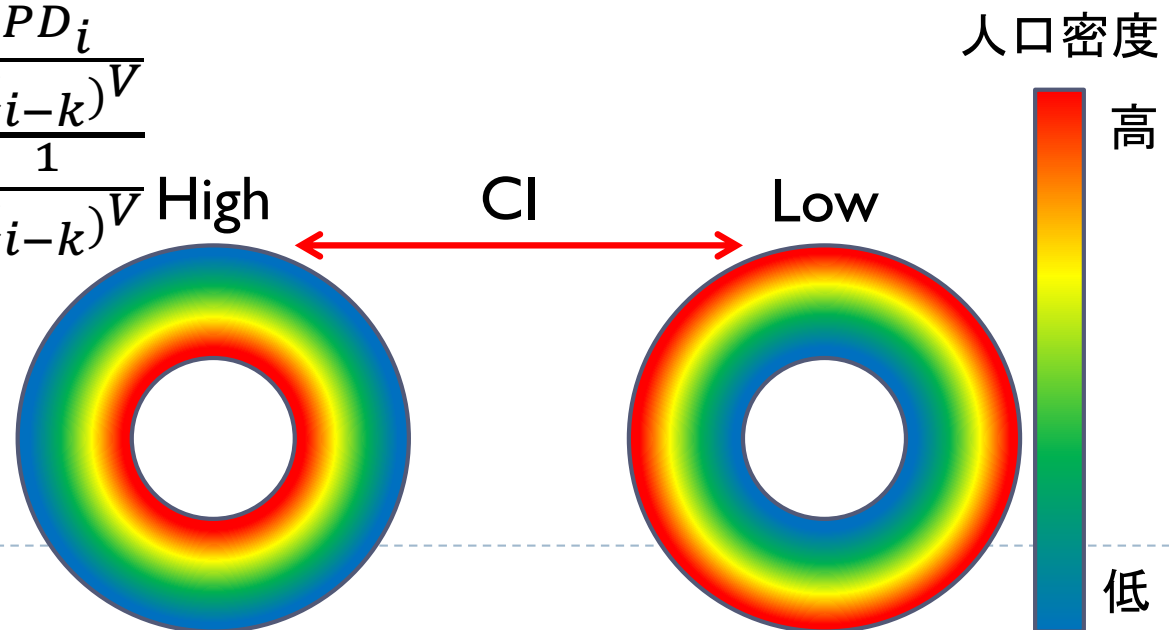
## 3. 人口集中度: CP [都市単位]

$$CP = \frac{(DID \text{ area})}{\pi RP^2}$$

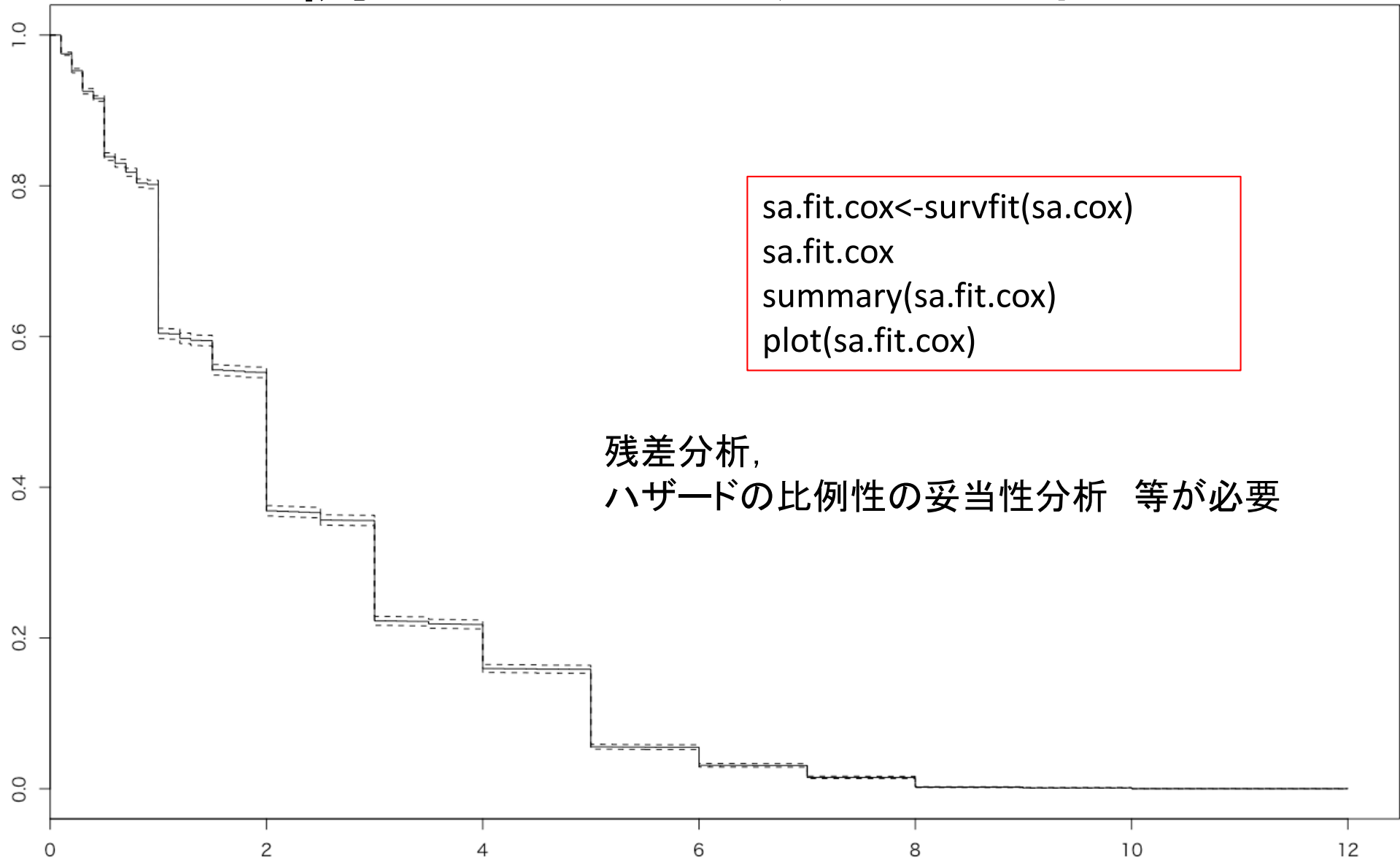


## 4. 他地域人口密度集積性: CI (people/km<sup>2</sup>) [大字(個人)単位]

$$CI_k^V = \frac{\sum_{i \neq k}^N \frac{PD_i}{(d_{i-k})^V}}{\sum_{i \neq k}^N \frac{1}{(d_{i-k})^V}} \text{ High}$$

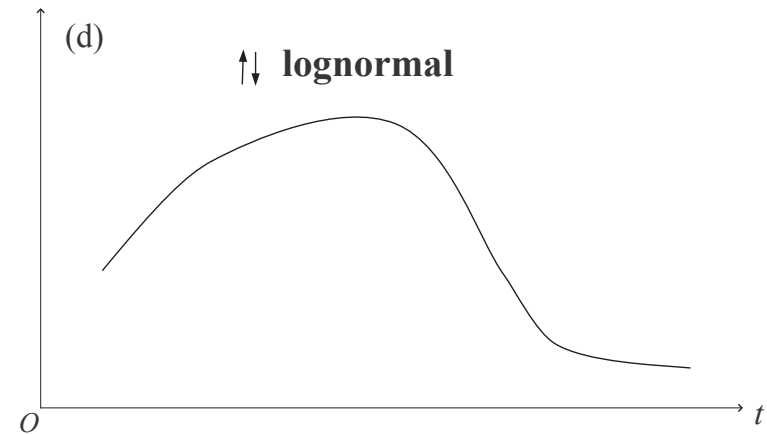
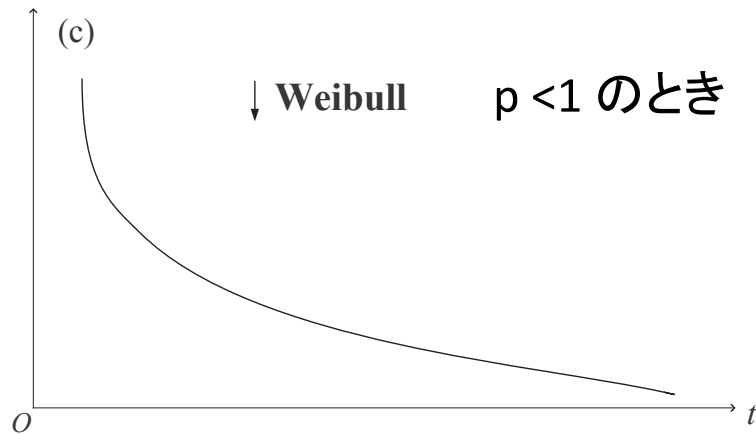
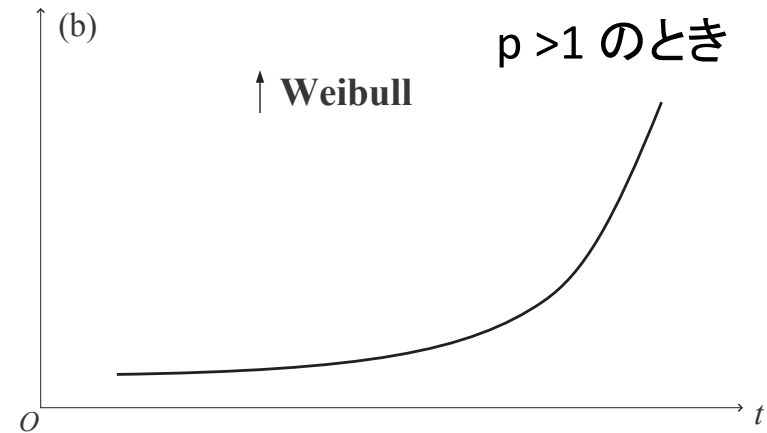
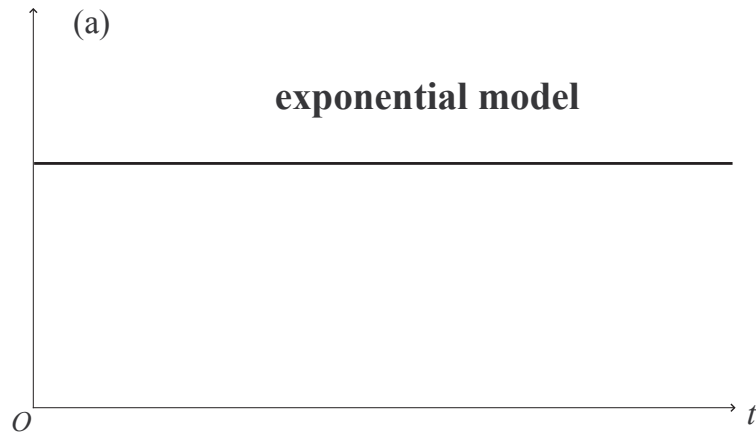


# Cox比例ハザードモデルの当てはめ





# いろいろなハザード関数 $h(t)$



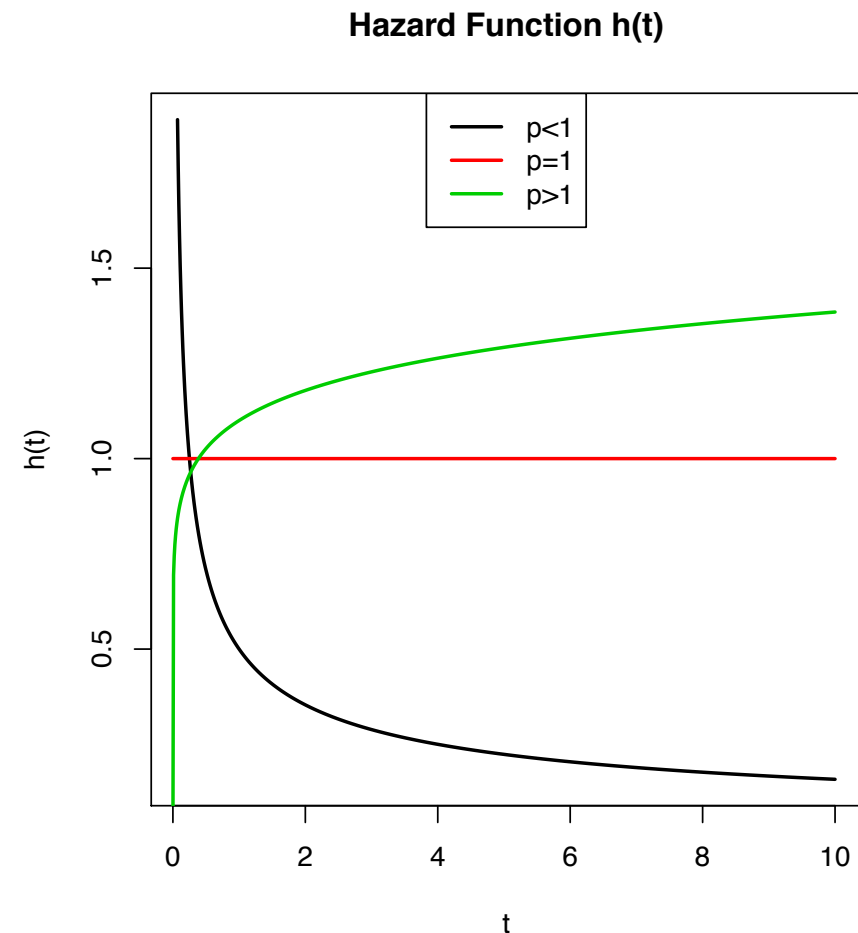
Distribution	$f(t)$	$S(t)$	$h(t)$
Exponential	$\lambda \exp(-\lambda t)$	$\exp(-\lambda t)$	$\lambda$
Weibull	$\lambda p t^{p-1} \exp(-\lambda t^p)$	$\exp(-\lambda t^p)$	$\lambda p t^{p-1}$
Log-logistic	$\frac{\lambda p t^{p-1}}{(1+\lambda t^p)^2}$	$\frac{1}{1+\lambda t^p}$	$\frac{\lambda p t^{p-1}}{1+\lambda t^p}$

# Weibull型 Hazard-Function

$$h(t) = \lambda p t^{p-1}.$$

$p$  is called **shape parameter**:

- ▶ If  $p > 1$  the hazard increases
- ▶ If  $p = 1$  the hazard is constant (exponential model)
- ▶ If  $p < 1$  the hazard decreases



# パラメトリックモデルの推定

ワイブル分布のハザード関数:

$$h(t) = \lambda p t^{p-1}$$

ワイブル加速時間故障モデル:

$$\frac{1}{\lambda^{1/p}} \equiv \exp(\beta_0 + \beta_1 x_{1n} + \dots + \beta_M x_{Mn})$$

```
#パラメトリックモデル
```

```
sa.weib <- survreg(Surv(日常買い物距離) ~  
  年齢+男性+職業学生+バス時間+バス本数+駅時間+駅本数+家族人員+  
  市街化区域+政令指定都市+首都圏+中京圏+関西圏+DID人口密度+  
  都市化度+市街化度+小売業密度+CP+RF+CI2,  
  data=traveldata,  
  dist="weibull")  
summary(sa.weib)
```

指数分布 → "exponential"  
対数正規分布 → "log-normal"  
対数ロジスティック分布 → "log-logistic"

	Value	Std. Error	z	p
(Intercept)	6.69e-01	5.01e-02	13.37	9.47e-41
年齢	-1.43e-03	4.38e-04	-3.26	1.10e-03
男性	2.23e-01	1.16e-02	19.18	4.98e-82
職業学生	7.38e-02	1.31e-02	5.64	1.69e-08
バス時間	1.29e-02	6.65e-03	1.94	5.25e-02
バス本数	-2.14e-02	6.82e-03	-3.14	1.72e-03
駅時間	3.74e-02	4.78e-03	7.81	5.65e-15
駅本数	-4.04e-02	7.96e-03	-5.08	3.82e-07
家族人員	3.27e-02	4.43e-03	7.38	1.57e-13
市街化区域	2.65e-01	1.86e-02	14.27	3.37e-46
政令指定都市	3.78e-02	1.63e-02	2.32	2.05e-02
首都圏	3.41e-02	2.27e-02	1.50	1.34e-01
中京圏	1.87e-01	2.14e-02	8.72	2.70e-18
関西圏	1.25e-01	1.84e-02	6.80	1.07e-11
DID人口密度	5.91e-06	5.84e-06	1.01	3.12e-01
都市化度	-1.50e-03	4.57e-04	-3.28	1.03e-03
市街化度	2.94e-03	5.46e-04	5.38	7.47e-08
小売業密度	-9.36e-04	6.22e-04	-1.51	1.32e-01
CP	-1.22e-01	2.61e-02	-4.67	3.04e-06
RF	1.04e-01	1.43e-02	7.26	3.97e-13
CI2	-3.88e-05	3.70e-06	-10.48	1.09e-25
Log(scale)	-2.59e-01	5.50e-03	-47.10	0.00e+00

Scale= 0.772

Weibull distribution

Loglik(model)= -34733.2    Loglik(intercept only)= -36166

Chisq= 2865.61 on 20 degrees of freedom, p= 0

Number of Newton-Raphson Iterations: 5

n=19548 (17268 observations deleted due to missingness)

Cox比例ハザードモデルの場合と多くの共変量の係数の符号が逆になっていることに注意

注意

# survreg's scale = 1/(rweibull shape)  
# survreg's intercept = log(rweibull scale)  
であるので,

$$\hat{\alpha} = 1/0.772 = 1.295$$

$$\hat{\lambda} = \exp(0.669) = 1.952$$

となる

どの確率分布を用いるのがベストかに関しては、  
**AIC = -2\*最大対数尤度 + 2\*パラメタ数**  
を用いて評価することができる

表5-3 保有期間モデルの推定結果

		重み付き モデルA		重みなし モデルB		t 値**
		推定値	t 値	推定値	t 値	
定数項		4.348		4.821		-2.27
自動車属性						
保有形態	リース車	-0.462	-4.63	-0.498	-3.19	0.19
	社用車	-0.625	-5.46	-0.402	-1.78	-0.88
購入時	中古車	-0.531	-15.77	-0.619	-12.55	1.48
車種	Mini, Subcompact	0.169	4.22	0.131	1.73	0.44
	Large, Luxury	0.132	2.85	0.119	1.32	0.13
	Sport car	0.288	5.75	0.281	2.94	0.06
	Pick-up truck	0.057	1.22	0.046	0.51	0.11
	Van	0.008	0.13	-0.004	-0.03	0.09
	Utility vehicle	-0.070	-1.16	-0.098	-0.85	0.22
年間走行距離/10,000 mile		-0.387	-3.52	-0.546	-6.27	1.13
世帯属性						
世帯構成	夫婦と子供 (0-15才)	0.089	1.56	0.119	1.36	-0.29
	夫婦と子供 (16-20才)	0.152	2.23	0.213	1.86	-0.46
	夫婦と子供とその他の大人	-0.012	-0.13	-0.024	-0.16	0.07
	独身者	0.043	0.80	0.265	3.05	-2.18
	夫婦とその他の大人	0.099	1.75	0.128	1.40	-0.27
子供数		-0.026	-1.13	-0.051	-1.46	0.60
大人数		-0.088	-2.84	-0.110	-2.14	0.38
賃貸		-0.113	-3.12	-0.186	-3.40	1.11
自動車保有台数		0.167	8.81	0.322	9.16	-3.87
運転者属性						
年齢/100才		1.697	11.78	1.702	7.28	-0.02
性別	女性	-0.100	-3.22	-0.080	-1.60	-0.34
職業	営業	-0.528	-4.00	-0.698	-3.99	0.78
	自営業	-0.189	-1.43	-0.252	-1.27	0.27
	退職者	-0.092	-1.51	-0.028	-0.29	-0.54
個人年収	\$20,000未満	0.048	1.30	0.098	1.65	-0.70
	\$100,000以上	-0.517	-6.24	-0.546	-4.27	0.19
形状パラメータ(1/γ)		0.685	10.55*	0.664	20.45*	0.63
サンプル数		8143		8143		
L(C)		-6377		-2973		
L(β)		-5883		-2706		
χ <sup>2</sup> (df)		988(26)		536(26)		

\*H<sub>0</sub>: γ=1 に対する t 値, \*\*H<sub>0</sub>: β<sub>a</sub>=β<sub>b</sub> に対する t 値(β<sub>a</sub> はモデル A による推定値, β<sub>b</sub> はモデル B による推定値)

# 自動車保有 期間モデル (山本先生D 論第五章)

ワイブル分布を想定

Rではないので  
符号は直接  
解釈してよい