

セダン型小型乗用車の死亡事故発生要因の ポアソン回帰による分析[†]

縄田和満*, 嶋村宗正**

An Analysis of Fatal Accidents in Sedan-Type Small Passenger Cars by the Poisson Regression Model

Kazumitsu Nawata* and Munemasa Shimamura**

This paper analyzed fatal accidents in sedan-type small passenger cars. The factors which may affect fatal accidents are analyzed using the regression type model. The nine factors, which describe characteristics of drivers and driving conditions, are used as explanatory variables in the analysis. The maximum likelihood method based on the Poisson distribution and weighted least squares method are used to estimate the model. The assumption of the Poisson distribution and exogeneity of an explanatory variable are also tested.

交通事故の発生要因を分析することは、交通事故発生の低減対策を考える上で重要であり、マイクロデータやマクロデータを用いた分析が行われている。本論文では、財団法人交通事故総合分析センターで集計された1995年末現在発売・使用されているセダンAクラスの乗用車のうち事故要因を表す変数に欠損値のない28車種のデータを使い死亡事故台数に影響を与えると考えられる要因の分析を行った。我が国においては、これまでは事故要因を集計した統合データベースは存在せず、本論文は死亡事故と事故要因の関連を分析した最初の試みである。データでは、死亡事故が起らなかった車種や死亡事故台数が小さい車種がかなり多く存在することから(死亡事故が0台は2車種、1-5台は9車種)離散データとして取り扱うことをが必要となっている。また、車種によって登録台数に大きな差があることを考慮して分析を行う必要がある。ここでは、分析にポアソン分布に基づく回帰モデルを用い、モデルの推定には最尤法および加重最小二乗法による推定を行った。また、線形確率モデルと事故要因を表す変数の標本選択による偏りについての考察を加えた。

1. はじめに

交通事故の発生要因を分析することは、交通事故発生の低減対策を考える上で重要であり、マイクロデータやマクロデータを用いた分析が行われている。マイクロデータを用いた分析としては、個々の事故例を調査し、事故原因を多様な角度から分析して対策に結び付ける試みが行われている。しかしながら、このような分析には交通事故の全体像がつかみにくいという欠点がある。

論文受付: 1998年7月 受理: 1998年10月

[†] この論文執筆にあたり、財団法人交通事故総合分析センターにご協力頂いた。また、この論文に対して横浜国立大学松浦克巳教授および本誌レフリーより貴重なコメントを頂いた。ここに記して感謝の意を表したい。なお、この論文の内容・見解は筆者個人のものであり、財団法人交通事故総合分析センターの見解ではない。

* 東京大学大学院総合文化研究科国際社会科学専攻, 〒153-8902 目黒区駒場 3-8-1
e-mail: nawata@waka.c.u-tokyo.ac.jp

** 財団法人交通事故総合分析センター, 〒136-0071 江東区亀戸 1-4-2

ある。Department of Transportation, United Kingdom (1992) では、マクロデータを使い交通事故要因の分析を行っている。この結果、運転者の性別、年齢層、制限速度、車の大きさなどが傷害事故の発生に影響することが示されている。この研究は英国で行われたものであり、交通環境の異なる日本のデータを用いた分析が必要となっている。

日本におけるマクロデータについては、事故データと車両データを統合した交通事故統合データベースが財団法人交通事故総合分析センターによって作成され、人・車・交通環境の観点からの総合的な分析が可能となってきている。交通事故総合分析センター (1997) では、主として 1500 cc 以下のセダン型普通乗用車 (セダン A クラス) に属する事故の発生率が公表され、運転者の特性、交通環境、車の特性についての分析が行われている。しかしながら、交通事故要因間の相互の影響について検討がなされていないことなど交通事故の発生要因について十分な分析が行われているとは言い難い。

本論文では、交通事故総合分析センター (1997) で示された 1995 年末現在発売・使用されているセダン A クラスの乗用車のうち事故要因を表す変数に欠損値のない 28 車種のデータを使い死亡事故台数に影響を与えると考えられる要因の分析を行った。このデータでは、死亡事故が起らなかった車種や死亡事故台数が小さい車種がかなり多く存在することから (死亡事故が 0 台は 2 車種、1-5 台は 9 車種) 離散データとして取り扱うことをが必要となっている。(離散データ分析モデルおよびその応用例については、Cameron and Trivedi (1986), Mondel (1993), Winkelmann (1994), Cameron and Johansson (1997) 等を参照せよ。) また、車種によって登録台数に大きな差があることを考慮して分析を行う必要がある。ここでは、分析にポアソン分布に基づく回帰モデルを用い、モデルの推定には最尤法および加重最小二乗法による推定を行った。さらに、ポアソン分布および説明変数の外生性の検定を行い、最後に線形確率モデルと事故要因を表す変数の標本選択による偏りについての考察を加えた。

2. 分析モデルおよび推定方法

2.1 ポアソン分布に基づく最尤法

本章では、マクロデータに基づく分析モデルを与える。いま、 k 種の異なった車種があり、 y_{ij} を第 i 車種の j 番目の登録車両が観測期間内に死亡事故を起こしたかどうかのダミー変数、すなわち、

$$(2.1) \quad y_{ij} = \begin{cases} 0 & \text{死亡事故を起こさなかった場合} \\ 1 & \text{死亡事故を起こした場合} \end{cases}$$

とする。通常の二値データの分析と同様、事故を起こすかどうかは、

$$(2.2) \quad y_{ij}^* = x'_{ij}\alpha + u_{ij}$$

の符号によって決定され、 $y_{ij}^* > 0$ の場合死亡事故が起こり、 $y_{ij}^* \leq 0$ の場合死亡事故は起こらないとする。 x_{ij} , α は事故要因を表す変数、未知のパラメータの k 次元のベクトルである。 F を $-u_{ij}$ の分布関数とすると、死亡事故の起こる確率 $p_{ij} \equiv P(y_{ij}=1|x_{ij})$ は、

$$(2.3) \quad p_{ij} = F(x'_{ij}\alpha)$$

と表すことができる。本分析で用いたデータでは x_{ij} の分布は有界であり、 $1 \gg p_{ij}$, $n_i \gg 0$ となっている。したがって、小数の法則及びポアソン分布の再生性から、第 i 車種の事故発生数 $Y_i = \sum_{j=1}^{n_i} y_{ij}$ は $\lambda = P_i \equiv \sum_{j=1}^{n_i} p_{ij}$ のポアソン分布に従うとみなすことができる。

通常の分析では、 F を標準正規分布またはロジスティック分布と仮定し、分析が行われるが、各

変数を集計したマクロデータしか得られないためこれらの仮定を直接使うことはできない。ここでは、F は線形関数で近似でき、線形確率モデルを使うことが可能であると仮定して分析を行う。(線形確率モデルの妥当性については6章を参照せよ。) F が線形関数で近似されるとすると、

$$(2.4) \quad p_{ij} \approx x'_{ij}\beta$$

となり、 n_i を第 i 車種の登録台数とすると、

$$(2.5) \quad P_i \approx X_i'\beta, \quad X_i = \sum_{j=1}^{n_i} x_{ij}$$

が得られる。

したがって、 β はポアソン分布に基づく対数尤度

$$(2.6) \quad \log L(\beta) = \sum_{i=1}^k \{-X_i'\beta + Y_i \log(X_i'\beta) - \log(Y_i!)\}$$

を最大にすることによって推定することができる。なお、変数の対数変換などを行うと個々のマイクロレベルの行動モデルとそれを集計したマクロレベルのモデルの整合性が失われてしまうことがあり、注意が必要である。

2.2 分散の不均一性を考慮した加重最小二乗法

前節では、ポアソン分布に基づく最尤法によってモデルの推定を行う方法について述べたが、このモデルでは、加重最小二乗法によってもモデルの推定を簡単に行うことができる。(2.4) 式の関係は、

$$(2.7) \quad \begin{cases} y_{ij} \approx x'_{ij}\beta + \varepsilon_{ij}, \\ E\varepsilon_{ij} = 0 \end{cases}$$

と書き直すことができる。 $p_{ij} \approx 0$ であるから、

$$V\varepsilon_{ij} = p_{ij} \cdot (1 - p_{ij}) \approx p_{ij}$$

である。

(2.7) 式を車種ごとに合計すると、第 i 車種の事故発生数 Y_i は、

$$(2.8) \quad \begin{cases} Y_i \approx X_i'\beta + \varepsilon_i \\ Y_i = \sum_{j=1}^{n_i} y_{ij}, \quad X_i = \sum_{j=1}^{n_i} x_{ij}, \quad \varepsilon_i = \sum_{j=1}^{n_i} \varepsilon_{ij} \end{cases}$$

となる。(2.8) 式の両辺を登録台数 n_i で割ると事故発生率に関するモデル、

$$(2.9) \quad \begin{cases} y_i \approx x_i'\beta + \varepsilon_i^* \\ y_i = Y_i/n_i, \quad x_i = X_i/n_i, \quad \varepsilon_i^* = \varepsilon_i/n_i \end{cases}$$

が得られる。ここで、

$$(2.10) \quad \begin{cases} E\epsilon_i^* = 0, \\ V(\epsilon_i^*) = \sum_{j=1}^m p_{ij}(1-p_{ij})/n_i \approx \sum_{j=1}^m p_{ij}/n_i = x_i' \beta / n_i \end{cases}$$

であるので、モデルは加重最小二乗法を使った次のような2段階の手順によって求めることができる。

- i) $S_1 = \sum n_i (y_i - x_i' \beta)^2$ を最小にする1段階目の推定量 $\hat{\beta}_1$ を求める。
- ii) 1段階目の推定量から重みを与える変数 $w_i = n_i / x_i' \hat{\beta}_1$ を計算し、 $S_2 = \sum w_i (y_i - x_i' \beta)^2$ を最小にする推定量 $\hat{\beta}_2$ を求め最終の推定量とする。

3. データ

3.1 死亡事故台数及び登録台数

データは、財団法人交通事故分析センターの交通事故統合データベースによった。死亡事故台数は、平成5年から平成7年までの3年間の合計であり、セダンAのうち事故要因のデータに欠損値のない28車種のデータを用いた。対象28車種のこの期間の死亡事故台数は901台で、その分布は表3.1、図3.1の通りである。車種ごとの死亡事故台数は、死亡事故台数が0台が2車種、1-5台が9車種、6-20台が1車種、21-50台が9車種、51-100台が9車種、101-150台が2車種

表3.1 死亡事故台数の分布

死亡事故台数	車種数
0	2
1-5	9
6-20	1
21-50	9
51-100	5
101-150	2
計	28

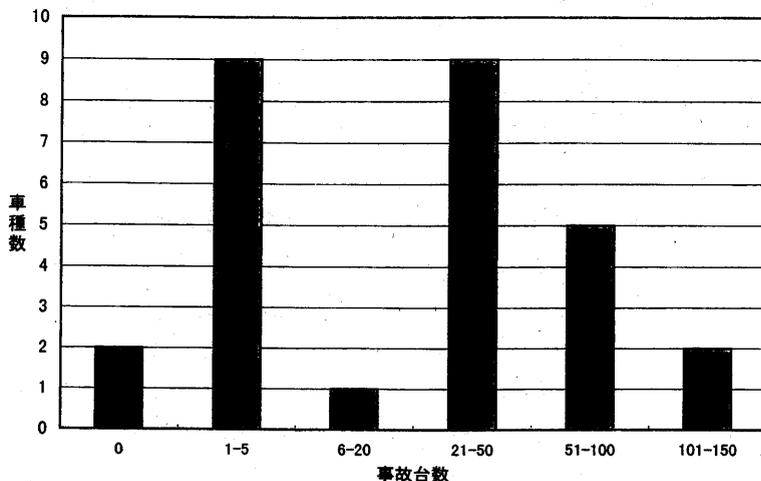


図3.1 事故台数の分布

であり、最大144台、平均32.2台である。死亡事故台数0台のものが2車種、1-5台と小さいものが9車種存在するが、このことは、データを離散データとして扱う、ポアソン分布を仮定した分析の必要性を意味しているといえる。

表3.2、図3.2は登録台数の分布であるが、28車種合計の登録台数は、1,592万台である。車種ごとの登録台数は、5万台以下が5車種、5-10万台が4車種、10-50万台が7車種、50-100万台6車種、100-200万台が5車種、200-300万台が1車種で、平均56万8,000台、最大は273万3,000台となっている。

登録台数が増加するに従い死亡事故台数は増加するが、登録台数1万台当たりの死亡事故率(死亡事故台数/登録台数 \times 10,000)の分布は表3.3および図3.3の通りである。登録台数1万台当たりの死亡事故率は、0-0.3が4車種、0.3-0.6が13車種、0.6-0.9が7車種、0.9-1.2が2車種、1.2-1.5が2車種となっており、平均0.57、標準偏差0.57、最大1.34である。死亡事故率と登録台数の関係を比較すると図3.4のようになる。登録台数が増加するに従い死亡事故台数は増加するが、登録台数が少ない場合、分散が大きくなっていることが分かる。(事故発生件数がポアソン分布に従うとすると、登録台数で割った死亡事故率では分散は登録台数に反比例する。)

3.2 交通事故発生要因

交通事故の発生要因については、運転者、交通環境、車の観点から分析する必要がある。こ

表3.2 登録台数の分布

死亡事故台数	車種数
0-50	5
50-100	4
100-500	7
500-1,000	6
1,000-2,000	5
2,000-3,000	1
計	28

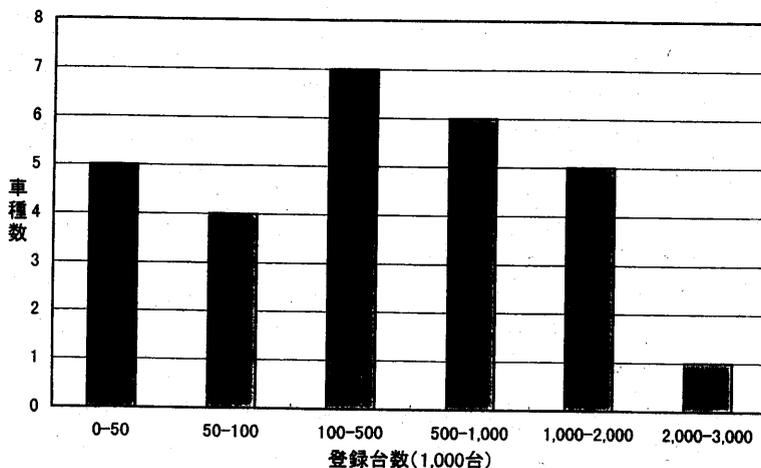


図3.2 登録台数の分布

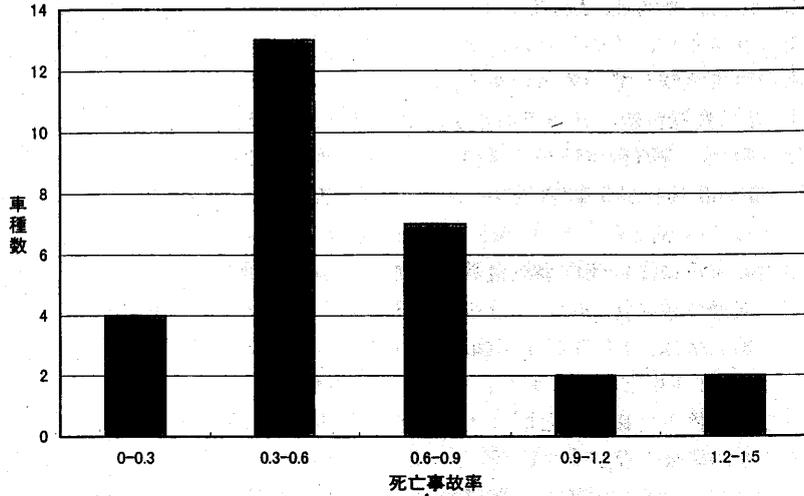


図 3.3 登録車両1万台当たりの死亡事故率

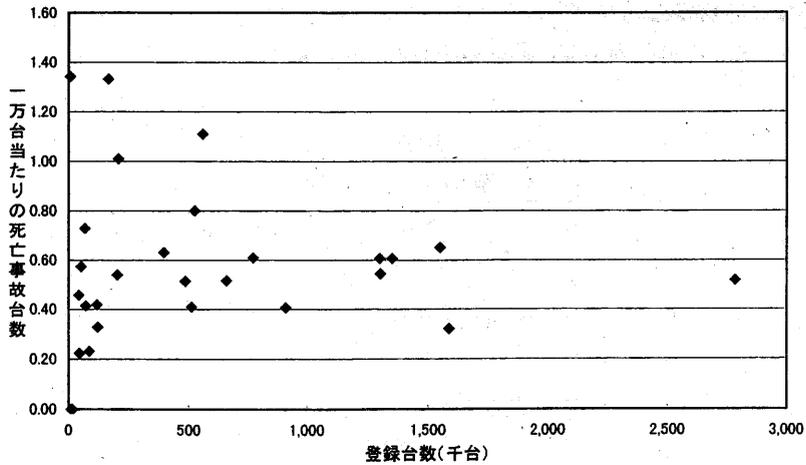


図 3.4 自動車登録台数と死亡事故率

表 3.3 登録車両1万台当たりの死亡事故率の分布

死亡事故台数	車種数
0-0.3	4
0.3-0.6	13
0.6-0.9	7
0.9-1.2	2
1.2-1.5	2
計	28

ここでは、交通事故統合データの中で、これらを示すことのできるデータを可能な限り選んだ。今回の分析で利用する事故要因は次表の通りであり、死亡事故台数と同様、平成5年から7年までの3年間のデータから計算している。単独事故比率 *SINGLE* は事故の結果であり、運転者

表 3.4 事故要因を表す変数の定義

変 数	定 義
<i>MALE</i>	運転者の性別が男である比率
<i>AGE24</i>	運転者の年齢が 24 歳以下である比率
<i>AGE65</i>	運転者の年齢が 65 歳以上である比率
<i>BUSINESS</i>	通行の目的が業務である比率
<i>DRIVE</i>	通行の目的がドライブである比率
<i>BELT</i>	シートベルトの着用率
<i>NIGHT</i>	夜間運転の比率
<i>SPEED</i>	危険認知速度が 41 km/h 以上の比率
<i>SINGLE</i>	事故類型が単独事故である比率

や運転条件によって決定され外生変数とみなすことのできる他の事故要因と異なるが、これは運転の「乱暴さ・不注意の度合い」を表す指標と考えられるため分析に加えた¹⁾。

事故発生要因を表す変数は、すべての登録車を調査した値を用いるべきであるが、残念ながら現状ではそのようなデータは存在しない。ここでは、入手可能なデータとして、少なくとも 1 人以上の死傷者がいた事故について関与した車について集計したものから求めた。このデータはランダムサンプリングによって得られたものではなく、標本選択による偏りを生じている可能性は否定できない。しかしながら、死傷者のいた事故件数は死亡事故件数に比べて非常に多く、1万台当たりの死傷事故関与台数は 150.3~263.3、平均 187.5 台で、登録台数の 2% 程度を調査したことになる。また、登録台数も大きいため、不完全であるが、このデータは事故要因をかなり正確に表していると考えて分析を行う。事故要因のうち、ベルト着用率と事故台数については関与した車すべてにおける値であるが、その他の事故要因は少なくとも 1 人以上死傷した乗員がいた車においての値である。予想される符号は、*NIGHT*、*SPEED*、*SINGLE* が正、*BELT* が負であり、これらについては片側検定を行った。事故要因を表す変数の 28 車種における平均・標準偏差は次の通りである。一部の変数の標準偏差が小さいが、これは車種間での値の差が小さいことを示しており、これらの変数の影響は有意に観測されにくい可能性がある。

表 3.5 事故要因を表す変数の平均と標準偏差

変 数	平 均	標準偏差
<i>MALE</i>	50.9 %	12.4 %
<i>AGE24</i>	29.1 %	10.9 %
<i>AGE65</i>	3.1 %	2.0 %
<i>BUSINESS</i>	9.6 %	3.6 %
<i>DRIVE</i>	8.5 %	3.1 %
<i>BELT</i>	86.0 %	2.6 %
<i>NIGHT</i>	33.5 %	3.3 %
<i>SPEED</i>	71.6 %	14.6 %
<i>SINGLE</i>	5.7 %	1.7 %

4. モデルの推定結果

4.1 最尤方法による推定結果

ここでは、2章で述べたポアソン回帰モデルの最尤方法による推定結果を示す。推定はTSP4.4によって行った。まず、事故要因を表す変数9個すべてを含んだモデルの推定を行った。次に、係数の推定値をその標準偏差で割った t 値の絶対値が小さい変数を除いていくことによって、影響の少ない事故要因を除外したモデルの推定を行った。(なお、変数選択の方法としてはこの他、対数最大尤度に基づく方法などが考えられるが手順が煩雑な上に、この場合選択される変数に大きな差はない。)事故要因を表す変数は、その総和(=各車種における平均値×登録台数)を求め、 X_i とする必要があるが、登録台数が大きいため、登録台数は1万台を単位として X_i を計算した。このため、事故が起こる確率は1万台当たりのものとなっている。推定結果は、表4.1の通りである。

t 検定によって各係数の検定を行うと、すべての事故要因を含むモデルでは、*AGE24*と*DRIVE*が5%の有意水準で有意である。また、*SINGLE*は予想される通りの符号であり、1%の有意水準で有意である。 t 値の絶対値は小さく5%の有意水準で有意ではないが、*NIGHT*は予想される通りの符号であるが、*BELT*と*SPEED*は反対の符号となっている。(ここでの検定は漸近的なものであるが、*BELT*、*NIGHT*、*SPEED*、*SINGLE*で片側検定を行うため尤度比検定等でなく、自由度を28-(説明変数の数+1)として t 検定を行った。)

影響の少ない事故要因を除外していったモデルの推定では、すべてのケースにおいて*SINGLE*は予想される符号通りで1%の有意水準で有意となっており、重要な事故要因であると考えられる。これらのモデルの推定方法が最尤法であることを考慮して、モデルをAIC(Akaike Information Criterion)によって選択すると、事故要因を6変数としたものが選択される。選

表4.1 最尤法による推定結果 (カッコ内は t 値)

変数の数	9	8	7	6	5	4
定数項	0.047 (0.312)	-0.034 (0.249)	-0.020 (-0.167)	-0.068 (-1.563)	0.004 (0.249)	0.026 (1.830)
<i>MALE</i>	-1.409 (-1.869)	-1.468 (-1.962)	-1.417 (-2.202)	-1.441 (-2.073)	-1.025 (-2.111)	-0.655 (-2.409)
<i>AGE24</i>	-2.406 (-2.303)	-2.410 (-2.336)	-2.392 (-2.462)	-2.413 (-2.476)	-1.576 (-2.429)	-1.481 (-2.457)
<i>AGE65</i>	1.019 (0.315)	1.324 (0.446)				
<i>BUSINESS</i>	3.482 (1.797)	3.499 (1.915)	3.455 (1.901)	3.358 (1.858)	2.568 (1.722)	
<i>DRIVE</i>	8.271 (2.246)	8.158 (2.352)	7.975 (2.449)	7.861 (2.569)	6.829 (2.301)	5.685 (2.329)
<i>BELT</i>	-0.427 (-0.267)	-0.617 (-0.459)	-0.550 (-0.412)			
<i>NIGHT</i>	3.949 (1.353)	3.904 (1.363)	3.380 (1.578)	3.382 (1.601)		
<i>SPEED</i>	-0.128 (-0.196)					
<i>SINGLE</i>	9.896 (2.803)	9.668 (3.606)	9.888 (3.817)	10.595 (4.679)	12.450 (5.660)	10.546 (4.954)
<i>Log L</i>	-83.366	-83.391	-83.489	-83.537	-84.694	-85.796
<i>AIC</i>	184.731	182.782	180.978	179.074	179.387	179.592

択された事故要因は、*MALE*、*AGE24*、*BUSINESS*、*DRIVE*、*NIGHT*、*SINGLE*である。各係数のt検定では *MALE*、*AGE24*、*DRIVE* が有意水準5%で有意となっており、*SINGLE* のt値は4,679と大きく、*p*値(片側)は $6,416 \times 10^{-5}$ で、常識的な有意水準では有意となっている。すなわち、この推定結果からは、運転者が男性や24歳以下であることは死亡事故の発生確率を減少させ、運転目的がドライブであることや単独事故であることは死亡事故の発生確率を増加させることが認められる。

4.2 加重最小二乗法による推定結果表

表4.2は加重最小二乗法による推定結果である。ポアソン分布に基づく最尤法の場合と同様、事故要因を表す変数9個すべてを含んだモデルの推定を行い、次に、係数の推定値のt値の絶対値が小さい変数を除いていくことによって、影響の少ない事故要因を除外したモデルの推定を行った。

表4.2 加重最小二乗法による推定結果 (カッコ内はt値)

変数の数	9	8	7	6	5	4
定数項	-0.289 (-0.094)	-0.741 (-0.623)	-0.654 (-0.711)	-0.609 (-0.677)	0.046 (0.124)	0.276 (1.008)
<i>MALE</i>	-1.331 (-1.407)	-1.318 (-1.450)	-1.275 (-1.486)	-1.375 (-1.691)	-0.975 (-1.366)	-0.685 (-1.045)
<i>AGE24</i>	-2.357 (-1.443)	-2.236 (-1.485)	-2.348 (-1.513)	-2.328 (-1.520)	-1.489 (-1.108)	-1.531 (-1.151)
<i>AGE65</i>	0.856 (0.169)	0.681 (0.142)				
<i>BUSINESS</i>	3.211 (0.983)	3.166 (1.053)	3.178 (1.083)	3.106 (1.082)	2.492 (0.940)	
<i>DRIVE</i>	8.247 (1.360)	8.175 (1.393)	8.181 (1.430)	7.720 (1.418)	6.521 (1.205)	5.394 (1.111)
<i>BELT</i>	-0.396 (-0.156)					
<i>NIGHT</i>	3.765 (0.866)	3.790 (0.898)	3.600 (0.963)	3.173 (0.891)		
<i>SPEED</i>	-0.161 (-0.192)	-0.226 (-0.306)	-0.270 (-0.391)			
<i>SINGLE</i>	9.431 (1.565)	10.096 (2.2301)	9.888 (3.817)	10.215 (2.395)	12.068 (3.030)	10.423 (2.826)
R^2	0.448	0.443	0.444	0.445	0.420	0.388

注) R^2 は変換された変数から計算したもの。

推定結果は表4.2の通りであるが、係数の推定値は、最尤法の推定結果と似た値となっている。また、影響の少ない事故要因を除外して行く順序も *BELT* と *SPEED* が変わるだけで(最尤法では *SPEED* が最初に *BELT* が3番目に除外されるが、加重最小二乗法では *BELT* が最初で *SPEED* が3番目に除外される)他は同一である。事故要因が6以下のモデルでは2つの推定方法で同一の変数が選択されている。しかしながら、最尤法に比較して、標準誤差が大きくt値の絶対値が小さくなっている。このため、事故要因を表す変数すべてを含んだモデルでは5%の有意水準で(*SINGLE*を含め)有意となるものはない。影響の少ない事故要因を除外したモデルにおいても、5%の有意水準で有意となるのは *SINGLE*のみであり、他はいずれの場合も有意とはならない。このことは、最尤法による推定の重要性を示していると考えられる。

4.3 単独事故比率の他の説明変数との関係

前節までの分析から単独事故比率 *SINGLE* は死亡事故の重要な事故要因であることが認められる。しかしながら、*SINGLE* は他の事故要因と異なり、事故の結果であり、運転者、運転条件を特定しても決定することはできない。ここでは、*SINGLE* と他の事故要因との関係についての分析を行う。分析には *SINGLE* を非説明変数、他の事故要因を説明変数とする重回帰モデルを用い、登録台数を重みとする加重最小二乗法によって推定を行った。これまでと同様、まず、すべての説明変数を加えたモデルの推定を行い、次に *t* 値の絶対値の小さい説明変数を除いていくことによって、影響の少ない説明変数を除外したモデルの推定を行った。予想される符号は、*AGE24*、*NIGHT*、*SPEED* が正、*BELT* が負であり、これらについては片側検定を行う。

推定結果は、表 4.3 の通りである。*BELT* はすべてのモデルにおいて負の値であり、有意水準 1% で有意 (片側検定) であり、*SINGLE* との強い関係が示唆されている。また、*SPEED*、*NIGHT* はすべてのモデルで正の値となり、説明変数が 6 個以下のモデルでは、*SPEED*、*NIGHT* が 5% で有意 (片側検定) である。(説明変数が 4 個のモデルでは *NIGHT* は 1% で有意。)したがって、シートベルト着用、速度 41 km 以上、夜間運転などの事故要因は、*SINGLE* という運転の「乱暴さ、不注意の度合い」を示す変数を通して死亡事故発生に影響していると考えられる。

表 4.3 *SINGLE* の分析モデルの推定結果 (カッコ内は *t* 値)

説明変数の数	8	7	6	5	4
定数	0.2368 (2.307)	0.2368 (2.368)	0.2315 (2.439)	0.2219 (2.412)	0.2126 (2.394)
<i>MALE</i>	-0.0008 (-0.023)				
<i>AGE24</i>	0.0156 (0.254)	0.0142 (0.379)	0.1979 (0.711)	0.0128 (0.520)	
<i>AGE65</i>	0.2180 (1.173)	0.2194 (1.278)	0.2273 (1.384)	0.2367 (1.471)	0.2085 (1.399)
<i>BUSINESS</i>	0.0522 (0.428)	0.0536 (0.512)	0.0579 (0.576)		
<i>DRIVE</i>	0.0369 (0.163)	0.0401 (0.224)			
<i>BELT</i>	-0.3501 (-3.351)	-0.3506 (-3.529)	-0.3490 (-3.605)	-0.3305 (-3.674)	-0.3261 (-3.700)
<i>NIGHT</i>	0.2245 (1.481)	0.2261 (1.718)	0.2340 (1.890)	0.2290 (1.883)	0.2695 (2.936)
<i>SPEED</i>	0.0345 (1.122)	0.0348 -1.2730	0.0383 (1.749)	0.0419 (2.026)	0.0373 (2.030)
R^2	0.8253	0.8253	0.8490	0.8221	0.8200

5. ポアソン分布および *SINGLE* の外生性の検定

5.1 ポアソン分布の検定

ここでは、ポアソン分布を仮定して推定を行った。しかしながら、ポアソン分布の仮定が満足されず、期待値と分散が等しくならない場合があり、いくつかの検定方法が提案されている。また、期待値と分散が等しくない場合は、二項分布・負の二項分布や (これらを含む) Katz シ

ステムなどを分析に用いることが考えられている。(詳細については, Cox (1983), Collings and Margolin (1985), Hausman, Hall and Griliches (1984), Cameron and Trivendi (1986, 1990), Lee (1986), Dean and Lawless (1989), Consul ana Famoye (1992) 等を参照せよ。)ここでは, Cameron and Trivendi (1990) の検定方法に基づき, ポアソン分布の検定を行う。

$\mu_i = EY_i$, $\sigma_i^2 = V(Y_i)$ とするとポアソン分布では $\sigma_i^2 = \mu_i$ である。帰無仮説・対立仮説を Collings and Margolin (1985) に従い,

$$H_0: \sigma_i^2 = \mu_i, \quad H_1: \sigma_i^2 = \mu_i + \alpha \mu_i^2, \quad \alpha \neq 0$$

とする。なお, ポアソン最尤推定量は帰無仮説・対立仮説のいずれのもとにおいても一致推定量である。(McCullagh and Nelder (1983), Gourieroux, Montfort, and Trongnon (1984).)

この場合, 検定統計量は,

$$(5.1) \quad T = \left[\sum \left(\frac{\hat{\mu}_i^2}{2} \right) \right]^{-1/2} \left[\sum \frac{1}{2} \{ (Y_i - \hat{\mu}_i)^2 - Y_i \} \right],$$

である。 $\hat{\mu}_i$ は μ_i の推定量である。帰無仮説もとで, T は漸近的に標準正規分布に従う。 AIC の値が最小になった事故要因を6変数とするモデルについて T の値を計算すると, $T=0.5448$ となる。有意水準を5%で帰無仮説は棄却されず, ポアソン分布との相違は認められない。

5.2 SINGLE の外生性の検定

すでに述べたように *SINGLE* は, 運転の「乱暴さ・不注意の度合い」を表す変数と考えられるが, 事故の結果であり, これらを直接測定するものではない。このため, 推定式の誤差項と相関が生じ, 外生変数として扱えない可能性がある。このため, ここでは Hausman (1978) の検定によって, *SINGLE* の外生性の検定を事故要因が6変数であるモデルに対して行う。検定方法は次の通りである。

- i) *SINGLE* 以外のすべての事故要因で *SINGLE* を回帰し回帰残差 e を求める。
- ii) 事故要因が6変数であるポアソン回帰モデルに e を説明変数として加え, 推定を行い e の係数 β_e について, $H_0: \beta_e = 0$, $H_1: \beta_e \neq 0$ の t 検定を行う。(この帰無仮説が棄却された場合, 外生変数である仮説は棄却される。この検定の実証研究への応用例については Sander (1992) 等を参照せよ。)

e を加えたモデルの推定結果は次の通りである。

表 5.1 e を加えたモデルの推定結果

変数	定数項	MALE	AGE24	BUSINESS	NIGHT	SINGLE	e
推定値	-3.464	-1.506	-2.663	2.673	8.928	12.144	-2.226
t 値	-0.377	-2.132	-2.053	0.951	1.898	2.441	-0.374

t 値は -0.374 であり, 有意水準5%で $H_0: \beta_e = 0$ は棄却されず, *SINGLE* を外生変数でないとして取り扱う必要性は認められない。

6. 推定における問題点

本論文では, 「交通事故データベース」に基づき推定を行ったが, データの制約上, 2つの問題点が生じている。第一は線形確率モデルとした仮定の是非である。この仮定が妥当であれば, 2章で示した通り, (推定量の効率の問題はあるが) 事故要因について各車種の平均のみによって分析を行うことが可能となる。しかしながら, 線形確率モデルには, 特に死亡事故確率の変

動が大きい場合、しばしば、適当なモデルでないという問題がある。第2は、事故要因が死傷事故を起こした車のデータのみから求められており、この値がすべての車から計算したものと異なる、すなわち、標本選択による偏りを生じている可能性である。本章では、これら2つの問題について理論的な考察を加える。

6.1 線形確率モデル

2章では、 $p_{ij} \sim x_{ij}\beta$ として線形確率モデルを分析に用いたが、現在では、線形確率モデルは一般的とはいえない。ここでは、誤差項 u_{ij} の分布が標準正規分布に従う場合に得られるプロビットタイプのモデルと比較し、モデルの妥当性について考察する。いま、誤差項 u_{ij} の分布が標準正規分布、事故要因を表す説明変数が

$$(6.1) \quad \begin{cases} x_{ij}a \sim N(\mu_i, \sigma_i^2), \\ \mu_i = E(x_{ij}), \\ \sigma_i^2 = V(x_{ij}a) \end{cases}$$

であるとする。(事故要因を表す説明変数はダミー変数であるが、それを $x_{ij}a$ の分布を変化させても得られる結論は正規分布を仮定した場合と大筋で変わらない。このため、ここでは、計算・表記を容易にするために正規分布を仮定する²⁾。)

この場合、第 i 車種において死亡事故が発生する確率 p_i は、

$$(6.2) \quad \begin{aligned} p_i &= E_x E_u(y_{ij}|x_{ij}) = E_x\{P(y_{ij}=1|x_{ij})\} \\ &= E_x\{\Phi(x_{ij}a)\} = \int_{-\infty}^{\infty} \Phi(x) \frac{1}{\sigma_i} \phi\left(\frac{x - \mu_i' a}{\sigma_i}\right) dx \\ &= \Phi(\mu_i' a / \sqrt{1 + \sigma_i^2}) \end{aligned}$$

である。ここで、 ϕ 、 Φ は標準正規分布の確率密度関数および累積分布関数であり、 E_u 、 E_x は u_{ij} および x_{ij} についての期待値を表す。いま、すべての車種において分散が等しく、 $\sigma_i^2 = \sigma^2$ 、 $i = 1, 2, \dots, k$ であるとする。 $\Phi(\mu_i' a / \sqrt{1 + \sigma^2}) = \Phi(\mu_i' a^*)$ 、 $a^* = a / \sqrt{1 + \sigma^2}$ となるから、この場合、(推定される係数の値は本来のもの $1/\sqrt{1 + \sigma^2}$ となるが) μ_i のみの情報から分析可能となる。また、事故の発生確率は一万台当たり 0.3~1.0 程度であるが、この範囲では $\Phi(z)$ は非常に良く直線近似され、線形近似による誤差は小さいと考えられる。したがって、推定された確率の評価には注意が必要であるが、どの事故要因が死亡事故に影響するかの分析は線形確率モデルを使って行うことが可能である。

また、 σ_i^2 が一定でない場合においても、その変動が $\mu_i' a$ に比較して著しく大きくない場合、線形確率モデルの妥当性について同様の分析を行うことができる。いま、 σ_i^2 が $\mu_i' a$ の関数であるとする。(ここでのデータでは P_i は小さく $\mu_i' a$ の値は負である。) σ_i^2 が $|\mu_i' a|$ の単調減少関数である場合、 P_i は $\mu_i' a$ の単調増加関数となる。 σ_i^2 が $|\mu_i' a|$ の増加関数である場合も、極端に変化しない場合は、 P_i は $\mu_i' a$ 単調増加関数となる。たとえば、

$$(6.3) \quad \sigma_i^2 = \gamma_2 \cdot |\mu_i' a| \quad \gamma_2 > 0$$

であるとする、

$$(6.4) \quad \mu_i' a / \sqrt{1 + \sigma_i^2} = \mu_i' a / \sqrt{1 + |\gamma_2 \cdot \mu_i' a|}$$

は、 $\mu_i' a$ の単調増加関数となり、 P_i は $\mu_i' a$ の単調増加関数となる。 P_i が $|\mu_i' a|$ の単調増加関数とならないためには、 σ_i^2 が $(\mu_i' a)^2$ より大きなオーダーで増加するなどの必要が生じる。

P_i が $\mu_i \alpha$ の単調増加関数ならば、ここで考えているデータの範囲では、分散が同一の場合と同様、事故要因の分析に関しては P_i の線形近似によって分析可能であると考えられる。さらに、 σ_i^2 に関して他の関数形を考慮しても、各係数の符号が逆転するためには分散が極端に変化することが必要となる。以上については x_{ij} がダミー変数であることを考慮してもほぼ同一の結果を得ることができる。本論文で扱ったのはセダン A クラスという同一タイプ・同一カテゴリーの車であり、その相関構造などが各車種によって大きく異なっているとは考えられず、線形確率モデルによって事故要因の分析を行うことが可能であると考えられる。

6.2 事故要因をあらわす変数の標本選択による偏りについて

ここでデータでは、事故要因が死傷事故を起こした車のデータのみから求められており、この値がすべての車から計算したものと異なる。このため、標本選択による偏りを生じている可能性があり、このことが事故要因の分析結果に影響を与えている可能性がある。ここでは、事故要因を表す変数が多変量正規分布に従うと仮定して、その影響について考察する。(前節と同様、事故要因を表す説明変数はダミー変数であるが、それを考慮しても得られる結論は正規分布を仮定した場合と大筋で変わらないと考えられる。)

いま、死傷事故を起こすか ($\eta_{ij}=1$) 起こさない ($\eta_{ij}=0$) かどうかを、

$$(6.5) \quad \begin{aligned} \eta_{ij}^* &= \gamma_0 + \gamma_1 x_{1j} + x_{2j}' \gamma_2 + \nu_{ij} \\ \eta_{ij} &= \begin{cases} 1 & \eta_{ij}^* > 0 \\ 0 & \eta_{ij}^* \leq 0 \end{cases} \end{aligned}$$

で与えられるとする。ここで、 ν_{ij} は x_{ij} と独立で標準正規分布に従う確立変数、 x_{1j} は事故要因の第 1 要因、 x_{2j} はその他の要因のベクトルで、 $x_{ij}' = (1, x_{1j}, x_{2j}')$ 、 $\gamma_0, \gamma_1, \gamma_2$ は未知のパラメータである。 $\mu_{i1} = E x_{1j}$ 、 $\mu_{i2} = E x_{2j}$ とすると、

$$(6.6) \quad \begin{cases} \eta_{ij}^* = \gamma_0^* + \gamma_1 x_{1j}^* + x_{2j}' \gamma_2 + \nu_{ij} \\ \quad = \gamma_0^* + w_{ij}, \\ \gamma_0^* = \gamma_0 + \gamma_1 \mu_{i1} + \mu_{2j}' \gamma_2, \\ x_{1j}^* = x_{1j} - \mu_{i1}, \\ x_{2j}^* = x_{2j} - \mu_{i2}, \\ w_{ij} = \gamma_1 x_{1j}^* + x_{2j}' \gamma_2 + \nu_{ij} \end{cases}$$

である。 $x_{ij}' = (x_{1j}, x_{2j}')$ の第 i 車種における分散共分散行列を Ω_i 、 $\gamma^* = (\gamma_1, \gamma_2')$ とすると、第 i 車種において、死傷事故が起こる確立は、

$$(6.7) \quad \begin{cases} P(\eta_{ij}=1) = E_x E_u(\eta_{ij}) = \Phi(\gamma_0^* / \sigma_{w,i}), \\ \sigma_{w,i}^2 = V(w_{ij}) = \gamma^{*'} \Omega_i \gamma^* + 1 \end{cases}$$

である。

ここで、 x_{ij}^* は多変量正規分布に従うから、 $x_{2j}' \gamma_2$ は

$$(6.8) \quad x_{2j}' \gamma_2 = a_i x_{1j} + \zeta_{ij}$$

と表すことができる。ここで、 ζ_{ij} は x_{1j} と独立な平均 0 の正規分布に従う確率変数である。(6.8) を (6.6) に代入すると、

$$(6.9) \quad \begin{cases} \eta_{ij}^* = \gamma_0^* + \xi_{ij} + \nu_{ij} \\ \xi_{ij} = (\gamma_1 + a_i)x_{i1j}, \\ \nu_{ij} = \nu_{ij} + \zeta_{ij} \end{cases}$$

を得る。
ここで、

$$(6.10) \quad \xi_{ij}^* = \begin{cases} \xi_{ij} & \eta_{ij} = 1 \\ 0 & \eta_{ij} = 0 \end{cases}$$

とすると、

$$(6.11) \quad \begin{cases} E(\xi_{ij}^* | \nu_{ij}) = \int_{-(\gamma_0^* + \nu_{ij})}^{-\infty} \frac{1}{\sqrt{2\pi}} \frac{\xi}{\sigma_{i,\xi}} \phi\left(\frac{\xi}{\sigma_{i,\xi}}\right) d\xi \\ = \sigma_{i,\xi} \phi\left(-\frac{\gamma_0^* + \nu_{ij}}{\sigma_{i,\xi}}\right), \\ \sigma_{i,\xi}^2 = V(\xi_{ij}) \end{cases}$$

である。 $\sigma_{i,\nu}^2 = V(\nu_{ij})$ とすると、 $\sigma_{i,w}^2 = \sigma_{i,\xi}^2 + \sigma_{i,\nu}^2$ であるから、

$$(6.12) \quad \begin{aligned} E\xi_{ij}^* &= \int_{-\infty}^{\infty} \sigma_{i,\xi} \phi\left(-\frac{\gamma_0^* + \nu}{\sigma_{i,\xi}}\right) \frac{1}{\sigma_{i,\nu}} \phi\left(\frac{\nu}{\sigma_{i,\nu}}\right) d\nu \\ &= \frac{\sigma_{i,\xi}^2}{\sigma_{i,w}} \phi\left(\frac{\gamma_0^*}{\sigma_{i,w}}\right) \end{aligned}$$

を得る。

ここで、 $E(x_{i1j}^* | \eta_{ij} = 1) = (1/\gamma_1^*) E\xi_{ij}^* / P(\eta_{ij} = 1)$ 、 $V(\xi_{ij}) = \gamma_1^{*2} V(x_{i1j})$ であるから、

$$(6.13) \quad \begin{cases} E(x_{i1j}^* | \eta_{ij} = 1) = \gamma_1^* \frac{\sigma_{i,1}^2}{\sigma_{i,w}} \lambda\left(\frac{\gamma_0^*}{\sigma_{i,w}}\right), \\ E(x_{i1j} | \eta_{ij} = 1) = \mu_{i1} + \gamma_1^* \frac{\sigma_{i,1}^2}{\sigma_{i,w}} \lambda\left(\frac{\gamma_0^*}{\sigma_{i,w}}\right) \end{cases}$$

$$\begin{aligned} \gamma_1^* &= \gamma_1 + a_i, \\ \sigma_{i,1}^2 &= V(x_{i1j}), \\ \lambda(z) &= \phi(z) / \Phi(z) \end{aligned}$$

である。したがって、 x_{i1j} には

$$(6.14) \quad b_{i1} = \gamma_1^* \frac{\sigma_{i,1}^2}{\sigma_{i,w}} \lambda\left(\frac{\gamma_0^*}{\sigma_{i,w}}\right)$$

だけの標本選択による偏りが生じていることになる。

すべての車種において同一の大きさの偏りがある場合は分析に影響はない。問題となるのは、標本選択による偏りによって、期待値の大きさの順序が変わってしまうこと、すなわち、

$$(6.15) \quad \mu_{i1} > \mu_{t1} \text{ であるにもかかわらず、 } E(x_{i1j} | \eta_{ij} = 1) < E(x_{t1j} | \eta_{ij} = 1)$$

となってしまうことである。したがって、標本選択による偏りが問題となるためには、

- i) 死傷事故の発生確率に特に大きな影響を与える要因があり、
 - ii) 事故要因の分散共分散構造が車種ごとに大きく異なる、
- という2条件が必要となり、極端な条件を想定する必要を生じる。

事故要因が1変数である場合、(6.15)となる条件を求めてみよう。 $\gamma_0 = -2\sqrt{2}$, $\gamma_1 = 0.1$ であり、第*i*車種において $\sigma_{i,1} = 1$, $\mu_{i1} = 0$ であるとする。(事故要因が1変数である場合、 $\gamma_{0i}^* = \gamma_0 + \gamma_1 \mu_{i1}$, $\gamma_i^* = \gamma_1$, $\sigma_{i,w}^2 = 1 + \sigma_{i,1}^2$ である。また、死傷事故の発生確率は1.50~2.63パーセントであるから、 $\gamma_0^*/\sigma_{i,w}$ は-2.17~-1.93であり、 $\lambda(\gamma_0^*/\sigma_{i,w}) = 2.31 \sim 2.52$ となる。)いま、第*t*車種において $\mu_{t1} = -0.01$ とすると、 $E(x_{i1j}|\eta_{ij} = 1) < E(x_{t1j}|\eta_{tj} = 1)$ となるためには、 $\sigma_{t,1} > 1.053$ であることが必要となる。また、 $\mu_{t1} = -0.1$ とすると、 $\sigma_{t,1} > 1.632$ であることが条件となっている。この例では、いずれの場合も平均の変動に比べて標準偏差は5倍以上変化しなければならないこととなる。

また、複数の事故要因が存在する場合においても、(6.15)のようになるためには、車種ごとに極端に異なった分散共分散の条件を考慮する必要を生じる。すなわち、前節の結論と同様、説明変数間の分散共分散構造が各車種によって特別な方向に大きく異なっていない限り、ここで用いたモデルによって事故要因の分析を行うことは妥当であると考えられる。

7. ま と め

本論文では、排気量が主に1500 cc以下である普通乗用車、セダンAクラスの乗用車について、死亡事故と事故要因についての関係の分析を行った。分析には財団法人交通事故総合分析センターによって作成された「交通事故統合データベース」を使い、各車種ごとに集計された28車種のデータを用いた。我が国においては、これまではこのようなデータベースは存在せず、本論文は死亡事故と事故要因の関連を分析した最初の論文である。死亡事故台数が少ない車種が存在することや登録台数が異なることを考慮し、分析にはポアソン分布に基づく回帰モデルを用いた。モデルの推定はポアソン分布に基づく最尤法および加重最小二乗法で行った。

事故要因としては、運転者の性別、年齢、通行の目的、シートベルト着用、夜間運転、速度、単独事故に関する9個の変数を考慮した。推定の結果、運転者が男性であること・運転者が年齢24以下であることは死亡事故発生率を減少させ、通行目的がドライブであること・単独事故であることは死亡事故発生比率を増加させることが認められた。また、通行目的が業務であること、夜間事故であることは有意水準5%では有意でないもののその係数は正となっている。加重最小二乗法による係数の推定値は最尤法によるものとよく似た結果となったが、標準誤差が大きく推定されており、この場合の最尤法による推定の重要性を示していると考えられる。単独事故比率の分析では、シートベルト着用率が負、夜間運転比率・運転速度41 km以上の比率が正の関係が認められた。また、ポアソン分布の妥当性および変数の外生性の検定を行い、最後に、線形確率モデルの妥当性および事故要因を表す変数の標本選択による偏りの影響についての考察を加えた。本分析では、得られるデータの制約上、事故要因は死傷者がいた場合の事故におけるデータを用いたが、死傷者のいない車のデータを加えた方がより一般的な車の利用を表すであろうし、(分析への影響はそれほど大きくないと考えられるものの)標本選択による偏りを生じるおそれもない。データの収集を含め、今後はこのような観点からの分析が必要となる。また、セダンAと異なる他のクラスの乗用車についても、死亡事故と事故要因の関係を分析することも今後の課題である。

〈注〉

- 1) 単独事故とは、最初の事故誘発行為が他の車両や歩行者、列車と関係せず、衝突相手が駐車車両や物体である場合あるいは衝突相手がない場合の事故である。事故車両のみによって引き起こされた事故であり、運転者の注意によって防げると考えられる事故である場合が多い。
- 2) 期待値等の計算には、標準正規分布の確率密度関数、累積分布関数の積分に関する次の公式を用いた。

$$\int_{-\infty}^{\infty} \phi(ax+b)\phi(x)dx = \frac{1}{\sqrt{1+a^2}} \phi\left(\frac{b}{\sqrt{1+a^2}}\right)$$

$$\int_{-\infty}^{\infty} \Phi(ax+b)\phi(x)dx = \Phi\left(\frac{b}{\sqrt{1+a^2}}\right)$$

$$\int_{-\infty}^{\infty} x\Phi(ax+b)\phi(x)dx = \frac{b}{\sqrt{1+a^2}} \phi\left(\frac{b}{\sqrt{1+a^2}}\right)$$

参 考 文 献

- [1] Cameron A. C., and P. Johansson, 1997, "Count Data Regression using Series Expansions with Applications," *Journal of Applied Econometrics*, 12, 203-223.
- [2] Cameron, A. C., and P. K., Trivendi, 1990, "Regression-based Tests for Overdispersion in the Poisson Model," *Journal of Econometrics*, 46, 347-364.
- [3] Cameron, A.C., and P.K. Trivedi, 1986, 'Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests,' *Journal of Applied Econometrics*, 1, 29-54.
- [4] Collings, B. J., and Margolin B. H., 1985, "Testing Goodness of Fit for Poisson Assumption when Observations are not Identically Distributed," *Journal of the American Statistical Association*, 80, 411-418.
- [5] Consul, P.C. and F. Famoye, 1992, "Generalized Poisson Regression Model," *Communications in Statistics: Theory and Method*, 21, 89-109.
- [6] Cox, D. R., 1983, "Some Remarks on Overdispersion," *Biometrika*, 70, 269-74.
- [7] Dean, C., and J. F. Lawless, 1989, "Tests for Overdispersion in Poisson Regression Models," *Journal of the American Statistical Association*, 84, 467-471.
- [8] Department of Transportation, United Kingdom, 1992, *Cars: Make and Models: The Risk of Driver Injury and Car Accident Rates in Great Britain*.
- [9] Gourieroux, C., A. Montfort, and A. Trognon, 1984, "Pseudo Maximum likelihood Methods: Applications to Poisson Models" *Econometrica*, 52, 701-720.
- [10] Hausman, J., 1978, "Specification Tests in Econometrics," *Econometrica*, 46, 1251-1271.
- [11] Hausman, J., B. H. Hall and Z. Griliches, 1984, "Econometric Models for Count Data with an Application to the Patents - R&D Relationships, *Econometrica*, 52, 909-938.
- [12] 交通事故総合分析センター, 「交通事故と運転者と車両の相関関係についての分析結果」, 1997.
- [13] Lee, L., "Specification Test for Poisson Regression Models," 1986, *International Economic Review*, 27, 689-706.
- [14] McCullagh, P., and J. A. Nelder, 1983, *Generalized Linear Models*, Chapman and Hall, London.
- [15] Model, 1993, "The Effect of Marijuana Decriminalization on Hospital Emergency Room During Esisodes: 1975-1978," *Journal of American Statistical Association*, 88, 737-747.
- [16] 縄田和満, 「TSPによる計量経済分析入門」, 朝倉書店, 1997.
- [17] Sander, W., 1992, "The Effects of Women's Schooling on Fertility," *Economic Letters*, 40, 229-233.
- [18] TSP International, 1996, *TSP Reference Manual*.
- [19] 和合肇・伴金美, 「TSPによる経済データの分析 [第2版]」, 東京大学出版会, 1995.
- [20] Wooldridge, J. R., 1991, "Specification Test and Quasi-maximum-likelihood Estimation," *Journal of Econometrics*, 48, 29-55.
- [21] Winkelmann, R., 1994, *Count Data Models: Econometric Theory and an Application to Labor Mobility*, Springer-Berlin.