

偶然誤差を考慮した事故率分析の区間設定方法

岡本 博*

回帰分析を用いた交通事故の要因分析に関して、事故率の持つ偶然誤差の大きさの評価方法並びに偶然誤差が回帰係数推定量と決定係数に与える影響を理論的に明らかにした上で、一連の道路での事故データ等が与えられたとき、偶然誤差の影響を抑えて有効な分析を行うための分析用区間の設定方法を提示する。東名・名神高速道路での事故データ等の分析により、この方法の有用性を検証した。

Key Words: traffic accident analysis, regression, accident rate, random errors, segmentation, error interval, error ratio

1. はじめに

(1) 本研究の目的

交通事故の発生と道路・交通状況との関連を分析する場合に、道路をいくつかの区間に分割し、その中で発生した事故の事故率（走行台キロ当たりの事故件数）と道路・交通等の区間特性との相関をクロス集計や回帰モデル等により分析する方法が多く用いられる。ところが、事故はごく希に起きる現象であるため、たまたま起きた（あるいは起きなかった）1件が、事故率に大きな影響を与える。このような大きな偶然誤差が存在することによって、観測された事故率と回帰モデルによる予測値との間に、偶然誤差とモデルの説明力不足の両方からなる大きな残差が生じ、現況再現力の高い分析結果を得ることが困難となっている。

事故率の持つ偶然誤差は、事故率を算定する区間の取り方に依存する。即ち、事故率を求める区間を短くすると、たまたま起きた1件の事故による偶然誤差のウェイトが高くなる。一方で、区間を長くとると多くの事故を平均化して偶然誤差は小さくなるが、事故の多い地点と少ない地点が1つの区間に取り込まれ、区間特性の違いによる事故発生の違いが分析できなくなる。従って、事故率の分析では、対象とする道路の分析区間の設定方法が、大きな意味を持つこととなる。

本研究は、事故率算定のための区間の取り方に着目し、偶然誤差の影響を抑えて有効な事故要因分析を行うための区間の設定方法を提示することを目的としている。

(2) 本研究の位置付けと基本的考え方

重回帰による事故要因分析の一般的なフローは、図1で表される。これは、分析用に区間を設定し、被説明変数である事故率を算定する「区間設定ステップ」と、説明変数を考えて回帰モデルを設定し（図中※1）、決

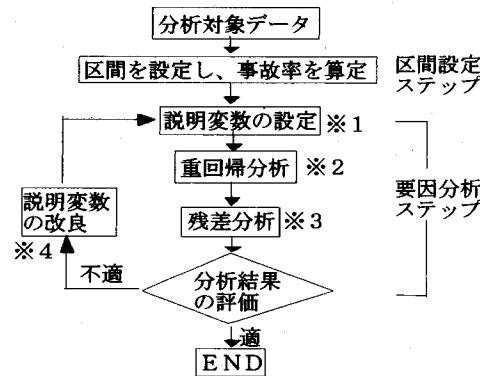


図-1 事故要因分析のフロー

定係数（重相関係数）を指標として変数選択を行い（図中※2），更に残差を分析する（図中※3）ことによって、説明変数を改良しモデルの説明力を上げていく（図中※4）「要因分析ステップ」からなる。

従来の事故要因分析に関する研究では、多くの場合、要因分析ステップでの説明変数の改善に力が注がれていた¹⁾が、現況再現力の高いモデルは得られていない。また、偶然誤差の影響を小さくするため、区間の設定方法を変える試みもなされている^{2), 3)}が、理論的根拠を持たず、試行錯誤の域をでていない。

要因分析ステップを有効なものとするには、区間設定ステップでは次の3要件が求められる。

- ・説明変数の選択等で重要な指標となる決定係数（重相関係数）に対して偶然誤差の影響が小さいこと。

（偶然誤差の影響が大きいと、適切な説明変数の選択ができないと同時に、回帰モデルの現況再現力が低いことが、説明変数が不適切なためか、偶然誤差によるためかが区別できない。）

- ・回帰分析の結果である回帰係数推定量の分散が小さいこと
- ・モデル改善の重要な手法である残差分析が有効に行え

* 岡本 博 建設省 関東地方建設局相武国道工事事務所長（〒192 東京都八王子市大和田町4-3-13）

ること、そのために、個々の事故率に含まれる偶然誤差が小さいこと。

(残差の大きな部分が偶然誤差では、新たな説明変数を導入して説明を試みることは、意味がないばかりでなく、誤った結論を導く)

そこで本研究では、決定係数及び回帰係数推定量に対する偶然誤差の影響を理論的に明らかにした上で、上述の3要件を満たす分析区間の設定方法を提案し、実際のデータにより検証する。

2. 基本概念

(1) 事故率・危険率・偶然誤差

長さ l の区間があり、この中の危険率（単位延長あたり事故発生確率）を p とするとき、 m 台の車両が通過する場合の事故の発生件数 k は、次式のポアソン分布で近似できる。

$$\Pr(k|\lambda) = \frac{\lambda^k}{k!} e^{-\lambda}, \lambda = mlp \quad \dots \dots \dots \quad (1)$$

着目する区間で通過車両 m 台に対して事故が k_o 件発生したとする。最尤推定法により p の点推定量 \hat{p} を求めるとき、次式のとおりとなる。

$$\hat{p} = k_o / ml \quad (= p_o) \quad \dots \dots \dots \quad (2)$$

$$\therefore \text{尤度関数 } L(k_o | p) = \frac{(mlp)^{k_o}}{k_o!} e^{-mlp}$$

また、 k_o の期待値は λ であるから、 k_o / ml の期待値は p となる。故に、事故率 $p_o = k_o / ml$ （走行台^{*}当りの事故件数）は、未知定数である危険率 p の最尤不偏推定量として意味付けられる。

事故率の持つ偶然誤差 v を両者の差と定義する。（単位は走行台^{*}当りの事故件数）

$$v = p_o - p \quad \dots \dots \dots \quad (3)$$

偶然誤差 v の平方の期待値（以下「偶然分散」と呼ぶ）を ϕ^2 で表すことになると、次式が得られる。

$$\phi^2 = E(v^2) = \frac{E[(\lambda - k_o)^2]}{m^2 l^2} = \frac{p}{ml} \quad \dots \dots \dots \quad (4)$$

(2) 単位区間と要素区間

分析対象の道路について、データ収拾の単位となる短い区間（例えば100m単位等）毎に事故件数と交通量が得られている場合を考える。以下、この区間を「単位区間」と呼ぶ。個々の単位区間の事故率では偶然誤差が大き過ぎるために、隣接する幾つかの区間をまとめて分析用のサンプルの1つの要素となる区間を作る。以下、この区間を「要素区間」と呼ぶ。

共通の記述方法として、図-2に示すように、ある指標 a の第 i 要素区間としての値を a_i とし、第 i 要素区間の中で第 j 番目の単位区間の値を $j a_i$ と記す。 h_i は第 i 要素区間内の単位区間数を表す。このとき、事故率・危険率・偶然誤差等の単位区間での値と要素区間での値は、表-1となる。ここで、交通量 m_i は、要素区間内で一様

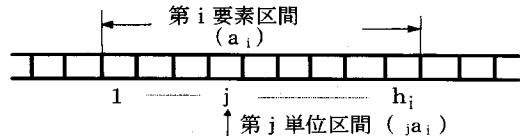


図-2 単位区間と要素区間

表-1 各指標値の単位区間と要素区間での値

| 項目 | 単位区間での値 | 要素区間での値 |
|------------------|--|--|
| 事故件数 | $\sum_j k_{oi}$ | $k_{oi} = \sum_j k_{oi}$ |
| 走行台 [*] | $m_i \delta$ | $m_i h_i \delta$ |
| 事故率 | $\sum_j p_{oi} = \frac{\sum_j k_{oi}}{m_i \delta}$ | $p_{oi} = \frac{1}{h_i} \sum_j p_{oi}$ |
| 危険率 | $\sum_j p_i$ | $p_i = \frac{1}{h_i} \sum_j p_i$ |
| 偶然誤差 | $\sum_j v_i$ | $v_i = \frac{1}{h_i} \sum_j v_i$ |
| 偶然分散 | $\sum_j \phi_i^2 = \frac{\sum_j p_i}{m_i \delta}$ | $\phi_i^2 = \frac{1}{h_i} \sum_j \phi_i^2$ |

δ は単位区間の長さ

とする。

(3) 要素区間の設定と事故率の分散

単位区間での事故率の分散 ST は、危険率の分散 PT と偶然誤差の平方 RT 及び分散の中心を事故率の平均に取ることによる調整項 DT から成る。この単位区間をまとめて要素区間を構成するとき、事故率の分散は、要素間の分散 SI と要素内の分散 SJ に分解され、前者だけが分析対象として抽出される。これを整理すると、表-2及び式(5)となる。単位区間の総数 N が十分大きいときは、調整項 DT は小さく、無視できる。

$$ST = PT + RT + DT \approx PI + RI + PJ + RJ \quad \dots \dots \dots \quad (5)$$

PI は要素区間の設定によって定まる PT の内訳である。要素区間の設定による危険率の抽出特性を表す指標として、単位区間での危険率分散 PT に対する要素区間での危険率分散 PI の割合を「有効危険率比 θ 」と定義する。

$$\theta = PI / PT \quad (\text{有効危険率比}) \quad \dots \dots \dots \quad (6)$$

要素区間の設定によって危険率分散 PI と共に偶然分散 RI も取り出される。 RI は確率変数で、その期待値は、

$$E(RI) = \frac{1}{N} \sum_i h_i \phi_i^2 = \phi_m^2 \quad \dots \dots \dots \quad (7)$$

ϕ_m^2 は、要素区間での偶然分散の、単位区間数重み付平均を表している。以下、「要素偶然分散」と呼ぶ。特に、単位区間の偶然分散の平均を全偶然分散 ϕ_{Nm}^2 と呼ぶ。

設定された要素区間の特性を偶然誤差の面から表す指標として、要素区間での偶然分散 ϕ_m^2 の危険率分散 PI に対する比を考えて、「偶然分散比 μ 」と定義する。

$$\mu = \phi_m^2 / PI \quad (\text{偶然分散比}) \quad \dots \dots \dots \quad (8)$$

式(4)をもとに、 p の最尤不偏推定量である p_o を用い、要素区間の偶然分散 ϕ_i^2 の推定量 $\hat{\phi}_i^2$ を次式で与える。

$$\hat{\phi}_i^2 = \frac{p_{oi}}{m_i h_i \delta} = \frac{k_{oi}}{m_i^2 h_i^2 \delta^2} \quad \dots \dots \dots \quad (9)$$

表-2 事故率分散の構成

| | T : 単位区間での分散 | I : 要素間の分散 | J : 要素内の分散 |
|-------------|--|---|--|
| S : 事故率の分散 | $ST = \sum_i \sum_j (p_{oi} - p_{om})^2 / N$ | $SI = \sum_i h_i (p_{oi} - p_{om})^2 / N$ | $SJ = \sum_i \sum_j (p_{oi} - p_{oi})^2 / N$ |
| P : 危険率の分散 | $PT = \sum_i \sum_j (p_i - p_m)^2 / N$ | $PI = \sum_i h_i (p_i - p_m)^2 / N$ | $PJ = \sum_i \sum_j (p_i - p_i)^2 / N$ |
| R : 偶然誤差の平方 | $RT = \sum_i \sum_j v_i^2 / N$ | $RI = \sum_i h_i v_i^2 / N$ | $RJ = \sum_i \sum_j v_i^2 / N - RI$ |
| D : 調整項 | $DT = \frac{2}{N} \sum_i \sum_j h_i (p_i - p_m) - v_m^2$ | $DI = \frac{2}{N} \sum_i h_i v_i (p_i - p_m) - v_m^2$ | $DJ = \frac{2}{N} \sum_i \sum_j h_i v_i (p_i - p_i)$ |

$$\text{注} \quad p_{om} = \frac{1}{N} \sum_i \sum_j p_{oi}, \quad p_m = \frac{1}{N} \sum_i \sum_j p_i, \quad v_m = \frac{1}{N} \sum_i \sum_j v_i$$

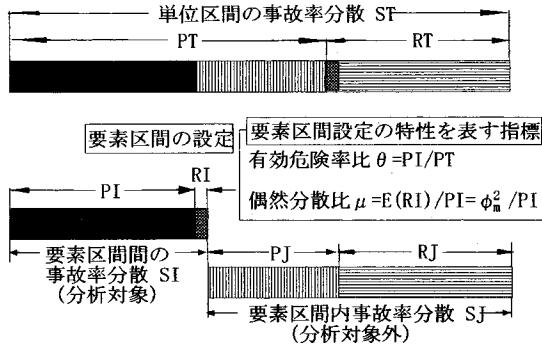


図-3 要素区間の設定による分析対象の抽出

要素偶然分散 ϕ_m^2 の推定量 $\hat{\phi}_m^2$ は次式となる。

$$\hat{\phi}_m^2 = \frac{\sum_i h_i \hat{\phi}_i^2}{N} = \frac{1}{N} \sum_i \frac{k_{oi}}{m_i^2 h_i^2 \delta} \quad \dots \dots \dots (10)$$

このとき

$$E(\hat{\phi}_m^2) = \frac{\sum_i h_i E(\hat{\phi}_i^2)}{N} = \frac{1}{N} \sum_i h_i \phi_i^2 \quad \dots \dots \dots (11)$$

$$Var(\hat{\phi}_m^2) = \frac{1}{N} \sum_i h_i \frac{h_i}{N} Var(\phi_i^2) \quad \dots \dots \dots (12)$$

上式によれば、要素偶然分散の推定量 $\hat{\phi}_i^2$ は不偏推定量であり、その分散は、 $Var(\hat{\phi}_i^2)$ の h_i/N を延長重み付き平均したもので、推定精度が高いことが分かる。これを用いて推定偶然分散比 $\hat{\mu}$ 及び推定有効危険率比 $\hat{\theta}$ を考えると、次式のように $PT, \phi_{Nm}^2, \theta, \mu$ は、観測されたデータから推定できる。これらの指標を用いると、図-3に示すように、与えられた単位区間データ全体の特性は PT, ϕ_{Nm}^2 によって表され、要素区間の設定による分析対象の抽出は、 θ, μ によって表される。

$$\hat{\theta} = \frac{PI_h}{PT_h} = \frac{SI - \hat{\phi}_m^2}{ST - \hat{\phi}_{Nm}^2} \quad \dots \dots \dots (13)$$

$$\hat{\mu} = \frac{\hat{\phi}_m^2}{PI_h} = \frac{\hat{\phi}_m^2}{SI - \hat{\phi}_m^2} \quad \dots \dots \dots (14)$$

3. 事故率説明モデル

(1) 単位区間重回帰モデルの設定

N 個の単位区間にについて、危険率 p と道路条件を表す $k-1$ 個の説明変数 \mathbf{x} との間に、次の線形関係が成立つと仮定する。 \mathbf{u} は p のうち $x_2 \dots x_k$ では説明しきれない部分を表す攪乱項で、以下、「方程式のはずれ」と呼ぶ。

\mathbf{u} の分布の母数及び回帰係数 \mathbf{b} は未知である。

$$\mathbf{p}_N = \mathbf{X}_N \mathbf{b} + \mathbf{u}_N \quad \dots \dots \dots (15)$$

$$\mathbf{p}_N = \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$

$$\mathbf{X}_N = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix}, \quad \mathbf{u}_N = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$$

方程式のはずれ \mathbf{u}_N について、分散が一様で無相関であるとする最も簡単な仮定を置く。

$$E(\mathbf{u}_N) = \mathbf{0} \quad \dots \dots \dots (16)$$

$$Var(\mathbf{u}_N) = \sigma^2 \mathbf{I}_N \quad (\mathbf{I}_N \text{ は単位行列}) \quad \dots \dots \dots (17)$$

各単位区間の事故率 p_o は、危険率 p と偶然誤差 v を用いて次式で表される。

$$\mathbf{p}_{oN} = \mathbf{p}_N + \mathbf{v}_N \quad \dots \dots \dots (18)$$

$$\text{ここで } \mathbf{p}_{oN} = \begin{pmatrix} p_{o1} \\ \vdots \\ p_{on} \end{pmatrix}, \quad \mathbf{v}_N = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

$$E(\mathbf{v}_N) = \mathbf{0} \quad \dots \dots \dots (19)$$

$$Var(\mathbf{v}_N) = \Phi_N = \begin{pmatrix} \phi_1^2 & & & \\ & \ddots & & \\ & & \ddots & \\ 0 & & & \phi_n^2 \end{pmatrix} \quad \dots \dots \dots (20)$$

ϕ_i^2 は単位区間毎の偶然誤差の分散
(偶然誤差は相互に無相関と仮定)

方程式のはずれと偶然誤差は無相関と仮定すると、事故率の単位区間での重回帰モデルは次式で表される。

$$\mathbf{p}_{oN} = \mathbf{X}_N \mathbf{b} + \mathbf{u}_N + \mathbf{v}_N \quad \dots \dots \dots (21)$$

$$E(\mathbf{u}_N + \mathbf{v}_N) = \mathbf{0} \quad \dots \dots \dots (22)$$

$$Var(\mathbf{u}_N + \mathbf{v}_N) = \sigma^2 \mathbf{I}_N + \Phi_N \quad \dots \dots \dots (23)$$

(2) 要素区間モデルへの変換

N 個の単位区間のうち隣接する h_i 個をまとめて合計 n 個の要素区間とする。次式のような n 行 N 列のグループ分け行列 \mathbf{G} を考え、この \mathbf{G} を単位区間での回帰モデルの両辺に左から乗じれば、要素区間での重回帰モデルが得られる。即ち、

$$\mathbf{G} = \begin{pmatrix} h_1 \text{ 個} & & \\ 1/h_1 \dots 1/h_1 & 0 & \\ \vdots & \ddots & \\ 0 & 1/h_n \dots 1/h_n & \end{pmatrix} \quad \dots \dots \dots (24)$$

率分析の決定係数に近いものとできる。

- ・危険率分析の決定係数は、同じ説明変数を用いたモデルでも要素区間の取り方によって異なる値となる。これは、平均単位区間数 h_m の違いと、有効危険率比 θ の違いの2つの要因による。 σ^2/h_m は要素区間レベルでの方程式のはずれの分散を表し、 h_m を大きくして θh_m も大きくできれば、決定係数が高くなるが、これは見かけ上のものである。一方、同じ h_m でも要素区間の取り方によって θ が大きいと、決定係数が高くなるが、これは、分析対象の危険率分散を巧く抽出できていることを示している。

(5) 回帰係数の推定に対する偶然誤差の影響

分析の結果得られる回帰係数推定量の良さは、その分散の小ささで評価される。回帰係数推定量を用いた予測値の分散の、全単位区間での平均値を考え、平均予測値分散 I_{EM} と定義する。この I_{EM} は重回帰モデルの回帰係数推定量の分散の大きさを表す指標となる。

$$I_{EM} = \frac{1}{N} \sum_{i=1}^n Var(\mathbf{c} \mathbf{b}_h) \quad \dots \dots \dots \quad (39)$$

説明変数が1つの場合を考えると

$$I_{EM} = \frac{\sigma^2}{N} \left(1 + \frac{1}{\gamma_x} \right) + \frac{\phi_{Nm}^2}{N} \left(1 + \frac{z_{dm}^2}{\gamma_x} \right) \quad \dots \dots \quad (40)$$

〈方程式のはずれの影響〉 〈偶然誤差の影響〉

$$\text{ここに } \gamma_x = \frac{\sigma_x^2}{\sigma_{xv}^2} = \frac{\sigma_x^2 \cdot N}{\sum_{i=1}^n \sum_{j=1}^m x_i^2} \quad \dots \dots \quad (41)$$

I_{EM} は、方程式のはずれの影響と偶然誤差の影響の和となる。上式のうち、 N は単位区間数、 σ^2 は方程式のはずれの分散、 ϕ_{Nm}^2 は全偶然分散であり、いずれも要素区間の取り方には依らない。 γ_x は説明変数 x の単位区間での分散に対する、要素区間として抽出された分散の割合で、以下、有効説明変数比と呼ぶ。

回帰係数推定量の分散を小さくするように要素区間を設定するためには、 γ_x を大きくする ($1/\gamma_x$ を小さく保つ) ことが必要となる。紙面の都合で省略するが、説明変数が複数の場合でも、同様に、説明変数の分散を大きくすることが求められる。

4. 方程式のはずれの系列相関

(1) 系列相関を考えた重回帰モデル

重回帰モデルとして想定した式は、限られた変数による単純な式形となるため、方程式のはずれは、取り入れられなかった変数あるいは式形の不備の影響を表す。一般に道路構造は滑らかであるため、短い区間での道路構造に関する変数は系列相関を持ち、方程式のはずれも系列相関を持つこととなる。

ここで、方程式のはずれ u が次式に示す1階の自己回帰過程に従うとした場合について考察する。 ρ は単位区間での方程式のはずれの系列相関の大きさを表す。

$$u_t = \rho u_{t-1} + \varepsilon_t \quad \dots \dots \dots \quad (42)$$

ただし $0 \leq \rho < 1$

$$E(\varepsilon_t) = 0, E(\varepsilon_t^2) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0 \quad (s \neq t)$$

重回帰モデルの分散共分散行列は次式となる。

$$Var(\mathbf{u}_N + \mathbf{v}_N) = \sigma^2 \Omega_N + \Phi_N \quad \dots \dots \dots \quad (43)$$

ここに $\sigma^2 = \sigma_\varepsilon^2 / (1 - \rho^2)$

$$\Omega_N = \begin{pmatrix} 1 & \rho & \cdots & \rho^{N-1} \\ \rho & 1 & \cdots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \rho^{N-1} & \cdots & 1 \end{pmatrix} = (\rho^{|t-s|})$$

(2) 決定係数

要素区間事故率の決定係数を考えると

$$R^2 \approx \frac{1}{1+\mu} \left(1 - \frac{\sigma^2 \bar{\omega}_{..}}{PT \theta h_m} \right) \quad \dots \dots \dots \quad (44)$$

$$\text{ここに } \bar{\omega}_{..} = \frac{1+\rho}{1-\rho} - \frac{2\rho}{(1-\rho)^2} \frac{1}{n} \sum \frac{1-\rho^{h_i}}{h_i}$$

後述の東名・名神でのケーススタディにおいて様々な要素区間の取り方について試算したところによれば、 $\bar{\omega}_{..}$ は要素内平均単位区間数 h_m を用いて次式で近似でき、 h_m と ρ の関数となる。

$$\bar{\omega}_{..} \approx \frac{1+\rho}{1-\rho} - \frac{2\rho}{(1-\rho)^2} \frac{1-\rho^{h_m}}{h_m} \quad \dots \dots \dots \quad (45)$$

前章と同様に式(44)の第1因子は偶然誤差の影響を表し、第2因子は要素区間危険率に対する回帰分析での決定係数となっている。系列相関を考えない場合と異なっているのは、第2因子で、方程式のはずれの分散 σ^2 に $\bar{\omega}_{..}$ が掛かっている点のみである。

決定係数に対する偶然誤差の影響を抑えるためには、偶然分散比 μ を小さくすることが求められる。

(3) 回帰係数推定量の分散

回帰係数推定量の分散に関しては、系列相関の影響は、方程式のはずれの部分に現れ、次の通りとなる。

$$I_{EM} = \frac{\sigma^2 \omega_m}{N} \left(1 + \frac{zz_{am}}{\gamma_x} \right) + \frac{\phi_{Nm}^2}{N} \left(1 + \frac{z_{dm}^2}{\gamma_x} \right) \quad \dots \dots \dots \quad (46)$$

$$\text{ここに } \omega_m = \frac{1}{N} \sum_{i=1}^n \sum_{s=1}^m \sum_{l=1}^k \sum_{k=1}^l \rho^{|r_i - r_s + k - l|} \approx \frac{1+\rho}{1-\rho}$$

$$zz_{am} = \sum_{i=1}^n \sum_{s=1}^m \sum_{l=1}^k \sum_{k=1}^l \rho^{|r_i - r_s + k - l|} \frac{x_i x_s}{\sigma_x^2} \cdot \frac{1}{N \omega_m}, \quad r_i = \sum_{s=1}^m h_i$$

上式の ω_m は系列相関による σ^2 の影響の拡大を示し、 N, σ^2, ϕ_{Nm}^2 と同様に、要素区間の取り方には依らない。

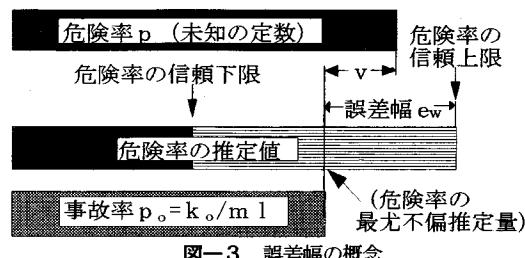
zz_{am}, z_{dm}^2 は要素区間の取り方に依存するが、未知数で、説明変数にも依り、要素区間設定の段階では考慮することはできない。

回帰係数推定量の分散を小さくするように要素区間を設定するためには、前章の結論と同じく、 γ_x を大きくする ($1/\gamma_x$ を小さく保つ) ことが求められる。

5. 要素区間の設定方法

(1) 事故率の偶然誤差の評価

第1章で述べたように、要素区間の設定には、モデル



改善の重要な手法である残差分析が有効に行えることが求められる。そのためには、個々の要素区間事故率に含まれる偶然誤差が、一定の範囲に納まっていることが必要となる。そこで、要素区間での偶然誤差の大きさを評価し、これが一定の範囲内に納まるよう要素区間を設定することを考える。

個々の事故率の偶然誤差 v の大きさを評価するため、ネイマンの信頼区間推定法を用いる。事故件数のボアソン分布は、 $\lambda > 5$ のとき正規分布 $N(\lambda, \lambda)$ で近似でき、次式が成り立つ。

$$\Pr(\lambda - z_{\alpha} \sqrt{\lambda} < k_0 < \lambda + z_{\alpha} \sqrt{\lambda} | \lambda) = 1 - \alpha \quad \dots \dots \dots (47)$$

ここに $1 - \alpha$ は信頼係数

この式の () 内の λ と k_0 を入れ換える、更に ml で除して偶然誤差 v の信頼区間を得る。 v の信頼限界のうち、絶対値の大きい方（信頼上限がこれにあたる）を用いて、個々の事故率の持つ偶然誤差の大きさを表す指標とする。これを誤差幅 e_w と呼ぶ。即ち

$$e_w = \frac{z_{\alpha}^2 + z_{\alpha} \sqrt{4k_0 + z_{\alpha}^2}}{2ml} \quad \dots \dots \dots (48)$$

図-3に誤差幅の概念を示す。

更に、事故率に対する偶然誤差の相対的な大きさを表すため、誤差率 e_r を次式のように定義する。

$$e_r = \frac{e_w}{p_0} = \frac{z_{\alpha}^2 + z_{\alpha} \sqrt{4k_0 + z_{\alpha}^2}}{2k_0} \quad \dots \dots \dots (49)$$

誤差率 e_r は事故件数 k_0 だけの関数となる。

(4) 要素区間の設定基準

a) 誤差幅基準

すべての事故率の誤差幅 e_w がある許容誤差幅 e_{wa} 以下となるように区間長を調整して、要素区間を設定する基準を考え、これを誤差幅基準と呼ぶ（図-4 a）。

$$e_w \leq e_{wa} \quad \dots \dots \dots (50)$$

式(48)より、この基準では、高い事故率の要素区間となるためには大きな走行台キロが必要となり、事故多発区間では要素区間長を長く取らなければならない。ところが事故多発区間はあまり長く続かないため、前後の事故の少ない区間を取り込んで一つの要素区間にせざるを得なくなる。つまり、事故多発区間を要素区間として上手く取り出すことが困難な基準となる。

b) 誤差率基準

相対値で見て、すべての事故率の誤差率 e_r がある許容誤差率 e_{ra} 以下となるように要素区間を設定する基準

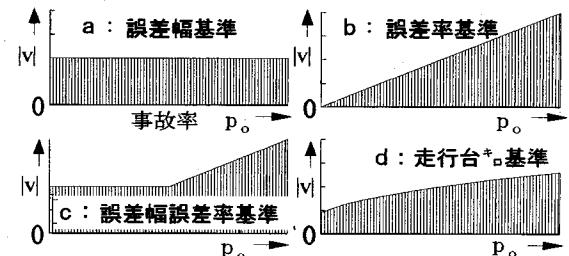


図-4 各基準での偶然誤差の範囲

を考え、これを誤差率基準と呼ぶ（図-4 b）。

$$e_r \leq e_{ra} \quad \dots \dots \dots (51)$$

この基準では、要素区間内の事故件数だけが問題となり、事故多発区間はそれだけで一つの要素区間となる。一方、事故の少ない区間に对しては偶然誤差の範囲を非常に小さいものとする制約が働き、前後の事故の多い区間を取り込んで一つの要素区間となる。

c) 誤差幅誤差率基準

これらの短所を補うものとして、許容誤差幅 e_{wa} 及び許容誤差率 e_{ra} を定め、偶然誤差がどちらかの範囲内に入るように要素区間を設定する基準を考え、これを誤差幅誤差率基準と呼ぶ（図-4 c）。すなわち

$$e_w \leq e_{wa} \quad or \quad e_r \leq e_{ra} \quad \dots \dots \dots (52)$$

この基準によれば、事故率の偶然誤差を一定の範囲に抑えつつ、事故の多い区間も少ない区間もそれぞれ無理なく一つの要素区間とすることができる。これが、本研究で最も有用なものとして提案する基準である。

d) 走行台キロ基準

誤差幅誤差率基準では、要素区間の設定がそれぞれの要素区間内の事故件数に依存することとなる。実際に適用する場合を考えると、事故の分布状況に応じて区間長を変化させながら要素区間を設定することはかなり面倒となる。そこでこれを簡略化し、要素区間内の事故件数に係わらず事故率の偶然誤差が一定の範囲内に入っていることを保証する基準を考える。これは走行台キロに関する基準として考えることができる。すなわち、

$$ml \geq M_a \quad \dots \dots \dots (53)$$

このとき、事故率の偶然誤差の範囲の概念は、(48)式より図-4 d のようになる。

走行台キロ基準の許容誤差幅は、誤差幅誤差率基準の許容誤差幅より小さくなり、事故率の小さい方も大きい方もそれぞれ厳しい基準となる。

(3) 分析に適した要素区間の設定方法

前節の基準を用いれば、1つの基準値に対して、個々の事故率の偶然誤差がその基準値の範囲内におさまるような1つの分析用サンプル（設定された要素区間のセット）が得られる。

第2章で示したように、要素区間の設定による分析対象の事故率分散の抽出のされ方は、偶然分散比 μ と有効危険率比 θ で表される。設定される要素区間全体に対し

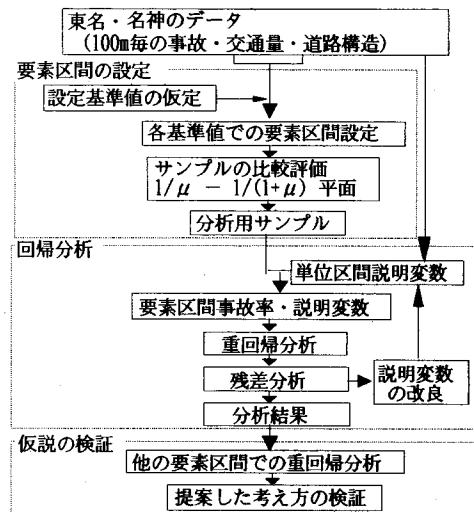


図-5 分析の流れ

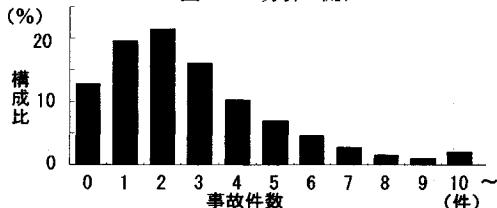


図-6 単位区間での事故件数ヒストグラム

では、第3・4章の重回帰モデルに関する考察により、次の2点が求められる。

- ・決定係数への偶然誤差による影響（低減及びバラツキ）を小さくし、危険率分析の決定係数に近い決定係数を得るという視点からは、 $1/(1+\mu)$ が大きいこと。
- ・回帰係数推定量の分散を小さくすると言う視点からは、 $1/\gamma_x$ が小さいこと。ただし、本研究では、説明変数を確定する以前で有効説明変数比 γ_x を直接用いることができない場合を考え、代りに有効危険率比 θ を用いる。 $1/\theta$ が小さいことは、原データの危険率の分散のうちできるだけ多くの部分を分析対象とすべきであることを示している。

そこで、分析対象となる単位区間にに対して、幾つかの基準値を仮定してそれぞれサンプルを作成した上で、これらの複数のサンプルを偶然分散比 μ 及び有効危険率比 θ を用いて $[1/\theta] - [1/(1+\mu)]$ 平面上で比較評価すれば、決定係数及び回帰係数推定量の分散から見て最も有効に危険率の分析が行なえるサンプルを選定することができる。これが、分析に適した要素区間を設定する方法である。

6. ケーススタディ (東名・名神高速道路の事故分析)

(1) 分析の流れ

実際の事故データ及び道路構造、交通量データを用い

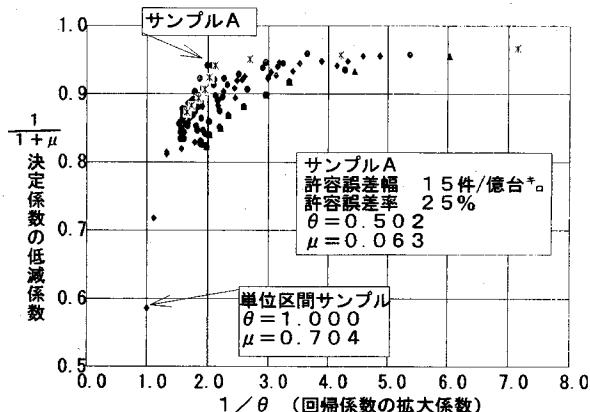


図-7 サンプルの評価平面

て重回帰分析を行ない、提案した方法の有効性を実証する。分析の流れを図-5に示す。

(2) 分析対象

分析対象道路は、東名・名神高速道路の厚木・西宮間(約500km)で、分析対象事故は、昭和45年から50年の6年間にこの区間の高速道路本線上で発生した事故の中から、道路構造との関連を考えることが不適当と思われる特殊事故（車両火災、人の転落等）及び路面凍結、積雪時の事故を除いた29,082件である。

道路構造データは、100mポスト毎の平面曲線半径、縦断勾配、構造種別(土工・橋梁・トンネル別)で、交通量は、平均約3,890万台/6年間(17,700台/日)、最大7,190万台/6年間(32,800台/日)、最小2,360万台/6年間(10,800台/日)である。図-6に100mの単位区間毎の事故件数の分布状況を示す。単位区間では、たまたま起きた1件が事故率に大きな差として現れる。

(3) 要素区間の設定

誤差幅誤差率基準の基準値として、許容誤差幅は0~50件/億台キロ、許容誤差率は0~50%の組み合せで99通り、走行台キロ基準の基準値として、0.1~2.5億台キロで11通り考え、また、比較のため等区間長(0.1~7.5km)16通りを加えて、合計126のサンプル(要素区間セットのバリエーション)を設定した。これらの要素区間設定にあたっては、道路の上流側から、各基準値を満足するようになるまで順次単位区間を取り込んでいく方法を取った。尚、等区間長サンプル以外では、事故発生地点と事故として記録される地点とのズレを少なくするために、単位区間数を4以上とした。

図-7にサンプルを評価する $1/\theta - 1/(1+\mu)$ 平面を示す。各サンプルは、 $1/(1+\mu)$ が大きいほど(偶然誤差が小さいほど) $1/\theta$ が大きい(回帰係数の分散が大きい)というトレードオフの関係を示している。この中から、分布全体の中で左上に突出した、即ち、 $1/(1+\mu)$ が大きいにもかかわらず $1/\theta$ が比較的小さいものとして、図中の

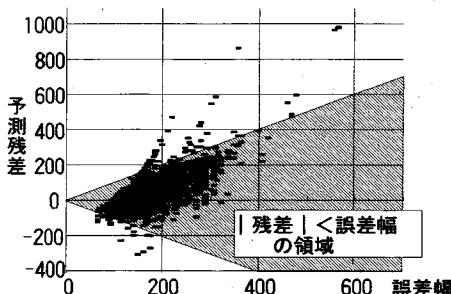


図-8 誤差幅と遠心加速度での残差（単位区間）

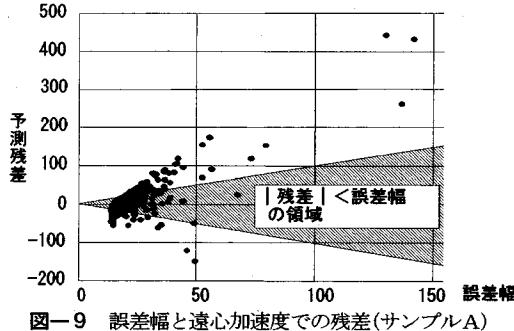


図-9 誤差幅と遠心加速度での残差（サンプルA）

誤差幅誤差率基準サンプルA ($e_{wa} = 15$ 件/億台キロ, $e_{ra} = 25\%$, 要素区間数: 404) を選定した。

(4) 要因分析

単位区間に毎に説明変数を定義し、サンプルAに対応する要素区間に変換して重回帰分析を行った。更に、残差分析によって、十分説明できていない要素区間を抽出し、内容を検討して説明変数を改良した。

図-8、図-9に有力な説明変数である遠心加速度指数での残差と誤差幅の関係を示す。単位区間のままの分析（図-8）では、多くの区間が残差より誤差幅の方が大きく、残差分析により説明変数の改良を行なうことは困難・不適切である。サンプルAの分析（図-9）では、誤差幅を越えた大きな残差をもつ要素区間が特定でき、残差分析が有効に行なえた。

(5) 分析結果

サンプルAに対して残差分析を含めた説明変数の選択・改良を行なった結果、4説明変数による事故率説明モデルが得られた。選ばれた説明変数の概要は次の通り。

a) 遠心加速度指数 [F]

一定の推進力（平坦部で60km/h走行）を維持し、縦断勾配によって自然に加減速されながら走行している車両を考え、これが曲線部で受ける横方向の加速度を示す。

$$F = \max(1/R, 1/R_{\max}) \times V^2 \quad \dots \quad (54)$$

R : 平面曲線半径 R_{\max} : R の上限値1500m

V : 自然に加減速される速度を表す指數

b) 急峻指数 [FF]

残差分析の結果、線形の特に厳しい区間については、事故率が遠心加速度指数と線形関係でなくなり、急に高

表-3 回帰分析結果

| | 単位区間分析 <単位区間数> | サンプルAでの分析 | |
|-----------------|-------------------|-------------------|-------|
| | | 重み付き | 重み無 |
| 要素区間数 | 9,959 | 404 | 404 |
| 平均区間長(km) | 0.10 | 2.46 | 2.46 |
| 事故率分散 | 71.0 ² | 39.5 ² | - |
| 偶然分散 | 45.6 ² | 9.6 ² | - |
| 危険率分散 | 54.4 ² | 38.3 ² | - |
| 偶然分散比 μ | 0.704 | 0.063 | - |
| 有効危険率比 θ | 1.000 | 0.502 | - |
| 決定係数 | 0.222 | 0.617 | 0.714 |
| 重相関係数 | 0.471 | 0.785 | 0.861 |
| 回帰係数 | | | |
| 遠心加速度指数 | 120.1 | 119.1 | 113.2 |
| 急峻指数 | 417.6 | 670.8 | 732.4 |
| トンネル指數 | 23.7 | 31.7 | 47.7 |
| 直線率 | 27.9 | 29.7 | 32.2 |
| 回帰定数 | 32.8 | 31.7 | 34.8 |

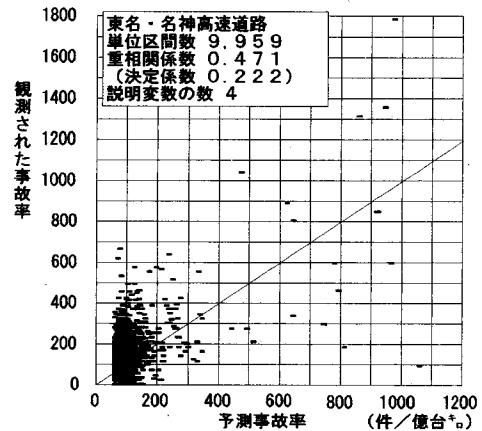
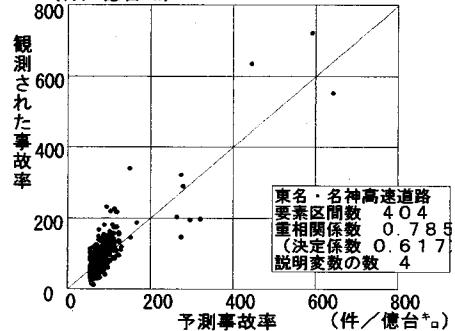
図-10 事故率の予測値と観測値（単位区間）
(件/億台*)

図-11 事故率の予測値と観測値（サンプルA）

くなることが明らかとなった。これを反映させるために導入した、急峻部での速度と曲率の組み合わせを示す。

$$\text{急カーブ・急下り坂で } FF = (V - V_{\min}) / R \quad (55)$$

(300m以下)・(4%以上), $V_{\min} = 60\text{km/h}$

c) トンネル指數 [Tn]

トンネル内及びトンネル上流1km以内で $Tn = 1$ (56)

d) 直線率 [L20]

長い範囲で直線がどの程度使われているかの指數

これらの説明変数を用いて推定された回帰係数及び決

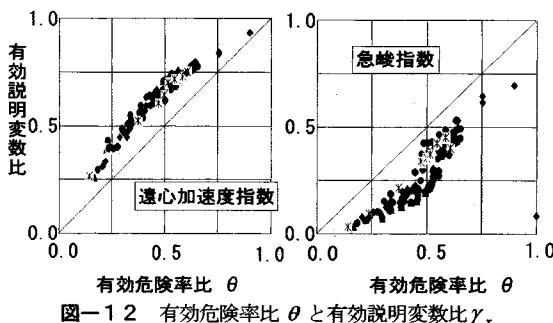
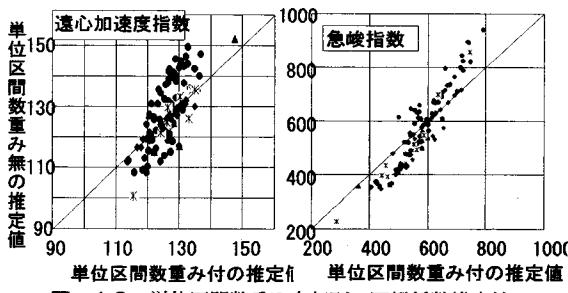
図-1-2 有効危険率比 θ と有効説明変数比 γ_x 

図-1-3 単位区間数重み無別の回帰係数推定値

定係数等を表-3に示す。参考のため、単位区間に直接回帰分析を行なった結果及び単位区間数の重みを付けない回帰分析結果を併記する。また、図-10に単位区間での予測値と実事故率との関係を、図-11にサンプルAでの予測値と実事故率との関係を示す。

偶然分散比が0.704と著しく高い単位区間での分析では、決定係数が0.222（重相関係数0.471）と低く、図-10でも予測の有効性は見られない。これに対して、同じ説明変数を用い、偶然誤差を取り除いて偶然分散比0.063としたサンプルAでは、決定係数は0.617（重相関係数0.785）と高く、図-11でも予測の有効性は明らかである。

また、重み無し回帰分析では、ほぼ同じ回帰係数で決定係数が0.741（重相関係数0.861）と著しく高いものとなっている。これは、事故率が高く区間長の短い要素区間が強調され、決定係数の分母である事故率分散が大きくなっているためである。従来の事故分析では、この重み無し分析が用いられている。そこでの決定係数と比べれば、本研究で得られた事故率説明モデルが、非常に高い説明力を持つものであることがわかる。

(6) 有効説明変数比と有効危険率比

図-1-2に、各サンプルでの有効危険率 θ と有効説明変数比 γ_x との関係を示す。本研究では、 γ_x の代わりに θ を用いることを提案したが、実際のデータで両者は良く対応しており、実用上は十分妥当と言える。

(7) 単位区間数重みの有無による比較

図-1-3は、各サンプルでの単位区間数（要素区間の延長重み）の有無による回帰係数推定値の違いを示したものである。これによれば、重み無し分析では、単位区間数の多い要素区間でも少ない要素区間でも同じ1として

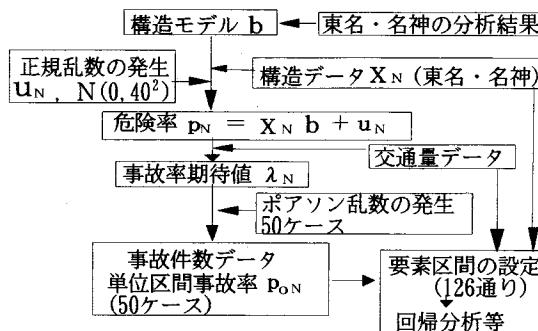
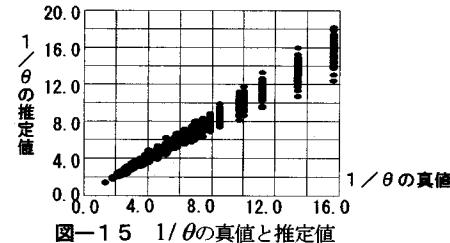
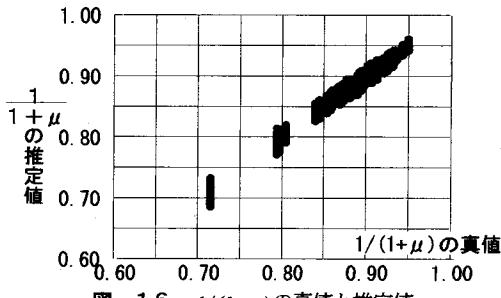


図-1-4 シミュレーションの流れ

図-1-5 $1/\theta$ の真値と推定値図-1-6 $1/(1+\mu)$ の真値と推定値

扱うため、特定の短い区間の影響を受けて結果が不安定となる。一方、重み付分析では、要素区間の取り方の違いによる回帰係数推定結果の違いは小さい。

7. シミュレーションモデルによる検証

(1) シミュレーションモデルの概要

提案した指標の推定精度や回帰分析に対する偶然誤差の影響の考え方を検証するため、図-1-4に示すシミュレーションを行なった。方程式のはずれの系列相関は考えていない。また、 $\sigma=40$ 件/億台キロとした。

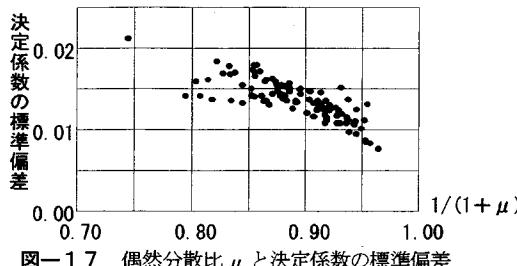
(2) 各指標の推定精度

a) 有効危険率比 θ の推定

サンプルの評価は $1/\theta$ で行うので、 $1/\theta$ の真値と推定値の関係で見ると、図-1-5に示す通りとなる。有効危険率比 θ が実用上十分な精度で推定でき、 $1/\theta$ が小さくなるほど推定誤差が小さくなることが分かる。

b) 偶然分散比 μ の推定

サンプルの評価は $1/(1+\mu)$ で行なうので、その真値と推定値を図-1-6に示す。これから、偶然分散比 μ が実用上十分な精度で推定でき、 $1/(1+\mu)$ が大きくなるほど推定誤差が小さくなることが分かる。

図-17 偶然分散比 μ と決定係数の標準偏差

c) 決定係数

事故率の決定係数のバラツキを見るため、図-17に、50ケースの事故データによる決定係数の標準偏差を示す。 $1/(1+\mu)$ が大きいほど、偶然誤差による決定係数のバラツキが小さいことが分かる。

8. むすび

本研究では、事故率を分析する際の偶然誤差に関し、回帰分析に与える影響の理論的検討を行なった上で、具体的な分析用の区間設定方法を示すことができた。

明らかとなったことをまとめると次の通りとなる。

- ・道路区間の事故率 p_a （走行キロ当たりの事故件数）の偶然誤差の大きさは、未知の母数である危険率 p （単位延長当たりの事故発生確率）と p_a との差として、信頼区間推定の信頼上限を用いた「誤差幅 e_w 」及び事故率に対する誤差幅の割合を示す「誤差率 e_r 」で評価できる。
- ・データ収集の単位となる短い区間である「単位区間」をまとめて、分析対象サンプルの要素となる区間である「要素区間」を設定する場合に、サンプルの特性は、単位区間での危険率分散に対する要素区間での危険率分散の割合を示す「有効危険率比 θ 」と、要素区間での危険率分散に対する偶然分散の比を示す「偶然分散比 μ 」の2つのパラメータで表すことができる。 θ , μ は未知の定数であるが、データから推定が可能である。
- ・事故率の回帰分析での決定係数は、危険率分析での決定係数と偶然誤差の影響との積となっており、偶然分散比 μ を小さくすれば、偶然誤差の影響を小さくできる。
- ・回帰係数推定量の分散は、方程式のはずれの影響と偶

然誤差の影響の和となっており、有効危険率比 θ を大きく保てば、回帰係数推定量の分散を小さくできる。

・残差分析を有効にするため、個々の事故率の誤差幅又は誤差率を一定の基準値内に納める「誤差幅誤差率基準」に従い、サンプル（設定された要素区間のセット）を作成する。幾つかの基準値を仮定して作成したサンプルの中から、偶然分散比 μ 及び有効危険率比 θ の各推定値を用いて分析に適したサンプルを選定する。これにより、残差分析が可能で、決定係数に対する偶然誤差の影響が小さく且つ回帰係数推定量の分散も小さく保った分析用の区間設定ができる。

・東名・名神のデータによる適用例及び乱数を用いたシミュレーションによりこの方法の有用性が検証された。

一般道路で交通量の多くない場合や、短期間のデータしか利用できない場合等偶然誤差の大きいデータの取扱いや、交通量に誤差がある場合、あるいは、線形重回帰モデル以外の解析手法への適用については、今後の検討課題としたい。

謝辞：本研究を進めるに際し、東京大学の越正毅教授には、終始適切な助言と励ましをいただきいた。記して感謝の意を表します。（東京大学審査学位論文要約）

参考文献

- 1) 松島憲昭：高速道路の事故多発区間分析、交通工学24巻1号、pp. 23～29、1989.
- 2) Siem Oppe: A comparison of some statistical techniques for road accident analysis, Accid.Anal.Prev., Vol. 24, No. 4, pp. 397～424, 1992.
- 3) 交通工学研究会：事故要因分析、1978.
- 4) 岡本博・越正毅他：事故発生の偶然変動を考慮した道路区間の事故危険度の評価手法、土木学会論文報告集、No. 326, pp. 115～127, 1982.
- 5) 斎藤和夫：事故危険度評価方法に関する調査研究の概観(1)(2)、交通工学、15巻、6号、pp. 47～54, 7号、pp. 37～48、1980。
- 6) 岡本博：偶然誤差を考慮した事故率の回帰分析手法に関する研究、1993。

(1994. 3. 16 受付)

A SEGMENTATION METHOD FOR REGRESSION ANALYSIS OF ACCIDENT RATES WITH RANDOM ERRORS

Hiroshi OKAMOTO

A traffic accident is a stochastic event and observed accident rates contains random errors that deteriorate the explanatory power and the reliability of the regression analysis. Random errors have to be appropriately controlled. When a data set of accident records, traffic records and geometric design elements of a certain stretch of a road is given, there are two steps for regression analysis: first, division of the road into a number of segments; and second, application of regression analysis to the set of segments. The main interest of this paper is the first step. The paper discusses how to evaluate random errors contained in accident rates of segments and what influence the random errors have on the regression. A segmentation method is proposed and examined by numerical examples of Tokyo-Kobe Expressway.