

生存時間モデルによる世帯の自動車走行距離の推定*

Estimation of household car traveling distances using duration model *

桑野将司**・藤原章正***・張峻屹****・岡英紀*****

By Masashi KUWANO**・Akimasa FUJIWARA***・Junyi ZHANG****・Hideki OKA*****

1. はじめに

近年、環境問題に対する関心の高まりの中、自動車車両から排出される CO₂ や NO_x 等が問題視されており、問題解決のための方法として、自動車利用の抑制と電気自動車や低公害車への転換（車両技術の向上）の推進が求められている。このような状況下、世帯の自動車保有・利用行動に関する研究が活発になされてきた。一般に、世帯の自動車保有・利用行動は「購入段階」、「利用段階（走行段階）」と「保有段階」の3段階に分けることができ、これら3つの段階におけるそれぞれの行動は互いに影響を及ぼしあい、一連の自動車保有・利用状態を構成している。分析に際しては、購入段階の車種選択モデル、利用段階の走行距離モデル、保有段階の保有期間モデルなど、それぞれの段階において特徴的なモデリング手法が提案されてきている。しかし、この中でも特に「利用段階（走行段階）」に関しては、モデル予測精度が低く、世帯の自動車利用行動を十分に記述できているとは言いがたい。

自動車の走行距離は、日々の交通行動の一環である自動車利用を一定期間内で集計したものであり、その一連の行動を簡略的にモデル化したものが走行距離モデルである¹⁾。分析に際しては、一定期間の走行距離の対数値を被説明変数とする回帰分析を基礎とした走行距離モデルが構築されてきた。しかし、多くの既往研究では、走行距離の予測精度は低い。この理由として、アンケート調査データを用いることによる走行距離の測定誤差¹⁾、走行距離に影響を及ぼす変数が十分に取り入れられていないこと（除外変数の影響）などが考えられる。また、その他に、モデルの特定化ならびに分析に際しての確率

分布の仮定が適切でないことも予測精度が低い要因として考えられる。一般に、走行距離モデルに回帰モデルを適用する場合、目的変数である世帯の走行距離を自動車属性、世帯属性、居住地属性等の説明変数を用いた線形式により定式化し、誤差項が正規分布に従うことを仮定することになる。しかし、世帯は自動車を様々な目的で利用しているために予測不可能な偶然的要素が多く存在し、必ずしも関数形に線形式を仮定することは適切ではない。

そこで本研究では、より予測精度の高い走行距離モデルの構築のため、確率分布の仮定が可能な生存時間分析の適用を試みる。具体的には地域内の世帯の自動車走行距離を集計し、地域内の自動車利用距離分布を推定することにより世帯の自動車利用行動の記述を行う。

2. 使用データの概要

2.1 調査の概要

本研究では、地方都市を対象に公共交通の利便性が低く、自動車交通に依存した都市の世帯の自動車保有・利用実態を把握するため、平成17年11月7日から11月23日に実施された広島県東広島市志和町での世帯アンケート調査結果を用いる（表1）。

当該調査は、配布・回収は訪問方式で行われた。具体的には地区自治体を通じた配布・回収、または地区自治体が配布し学生が回収に訪れるといったいずれかの方法が採られた。

対象世帯は調査対象地域内で無作為に抽出され、966世帯から回答を得た。調査では世帯構成員の個人属性、世帯属性、保有自動車特性、自動車利用実態、交通行動実態等について質問をした（表2）。自動車特性については、保有している自動車最大3台までの特性についての記入を求めた。

本研究では全サンプル966世帯のうち記入漏れ、無回答票を除いた自動車保有世帯212世帯を対象に分析を行う。なお、本調査では保有自動車の特性を表す変数として、年間走行距離を質問したが、この調査項目に対しての回答率が低く、また複数台保有世帯では全ての自動車についての記入が必要であることから有効サンプル数が

*キーワード：交通行動分析，自動車保有・利用
** 正会員，修（工）広島大学大学院工学研究科
（〒739-8527 東広島市鏡山1-4-1）
TEL/FAX: 082-424-7825, E-mail: kuwano@hiroshima-u.ac.jp
*** 正会員，博（工）広島大学大学院国際協力研究科
（E-mail: afujiw@hiroshima-u.ac.jp）
**** 正会員，博（工）広島大学大学院国際協力研究科
（E-mail: zjy@hiroshima-u.ac.jp）
***** 学生員，学（工）広島大学大学院国際協力研究科
（E-mail: hideki-oka@hiroshima-u.ac.jp）

非常に少なくなるといった問題が明らかとなった。したがって、このサンプルの母集団代表性については注意が必要である。自動車利用に関する情報収集の効率的な調査方法および質問票設計の改善が今後の課題の1つである。

表1 調査の概要

調査対象	広島県 東広島市 志和町 住民
調査単位	世帯
調査期間	配布：平成17年11月7日 回収：平成17年11月17日～23日
配布数	1,447 世帯
回収数 (率)	966 世帯 (67%)
有効回答数 (率)	212 世帯 (15%)

表2 調査内容

世帯属性	居住年数, 自動車保有台数, 最寄りバス停までの距離など.
個人属性	性別, 年齢, 職業, 免許の有無など.
保有自動車特性	車両タイプ, 排気量, 乗車人数, 購入時期, 燃費, 利用目的, メインドライバー, 利用頻度, 年間走行距離など.
交通行動	出発地 (時間), 到着地 (時間), 目的, 交通手段, 同伴者など.

2.2 自動車利用特性

自動車利用特性として走行距離を見てみると、1世帯あたりの平均年間総走行距離は17,737kmであった。年間走行距離の分布に着目すると10,000kmから25,000km走行している世帯が多く、同じ町内に居住している世帯でも、自動車利用距離が大きく異なることが図1から観測された。さらに、25,000km以上走行している世帯も存在し、分布の裾が右に長くなっているのが特徴である。

1台あたりの走行距離に着目すると、平均は9,143kmであり、保有台数ごとに分けてみると、1台保有世帯では11,497km、2台保有世帯では9,182km、3台保有世帯では8,686kmである。

保有台数によって1台あたりの年間走行距離に差があるか否かについて5%の棄却域による検定を行った結果、1台保有世帯と複数台保有世帯の1台あたりの年間走行距離には有意な差が確認され、「複数台保有世帯の方が1台あたりの利用距離は短い」ことがわかった。しかし、2台保有世帯と3台保有世帯では差があるとは言えないという結果が得られ、複数台保有世帯ではその数が2台、3台と増えても、その利用に有意な違いがないことが明らかとなった。このように、保有自動車特性、複数台保有世帯の1台あたりの走行距離に保有台数間で有意な差が現れなかったのは、免許保有者の多くが自分専用の自動車を保有しており、世帯内の他の構成員からの制約を受けることなく自由に自動車を利用しているからであると考えられる。

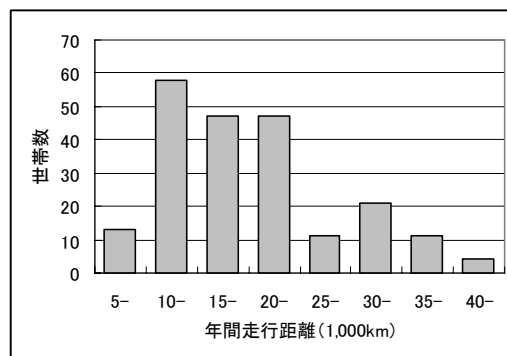


図1 年間走行距離の分布

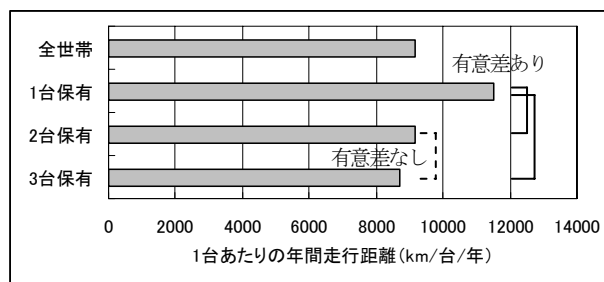


図2 保有台数ごとの1台あたりの年間走行距離

3. 生存時間モデルによる自動車走行距離モデル

2.2節の世帯の自動車利用特性からわかるように、自動車走行距離のモデル化に当たっては、目的変数が距離であるために負の値を取りえず、分布の裾が右に長くなるということが特徴となる。ゆえに通常正規性を前提とした手法は必ずしも適切とは限らず、適当な変数変換あるいは分布の形に強く依存しない手法を適用すべきである。

そこで本研究では、生存時間モデルによる世帯の自動車総走行距離（世帯が保有している全ての自動車の走行距離の総和）のモデル化を試みる。生存時間モデルは、対象とする事象が生起するまでの時間(生存時間)を解析の対象とし、交通の分野では一般的に活動時間や自動車保有期間のモデル化に適用される手法である²⁾。しかし、自動車総走行距離もまた、分布の裾が右に長いなど、解析の対象とされる生存時間と同様の性質を有していることから、十分に生存時間モデルの適用が可能であると考えられる。本研究では、対象とする事象を自動車の利用終了とし、生存時間モデルによる世帯の自動車総走行距離のモデル化を試みる。

生存時間モデル³⁾では、対象とする事象が生起するまでの時間 T の分布を生存関数 $S(t)$ 、およびハザード関数 $h(t)$ で表す。生存関数は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を表す関数、ハザード関数は対象とする事象がある時点 t までに生起していないという条件下で次の瞬間に事象が生起するという条件付き確率である。生存関数は式(1)、ハザード関数は式(2)で

表され、ハザード関数と生存関数の関係は式(3)のように表される。ここで、 $F(t)$ は累積分布関数、 $f(t)$ は確率密度関数である。

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

$$h(t) = f(t) / S(t) \quad (2)$$

$$S(t) = \exp\left(-\int_0^t h(t) dt\right) \quad (3)$$

本研究では、基準ハザードに特定の分布形を仮定するパラメトリックアプローチを採用し、説明変数の導入方法として比例ハザードモデルを用いる。このときハザード関数は式(4)のように表される。

$$h(t|x) = h_0(t) \exp(-\beta X) \quad (4)$$

ただし、 $h_0(t)$ は全ての説明変数が 0 の時のハザード関数であり、基準ハザード関数と呼ばれる。 X は説明変数ベクトル、 β はその未知パラメータベクトルである。

パラメトリックアプローチを用いる場合、基準ハザード関数 $h_0(t)$ に何らかの確率分布を仮定した上でモデルの推定を行うことが必要である。本研究では、基準ハザード関数の確率分布として、一般によく用いられる 1)指数分布、2)ワイブル分布、3)対数正規分布、4)対数ロジスティック分布を取り上げてモデルの推定を行い、どの分布形が最も適しているかについて実証的な検討を行う。それぞれの分布形の確率密度関数は式(5)から式(8)に示す通りである。

1)指数分布

$$f(t|X) = \exp(-\beta X) \exp\{-\lambda t \exp(-\beta X)\} \quad (5)$$

2)ワイブル分布

$$f(t|X) = \lambda \gamma t^{\gamma-1} \exp(-\gamma \beta X) \exp\{-\lambda t^\gamma \exp(-\gamma \beta X)\} \quad (6)$$

3)対数正規分布

$$f(t|X) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{(\ln t - \beta X)^2}{2\sigma^2}\right\} \quad (7)$$

4)対数ロジスティック分布

$$f(t|X) = \frac{\lambda \gamma t^{\gamma-1} \exp(-\beta X)}{\{1 + \lambda t^\gamma \exp(-\beta X)\}^2} \quad (8)$$

ここで、 λ は尺度パラメータ、 γ は形状パラメータである。

4. 実証分析

4.1 回帰モデルによる走行距離分析

ここでは、生存時間モデルによる走行距離分析の有効

性を検証するために、従来の回帰モデルの推定を行う。分析に際して、世帯の自動車総走行距離の対数値を被説明変数とする。説明変数としては、世帯属性と居住環境に関する属性を用いることとする。

回帰モデルによるモデル推定結果を表 3 に示す。表 3 より自由度調整済み決定係数が 0.392 と小さいことがわかる。これは、推定されたモデルによって世帯の自動車年間走行距離が説明される割合が低いことを示しており、走行距離の予測の困難性が表れている。

表 3 回帰モデルによる走行距離分析

説明変数	推定値	t 値
定数項	9.687	26.896
最寄のスーパーまでの距離(km)	-2.114	-2.703
最寄のバス停までの距離(km)	0.148	3.762
最寄の鉄道駅までの距離(km)	1.758	2.197
通勤先までの距離(km)	0.047	1.755
最寄の病院までの距離(km)	-0.211	-4.068
自動車保有台数	0.235	7.310
世帯収入(100 万円)	-0.178	-2.890
世帯人数	0.173	7.110
65 歳以上の世帯構成員の数	-0.211	-8.521
6 歳以下の世帯構成員の数	-0.264	-4.473
サンプル数		212
決定係数		0.412
自由度調整済み決定係数		0.392

4.2 生存時間モデルの確率分布仮定

パラメトリックな生存時間モデルを用いるため、基準ハザード関数に特定の確率分布を仮定する必要がある。本研究では有用性の高い 4 つの分布形を仮定する (1)指数分布、2)ワイブル分布、3)対数正規分布、4)対数ロジスティック分布)。

それぞれの分布形を仮定した場合の最終対数尤度を表 2 に示す。表 2 より、ワイブル分布を仮定した場合の最終対数尤度が最も高くなることが示された。ついで、最終対数尤度が高いのは対数ロジスティック分布、対数正規分布であり、指数分布の最終尤度が最も低くなった。

表 4 最終対数尤度の比較

確率分布	最終対数尤度
指数分布	-2297.67
ワイブル分布	-2149.54
対数ロジスティック	-2269.36
対数正規分布	-2274.35

4.3 モデル推定結果の比較

最も最終対数尤度が高かったワイブル分布を仮定した場合の生存時間モデルの推定結果のみを表 5 に示す。なお、パラメータの符号が正であれば絶対値が大きいほど走行距離が長くなり、反対に負であればパラメータの絶

対値が大きいほど走行距離が短くなることを示している。従来の分析手法である、回帰モデルを用いた走行距離モデルの推定結果（表 3）と比較すると、推定されたパラメータの符合は、すべて一致している。

パラメータ推定結果に着目すると、居住地環境に関する説明変数のうち、バス停や鉄道駅までの距離に関するパラメータは性で有意な値を得た。交通利便性が悪いほど、自動車利用は増すことを意味しており、理論的に妥当である。次に生活施設に着目すると、スーパーまでの距離に関するパラメータが負で有意の値となった。これは地方都市では、スーパーまでの距離が長くなると、買い物頻度は少なくなり、1度の買い物で大量に購入する「まとめ買い」の傾向を反映している結果であると考えられる。次に、世帯属性に着目すると、世帯人数が多い世帯ほど自動車の年間総走行距離は長くなる傾向にある。しかし、幼児がいる世帯では自動車利用は少ないという結果を得た。今後少子高齢化が進むにつれ、自動車利用は減少すると考えられるが、一方で、調査対象地域ではバスの利便性が低いことから、交通行動への制約が懸念される。このように例えば、スーパーまでの距離の増加や少子高齢化の進行が自動車利用抑制につながる事が明らかとなったが、反面これは世帯の生活の質（QOL: Quality of life）を下げる危険性もあり、自動車利用による環境への影響と QOL の水準維持を関連付けた研究を行うことが今後の研究課題である。

表 5 生存時間モデルによる走行距離分析

説明変数	推定値	t 値
最寄のスーパーまでの距離	-2.067	-2.791
最寄のバス停までの距離	0.145	3.601
最寄の鉄道駅までの距離	1.738	2.290
通勤先までの距離	0.057	2.092
最寄の病院までの距離	-0.204	-4.025
自動車保有台数	0.257	7.712
世帯収入	-0.161	-2.554
世帯人数	0.169	6.961
65 歳以上の世帯構成員の数	-0.201	-8.877
6 歳以下の世帯構成員の数	-0.273	-5.516
尺度パラメータ	9.571	26.880
形状パラメータ	0.166	14.351
サンプル数		212
最終対数尤度		-2149.54

4.4 モデル精度の比較

ここでは、回帰モデルと生存時間モデルによるモデルの現況再現性の比較を行う。それぞれ実測値と推定値の関係を図 3、図 4 に示す。両者を比較すると、生存時間モデルの方が、推定値と実測値の誤差が少なく、精度の高い予測がなされたと言える。

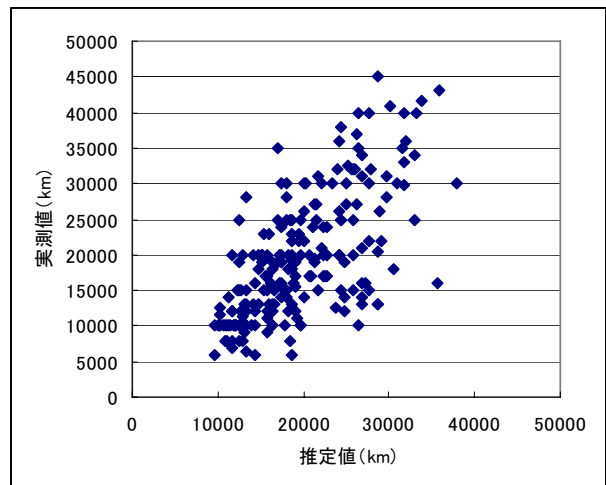


図 3 実測値と測定値の関係（回帰モデル）

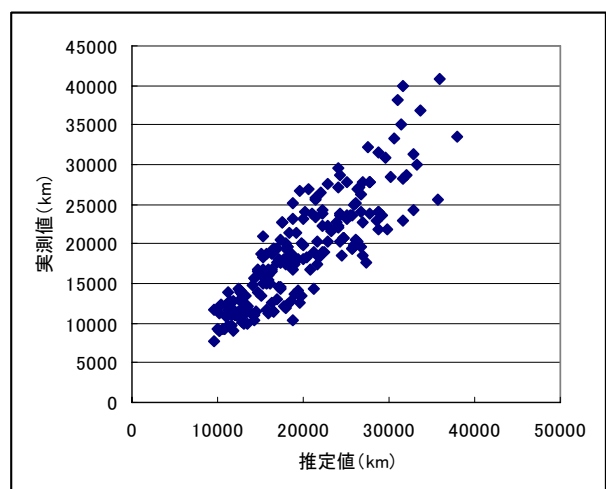


図 4 実測値と測定値の関係（生存時間モデル）

5. 結論

本研究では、従来の回帰モデルによる走行距離モデルの予測精度が低いことを問題視し、生存時間モデルによる年間走行距離の分析を試みた。平成 17 年に地方都市郊外部で実施された調査結果より、同じ町内に住んでいても世帯により年間総走行距離が大きく異なることが明らかとなった。実証分析の結果、ハザード関数にワイブル分布を仮定したときの現況再現性が最も高く、従来の回帰モデルに比べて、生存時間モデルを適用したほうが、誤差が少なく予測精度が高くなることが示された。

参考文献

- 1) 山本俊行, 中川展孝: 観測の不完全性を考慮した自動車年間走行距離の分析, 土木計画学研究・講演集, Vol. 32, CD-ROM, 2005.
- 2) Hensher, D. and Mannering, F.: Hazard-based duration models and their application to transport analysis, Transport Reviews Vol.14 No.1, pp.63-82, 1994.
- 3) 北村隆一, 佐々木邦明, 山本俊行, 森川高行, 藤井聡: 交通行動の分析とモデリング, 技報堂出版, pp.190-203, 2002.