

## SPパネル調査の調査期間決定のための基礎分析\*

### Determining the Duration of Survey Period for SP Panel Survey \*

張 峻屹\*\*・杉恵頼寧\*\*\*・藤原章正\*\*\*・奥村 誠\*\*

By Junyi ZHANG\*\*, Yoriyasu SUGIE \*\*\*, Akimasa FUJIWARA \*\*\* and Makoto OKUMURA\*\*

#### 1. はじめに

同一個人を複数時点にわたって追跡するパネル調査は個人の交通行動の時間的な変化をより正確に捉える手法として、1980年代以降盛んに調査、研究されている。これによって、数多くの成果が得られており、今後の交通計画のための有力なツールとして期待されている。しかし、パネル調査を普及し、実用化させるにはまだいくつかの課題が残されている。パネル調査の期間（インタビューの頻度）を一体何年（何回）実施すればよいかはその一つである。

パネル調査回数に影響する要因として、調査側のものと被調査側（個人）のものに分けられる。調査側の要因として主に費用制約、調査票の設計技術、設問の数などである。限られた費用のもとで最小限の設問数で、最大限の情報を得るために調査票をいかにうまく設計し、回答者を引きつけるかは無視できない問題である。最近、コンピュータをベースとした調査の狙いの一つはここにある。また、被調査側の個人はその社会経済属性、性格、態度などによって、調査への参加意欲が異なる。さらに、調査回数を重ねていくにつれて、疲れなどによる回答バイアス、消耗バイアスが存在する。しかも、これらのバイアスは時間とともに大きくなる可能性がある。このように、調査期間とデータの信頼性との間にはトレードオフが存在する。

そこで、本研究では当研究室が実施した広島市の新交通システムの利用意識に関する5時点SPパネルデータを用いて、被調査側の立場から適正なパネル調査期間について検討する。ある個人が初期調査に参加

してからいつまで調査に参加し続けてくれるかという問題は、経済学で扱う失業期間や医学における病人の生存時間などと似通う性質をもつため、これらの時間を取り扱う期間モデルを採用する。

#### 2. 期間モデルに関する概説<sup>1)</sup>

期間モデルは観測時点からある事象が発生するまでの期間（生存時間）を分析する際に用いられる。生存関数 $S(t)$ （確率変数 $T$ がある時点 $t$ を越える確率）とハザード関数 $h(t)$ （ $t \leq T$ の条件下で時点 $t$ の瞬間にある事象が起こる確率）によって表される。期間 $t$ の確率密度関数を $f(t)$ 、その分布関数を $F(t)$ とすると、以下のような関係式が成立する。

$$F(t) = \text{Prob}(T \leq t) = \int_0^t f(u) du \quad (1)$$

$$S(t) = \text{Prob}(T \geq t) = 1 - F(t) \quad (2)$$

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\text{Prob}(t \leq T \leq t + \Delta t \mid T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} \\ = -\frac{d \ln(S(t))}{dt} \quad (3)$$

以上の各式に示されるように、これらの関数はどれか1つが定まれば、残りの関数も自動的に決まるという意味で数学的には等価なものである。

また、期間 $t$ に影響する要因を取り入れる方法として比例ハザードモデルと加速寿命モデルがある。両者はそれぞれ式(4)、(5)によって表される。

$$h(t \mid \mathbf{X}) = h_0(t) \exp(\beta \mathbf{X}) \quad (4)$$

$$h(t \mid \mathbf{X}) = h_0(t \exp(\beta \mathbf{X})) \exp(\beta \mathbf{X}) \quad (5)$$

ただし、 $h_0$ は全ての説明要因が0の時のハザード関数であり、基準ハザードと呼ばれる。 $\mathbf{X}$ は説明要因ベクトル、 $\beta$ はそのパラメータベクトルである。

比例ハザードモデルは時間に依存する項と説明要因に依存する項に分割できるため、応用しやすいという利点をもつ。しかし、例えば、自動車の保有期間

\* キーワード：交通調査、SP、パネルデータ、調査論

\*\* 正会員、工博、広島大学工学部第四類（建設系）  
（東広島市鏡山1-4-1、TEL & FAX: 0824-24-7849）

\*\*\* 正会員、工博、広島大学大学院国際協力研究科  
（東広島市鏡山1-4-1、TEL & FAX: 0824-24-7826）

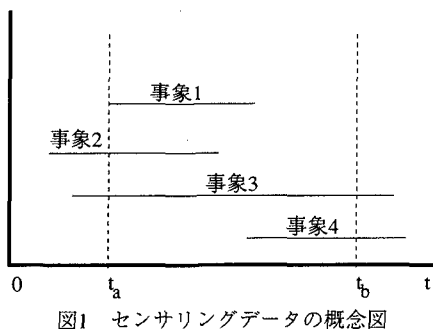


図1 センサリングデータ概念図

を考える場合、自動車保有者と非保有者のハザードの比率は時間的に変化しないという非現実的な結果になってしまう。これに対して、加速寿命モデルでは説明要因により基準ハザード関数の中において時間をrescaleするため、上述の問題点が存在しない。本研究では後者の加速寿命モデルを用いる。

基準ハザードに関する分布として、本研究では以下の3種類を取り上げて、個人の参加意思の継続期間（調査開始から継続参加しようとする時間の長さ）を表現する。

#### (1) 指数分布

$$f(t) = \gamma \exp(-\gamma t) \quad (6)$$

$$S(t) = \exp(-\gamma t) \quad (7)$$

$$h(t) = \gamma \quad (8)$$

#### (2) ワイブル分布

$$f(t) = \gamma \alpha t^{\alpha-1} \exp(-\gamma t^\alpha) \quad (9)$$

$$S(t) = \exp(-\gamma t^\alpha) \quad (10)$$

$$h(t) = \gamma \alpha t^{\alpha-1} \quad (11)$$

#### (3) 対数対数ロジスティック分布

$$f(t) = \gamma \alpha t^{\alpha-1} / (1 + \gamma t^\alpha)^2 \quad (12)$$

$$S(t) = 1 / (1 + \gamma t^\alpha) \quad (13)$$

$$h(t) = \gamma \alpha t^{\alpha-1} / (1 + \gamma t^\alpha) \quad (14)$$

ただし、 $\gamma$ 、 $\alpha$ は分布の尺度パラメータ、形状パラメータと呼ばれ、分布の形を決定する上で非常に重要なパラメータである。

そして、センサリングデータが存在するか否かによって、対数尤度関数の定式化が違ってくる。センサリングデータ概念図は図1に示されている。図1において、 $t_a$ 、 $t_b$ はそれぞれ調査の開始時刻と終了時刻を表す。事象1に関しては調査開始と同時に発生するため、それが観測できる。これに対して、事象2はいつ始まったかは分からないため、左側センサリ

ングの問題が存在する。事象3はいつ始まり、いつ終了するか両方が未知であるため、区間センサリングの問題を考慮する必要がある。事象4はいつ終了するかに関する情報が得られないため、右側センサリングが問題となる。

センサリング問題が存在しない場合に、加速寿命モデルの対数尤度関数は以下のように表される。

$$\begin{aligned} \ln L &= \sum_i \ln(h_i(t_i | X_i)) \\ &= \sum_i \{ \ln(h_{i,0}(t_i \exp(\beta X_i))) + \beta X_i \} \end{aligned} \quad (15)$$

ただし、 $i$ は個人を表す。

一方、データにセンサリング問題が存在する場合、対数尤度関数の定式化は複雑になる。

### 3. 新交通システムの利用意識に関する

#### 5時点パネル調査の概要

当研究室では1994年8月に開業する新交通システムの利用意識を明らかにするために、1987年から1994年の7年間に於いて予定建設ルートの沿線団地に住む通勤・通学者を対象に5回のSPパネル調査を実施した。新交通システムの代替手段として自家用車とバスを取り上げた。調査に参加した回答者数の時間的な変化状況を図2に示す。同図から分かるように、継続的に参加してくれる個人もいれば、いったん調査から離脱して再び復帰する個人もいる。さらに、離脱サンプルを補うために新規サンプルを時点ごとに追加している。このように、いろいろな参加パターンがこの5時点パネル調査に混在している。

まず、参加回数の分布を図3に示す。平均参加回数は2.1回である。

つぎに、個人のパネル調査への参加意思の継続期間の分布を図4に示す。例えば、5時点ともに継続参加した個人は1987年11月から1994年7月までの56ヵ月ずっと参加する意思があると考え、1987年11月に調査に参加し1990年11月に離脱して、1993年11月に調査に復帰し1994年7月に離脱する回答者に関しては、3時点で継続参加する意思があるケース（36ヵ月）と1時点しか参加する意思がないケース（8ヵ月）に分けて考える。参加意思の継続期間の平均値は18.5ヵ月である。

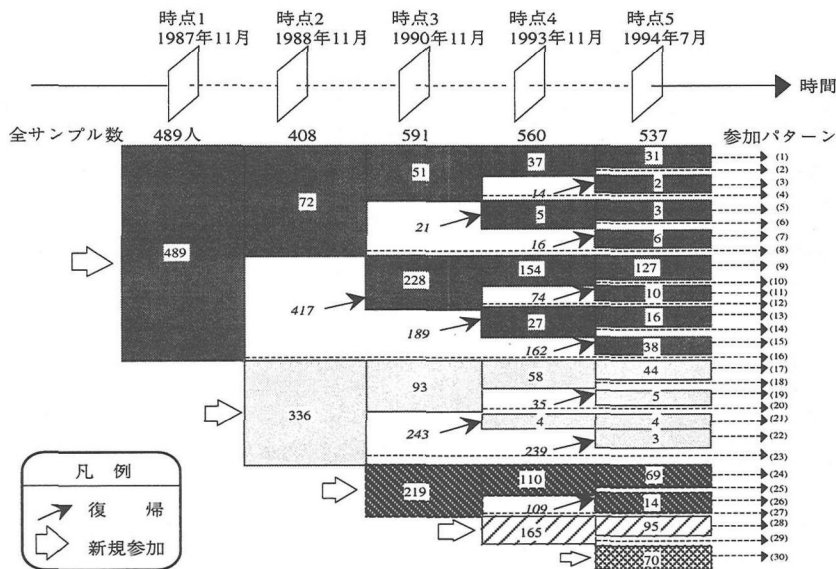


図2 5時点SPパネル調査に参加した回答者数

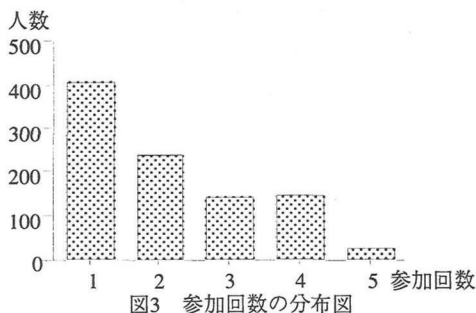


図3 参加回数の分布図

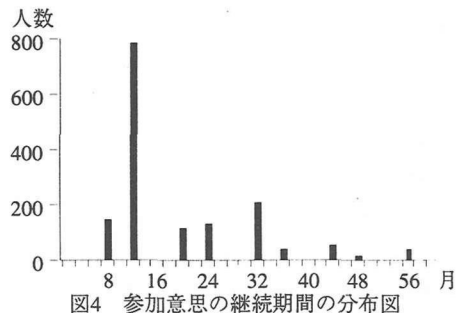


図4 参加意思の継続期間の分布図

#### 4. 参加回数に影響する要因の分析

被験者側から見てパネル調査の適正な実施期間を明らかにするために、まず被調査側の個人の参加回数に影響する要因を分析する必要がある。したがって、この節では参加回数に関する要因分析を行う。ここで、個人  $i$  が  $k_i$  回参加する確率  $P(k_i)$  がポアソン分布に従うと仮定する。

$$P(k_i) = \frac{(\lambda_i)^{k_i}}{k_i!} \exp(-\lambda_i) \quad (16)$$

$$\lambda_i = \exp(\beta X_i) \quad (17)$$

ただし、 $\lambda_i$  はポアソン分布のパラメータである。

参加回数に影響する要因  $X_i$  として、個人属性（性別：男性1、女性0；年齢、職業：有1、無0；免許：有1、無0；主要交通手段：自動車1、公共交通0；世帯人数）を考える。これらの要因の影響は式(17)で示

表1 参加回数モデルの推定結果

| 説明変数   | パラメータ値  | t 値    |
|--------|---------|--------|
| 性別     | 0.010   | 0.189  |
| 年齢     | 0.009   | 4.849  |
| 職業     | 0.229   | 2.515  |
| 免許有無   | 0.133   | 1.957  |
| 主要交通手段 | -0.110  | -2.133 |
| 世帯人数   | 0.059   | 3.796  |
| 初期尤度   | -1916.0 |        |
| 最大尤度   | -1456.3 |        |
| 尤度比    | 0.240   |        |
| 対象人数   | 939     |        |

される。ただし、用いる個人属性は一番最後に参加した時点のものとする。

ポアソン分布を仮定した参加回数モデルの推定結果を表1に示す。表1から年齢の高い人、有職者、免許保有者、普段公共交通を利用する人、世帯人数の

表2 参加意思の継続期間モデルの推定結果

| 説明変数     | パラメータ推定値 |          |             |
|----------|----------|----------|-------------|
|          | 指数分布     | ワイブル分布   | 対数ロジスティック分布 |
| 性別       | 0.009    | 0.001    | 0.027       |
| 年齢       | -0.007** | -0.007** | -0.006**    |
| 職業       | -0.221** | -0.292** | -0.118**    |
| 免許有無     | -0.137   | -0.137** | -0.128**    |
| 主要交通手段   | 0.038    | 0.040    | 0.025       |
| 世帯人数     | -0.057*  | -0.068** | -0.047**    |
| $\gamma$ | 0.106**  | 0.010**  | 0.000**     |
| $\alpha$ |          | 2.025**  | 3.524**     |
| 初期尤度     | -6109.8  | -8779.3  | -5832.4     |
| 最大尤度     | -5676.0  | -5210.9  | -5087.5     |
| 尤度比      | 0.071    | 0.406    | 0.128       |
| サンプル数    | 1456     |          |             |

(\*: 5%で、\*\*: 1%で有意)

多い人ほど、参加回数が多いことが分かった。また、推定した参加回数の期待値  $\lambda = \exp(0.71537) = 2.04$  であり、実際の平均参加回数を精度よく再現している。

## 5. 参加意思の継続期間に関する分析

参加意思の継続期間を分析するために、第2節で紹介した加速寿命モデルを利用する。そして、ハザード関数の基本分布として、第2節で挙げた指数分布、ワイブル分布と対数ロジスティック分布をそれぞれ仮定する。説明要因は第4節と同様である。

なお、対数尤度関数を定式化する際にセンサリングデータを考慮しない。例えば、図2の参加パターン(15)では途中の3時点が参加していないで最後の5時点目に復帰するため、明らかに区間センサリングを考慮する必要がある。しかし、センサリングを考慮すると尤度関数の定式化が複雑になる。本研究では参加意思の継続期間を分析するため、途中で離脱し再び復帰する個人の離脱前後の継続期間をそれぞれ異なる個人の継続期間として考えて、センサリングの影響を排除する。各モデルの推定結果は表2に示す。表2の初期尤度は分布パラメータ  $\gamma$  と  $\alpha$  をそれぞれの推定値、個人属性のパラメータを0として計算した対数尤度である。

表2から、参加意思の継続期間を最も精度よく説明できたのはワイブル分布を仮定した加速寿命モデルである。そして、参加意思の継続期間に有意に影響する要因として、年齢、職業、免許、世帯人数が挙

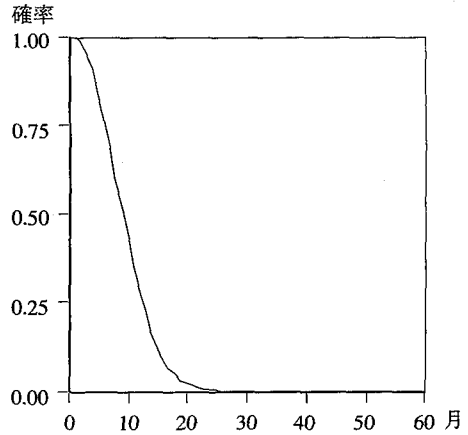


図5 参加意思の時間的長さの生存関数

げられる。また、 $\alpha = 2.025 > 1$  であることから、ハザード関数は増加関数であり、継続期間が長くなるにつれて参加から離脱する確率が高くなる。この場合の生存時間分布は図5に示される。図5から、平均的な個人の参加意思の継続期間は2年以内である。よって、2年以上調査を続ける場合、年齢が高い人、有職者、免許保有者、世帯人数が多い人は調査に継続参加する確率が高いため、回答には大きなバイアスが含まれる可能性がある。

## 6. おわりに

本研究では、適正なパネル調査期間を明らかにするために、個人の調査への参加意思の継続期間を、ワイブル分布を仮定した加速寿命モデルにより表現した。そして、参加意思の平均継続期間が約2年であることを示した。この結果は、被験者からみると初期サンプルの更新期間を2年と考えて設計することが妥当であることを意味している。一般性のある結論を得るためには異なるパネルデータを用いた実証研究がさらに必要であるが、その際には本研究の分析方法が有用であるものと考えられる。

また、今後センサリングデータを考慮した加速寿命モデルを構築することも必要である。

## 参考文献

- 1) Hensher D. A. and Mannering F. L. (1994): Hazard-based duration models and their application to transport analysis, Transport Review, Vol.14, No.1, pp.63-82.