

判別分析を用いたPSCターゲットシステムに関する研究

伊藤 博子*・松岡 猛†・岡田 光豊‡

PSC Targetting System with Discriminant Analysis

Hiroko ITOH, Takeshi MATSUOKA and Mitsutoyo OKADA

Abstract

This research applies discriminant analysis for ships-in-port classification as a new port state control (PSC) targetting system in order to promote the efficiency of PSC. Discriminant analysis generates classification functions for a sample of observations from pre-determined categories. We applied Fisher's linear discriminant function to historical inspection data to create a classification function that select ships with deficiencies worth detention out of object ships. Based on the developed function, a method to estimate the percentage of required inspection to cover a given rate of detention is determined. The result of comparison between proposed targetting system and 2003 revised targetting system shows that the former provides better capture rates in the range of 10–60% inspection rates.

Keywords: port state control(PSC), discriminant analysis, inspection, detention

キーワード: ポートステートコントロール, 判別分析, 検査, 拘留

1. 緒論

国際基準に満たないサブスタンダード船と呼ばれる船舶が安全上、環境保全上の問題を引き起こす事例が多く、寄港国側で入港船舶を監督するポートステートコントロール(PSC)の重要性が増している。PSCは国際海事機関(IMO)での採択(1981年)を契機に世界中で実施されるようになり、徐々に検査体制が強化されてきた。しかし、PSCの船舶検査は検査側船舶側の双方にとって多大な労力と時間がかかり、多くの港において全数検査は不可能であり、多数船舶が入港する箇所では検査率の向上も難しい。そこで、限られた資源を効率的に充当しようと、ターゲットシステムと呼ばれる基準が各種提案され、その算定結果にもとづいて検査対象を選定することが考えられている。しかし、これまでの算定方法は多分に経験者の勘によるものであり、かつ得られた評価値のどこまでを検査対象とすべきかといった明確な基準も設けられていない。

本論文では、PSC業務の効率化のために、統計的手法に基づいた判断基準を提供する新たなターゲットシステムを提案する。具体的には、拘留による改善を必要とする船舶を検査対象として捕捉する割合を向上させることを目的とし、過去の検査結果のデータをもとに推定する方法を提案する。まず、分類されるグループが既知の標本から分類の目安となる関数を求める判別分析(discriminant analysis)を用いて、検査が必要かどうかを判別する判別関数を導出する。次に、得られた判別関数を用いて、変動する検査率に対応させて検査対象を選出する方法を考え、評価値と捕捉率の関係を整理する。また、最新のターゲットシステムである「2003年修正案」による捕捉率と比較し、提案するターゲットシステムの有効性を確認する。

* 正会員 海上技術安全研究所 (〒181-0004 東京都三鷹市新川6-38-1) hiroko@nmri.go.jp
† 海上技術安全研究所
‡ 東京MOU事務局 (〒105-0001 東京都港区虎ノ門3-8-26)

2. PSC 検査履歴とターゲットシステム

2.1 検査履歴データ

PSC実施国は、地域ごとに「MOU (Memorandum of Understanding)」とよばれる覚書にもとづいたコミュニティを形成し、検査マニュアルの標準化や情報交換などで協力し、より危険と考えられる船舶を重点的に監督することを目指している。日本やアジア太平洋地域の各国が参加する東京 MOU では、2002 年に検査対象となる寄港が延べ 25,202 隻あり、78%にあたる 19,588 隻において検査が行われた。これにより、13,760 隻で欠陥が発見され、是正するまで出航を認められない重大な欠陥を抱えた船舶（以下「拘留された船舶」と呼ぶ）は 6.7%にあたる 1,307 隻あった。

ウラジオストックにある APMIAS (Asia Pacific Maritime Information and Advisory Services) では、東京 MOU 参加国における検査データの共有を目的に、AP-CIS (Asia Pacific Computerized Information System) を構築しデータを収録している。毎回の検査データとして、以下の 16 項目が収録されている。

まず、検査対象となる船舶の属性等として 9 項目が記録される。

- 検査当局 (name of reporting authority)
- IMO 番号 (IMO number)
- 船籍 (ship flag state)
- 船種 (ship type)
- 船級協会 (ship classification society)
- 建造年 (year of build)
- 検査日 (date of inspection)
- 検査場所 (place of inspection)
- 会社 (company)

ここで、「検査当局」、「船籍」、「船級協会」等はあらかじめ定められた ID 等の使用により統一的に表記されている。「会社」(ISM コードの company) は自由記述式となっており、約 9 割に何らかの記述がされている。それらのうち、ほとんどが運航管理会社の名前であるが、船主の場合や併記の場合もあり、一部記号や略号等の判読不能なものも混在している。

次に、前回検査までの履歴データとして 4 項目を設定し、その都度計算し記録した。なお、括弧内のセミコロン以降は次節以降の数式にて使用する際の変数名である。

- 前回検査からの経過月数 (time since last inspection (months); x_{time})
- 過去 4 回の検査における拘留数 (number of detentions during last 4 inspections; x_{det})

- 過去 4 回の検査における欠陥数 (number of deficiencies during last 4 inspections; x_{def})
- 前回検査における重大欠陥数 (number of outstanding deficiencies during last inspection; x_{out})

検査により得られる検査結果データとして 3 項目が記録される。

- 全欠陥数 (total number of deficiencies)
- 拘留に相当する欠陥数 (number of detainable deficiencies)
- 拘留されたか否か (detention (yes/no))

本研究では、検査前に得られる 13 項目のデータをもとに、「拘留されたか否か」を予測することとなる。

2.2 推定できる変数

検査履歴のデータが蓄積された場合、以下の統計量についておおよその値が推定できる。ただし、PSC が全数検査や無作為検査でないため、母集団に偏りがある点に注意する必要がある。

- 船籍毎の拘留率 (ship flag detention rate(%); x_{flag})
- 船級協会毎の拘留率 (classification society detention rate(%); x_{CS})

また、IACS に所属する船級協会のリストおよび建造年月日からの算出により以下の 2 項目が分かる。

- IACS であるか否か (IACS or Non-IACS; $x_{IACS} \in \{0, 1\}$, 0=false, 1=true)
- 船齢 (ship age; x_{age})

2.3 ターゲットシステム

ターゲットシステムとは、PSC を効率的に行うために検査対象となる寄航をする船舶群の中から、船種や船齢、過去の検査履歴等を考慮して検査すべき対象を選出する基準のことであり、評価値の算出法及び得られた評価値の扱い方に関する取り決めが中心的な役割を担っている。主要なターゲットシステムとしては、パリ MOU による方式や東京 MOU による「2003 年修正案」などが存在する。2003 年修正案では、船籍毎の拘留率 (x_{flag})、船種 (type)、IACS であるか否か (x_{IACS})、船齢 (x_{age})、前回検査からの経過月数 (x_{time})、過去 4 回の検査における拘留数 (x_{det})、過去 4 回の検査における欠陥数 (x_{def})、前回検査における重大欠陥数 (x_{out}) を用いて、評価値 z を次式により算出し、 z 値の大きい船舶ほど検査すべき度合いが高いと定めている。

$$z = z_{flag} + z_{type} + z_{IACS} + z_{age} + z_{time} + z_{det} + z_{def} + z_{out} \quad (1)$$

ここで、

$$z_{flag} = \max \{ x_{flag} - \overline{x_{flag}}, 0 \}$$

$$z_{type} = \begin{cases} 4 & \text{if } type \in \{13, 30, 40, 55, 60, 61, 70\} \vee x_{age} > 15 \\ 0 & \text{otherwise} \end{cases}$$

$$z_{IACS} = \begin{cases} 0 & \text{if } x_{IACS} = true \\ 10 & \text{otherwise} \end{cases}$$

$$z_{age} = \begin{cases} 0 & \text{if } 0 < x_{age} \leq 5 \\ 5 & \text{if } 5 < x_{age} \leq 10 \\ 10 & \text{if } 10 < x_{age} \leq 15 \\ 10 + 1/(x_{age} - 15) & \text{if } 15 < x_{age} \leq 20 \\ 15 + 2/(x_{age} - 20) & \text{otherwise} \end{cases}$$

$$z_{time} = \begin{cases} 3 & \text{if } 6 < x \leq 12 \\ 6 & \text{if } 12 < x \leq 24 \\ 50 & \text{if } 24 < x \\ 0 & \text{otherwise} \end{cases}$$

$$z_{det} = \begin{cases} 15 & \text{if } x_{det} = 1 \\ 30 & \text{if } x_{det} = 2 \\ 60 & \text{if } x_{det} = 3 \\ 100 & \text{if } x_{det} = 4 \\ 0 & \text{otherwise} \end{cases}$$

$$z_{def} = 0.5 \cdot x_{def}$$

$$z_{out} = 2.0 \cdot x_{out}$$

なお、 z_{type} における特定の船種に関して船齢15を超える場合に加点する部分などについては、経験者の勘というよりむしろ政策上の要請にもとづく算定が含まれているといわれている。

3. 判別分析

3.1 判別関数について

判別分析 (discriminant analysis) とは、分類されるグループとして既知の標本があるときに、与えられた新しい測定値がどのグループに属するかを判別する手法である。判別分析には、主として一次関数で表現される線形判別関数を用いる方法 (Fig. 1a) と、平均からの距離を各群の分散によって標準化したマハラノビスの距離による二次の判別関数を用いる方法 (Fig. 1b) がある。

一般には、各群の分散共分散行列が等しい場合、すなわちそれぞれの分散共分散行列を Σ_1, Σ_2 とおいたとき、 $\Sigma_1 = \Sigma_2$ が仮定できれば線形判別関数を使用し、そうでない場合は二次の判別関数を使用する⁽²⁾。実際に使用する

るデータを、拘留された船舶群 (以下「拘留船舶群」という) と拘留されなかった船舶群 (以下「非拘留船舶群」という) の2グループに分類し、分散共分散行列の同等性を自由度21の χ 分布において有意水準 $\alpha = 0.05$ で検定した結果、 $\chi^2 < \chi_{\alpha}^2$ となることから、有意確率 P は、 $P < \alpha$ となり、 Σ_1 と Σ_2 の間に統計的に有意な差があることが分かった。

以上から、上述の考えに従えば二次の判別関数を使用することになるが、二次の判別関数の場合、数式の項数が多くなる上、変動する検査率に対応させて判別の境界線を移動させることが難しい。そこで、ターゲットシステムとしての実用性を重視して、近似的に線形判別関数を用いた上で、検証用データにより判別結果を確認していくこととした。

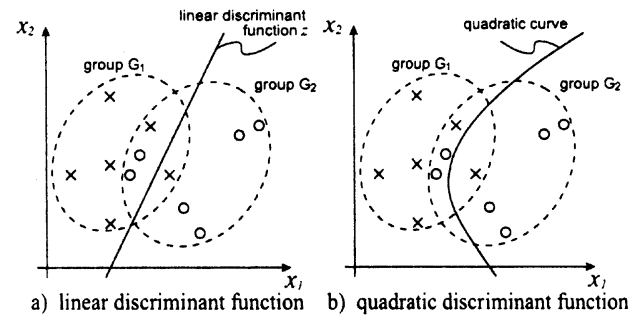


Fig. 1 Discriminant analysis.

3.2 線形判別関数による判別

p 個の変数 x_1, \dots, x_p に対して任意の係数 (a_1, \dots, a_p) を用いて作られる線形結合を z とする。

$$z = a_1 x_1 + \dots + a_p x_p \quad (2)$$

この値 z によって判別することを考える。係数 a_1, \dots, a_p は、観測されている2群のデータをできるだけ正確に判別するように定めたい。係数 a_1, \dots, a_p が定まると、 n 個のサンプルのそれぞれに対して、線形結合 z の値 $z_i^{(k)} = a_1 x_{i1}^{(k)} + \dots + a_p x_{ip}^{(k)}$ ($k = 1, 2; i = 1, \dots, n_k$) を計算することができる。これを判別得点と呼ぶ。全体の平均値を \bar{z} 、第 k 群における平均値を $\bar{z}^{(k)}$ とすれば、 z の平均値との差の総平方和 S_t は、群間平方和 S_b と群内平方和 S_w に分解される。

$$\sum_{k=1}^2 \sum_{i=1}^{n_k} (z_i^{(k)} - \bar{z})^2 = \sum_{k=1}^2 \sum_{i=1}^{n_k} (z_i^{(k)} - \bar{z}^{(k)})^2 + \sum_{k=1}^2 n_k (\bar{z}^{(k)} - \bar{z})^2 \quad (3)$$

ここで左辺が総平方和 S_t 、右辺第一項が S_w 、第二項が S_b である。群間の差が大きければ、 S_b が大きくなる

ことから、「良い判別」を全体に対して群間平方和 S_b が大きくなることと考へ、相関比 $\eta^2 = S_b/S_t$ を最大化するよう係数 a_1, \dots, a_p を定める。 η^2 を a_1, \dots, a_p について偏微分してゼロとおくことで、 η^2 を最大にする係数が導かれる。このようにして得られた係数を用いて判別得点を計算し、判別の分点としては、両群の平均の中点を用いたり、標本の判別に対する誤りの個数が少なくなるような点を用いたりする。

3.3 説明変数の選定方法

Table 1 はピアソンの相関係数による主な変数間の相関行列である。一般に「かなり強い相関がある」とされる ± 0.7 から ± 1 の相関が、 x_{CS} と x_{flag} 、 x_{IACS} と x_{CS} の間にあり、「やや相関あり」とされる ± 0.4 から ± 0.7 の相関が 9 組あることから、履歴データの変数間には高い相関があることが読み取れる。このように、独立変数相互間の相関が高い場合、相互に影響しあって判別分析を失敗させることがあるため、相関の高い変数を除外するか、変数選択を適用する必要がある。ここではステップワイズ法で変数選択を行うとともに、判別分析の失敗は求める判別係数の符号が実態と逆になる現象として表れることが多い点に注意して判別結果を確認することとする。

Table 1 Correlation matrix of main variables.

	x_{flag}	x_{CS}	x_{det}	x_{def}	x_{out}	x_{age}	x_{time}	x_{IACS}
x_{flag}	1	0.782	0.424	0.402	0.238	0.372	-0.113	-0.687
x_{CS}	0.782	1	0.450	0.403	0.227	0.374	-0.119	-0.902
x_{det}	0.424	0.450	1	0.618	0.318	0.289	-0.084	-0.409
x_{def}	0.402	0.403	0.618	1	0.599	0.371	-0.116	-0.402
x_{out}	0.238	0.227	0.318	0.599	1	0.220	-0.047	-0.220
x_{age}	0.372	0.374	0.289	0.371	0.220	1	-0.002	-0.351
x_{time}	-0.113	-0.119	-0.084	-0.116	-0.047	-0.002	1	0.115
x_{IACS}	-0.687	-0.902	-0.409	-0.402	-0.220	-0.351	0.115	1

4. 判別分析の結果

4.1 判別関数

APCIS のデータのうち、2002 年 1 月から 2003 年 12 月までの 37,106 件を用いて判別分析および結果の検証を行う。以下のようにデータを分割して使用した。

- 判別用…2002 年 1 月～2003 年 8 月検査分 (32,121 件)
- 検証用…2003 年 9 月～2003 年 12 月検査分 (4,985 件)

判別用データに対して、市販の統計解析用ソフトウェア SPSS を用いてステップワイズ法を使用した線形判別分析を行ったところ、6 種類の変数を用いた判別式が作成された。得られた判別関数の係数と定数項を、Table 2

に示す。

Table 2 Coefficients for the discriminant function.

x_i	standardized coefficient	coefficient
x_{flag}	0.409	0.053
x_{CS}	0.274	0.040
x_{det}	0.210	0.388
x_{def}	0.123	0.009
x_{age}	0.190	0.023
x_{IACS}	-0.115	-0.341
(const.)		-0.909

これにより、判別関数は、以下のように定義される。

$$z = 0.053 \cdot x_{flag} + 0.040 \cdot x_{CS} + 0.388 \cdot x_{det} + 0.009 \cdot x_{def} + 0.023 \cdot x_{age} - 0.341 \cdot x_{IACS} - 0.909 \quad (4)$$

なお、ステップワイズのための F 値は $F_{in} = 3.84$ 、 $F_{out} = 2.71$ とした。これは、説明変数の目的変数への影響の大きさを示す寄与度が 0.0 (寄与が無い) と置いても良いという仮説を統計的に検定し、自由度 (標本数に対応) ∞ 、有意水準 0.05 と 0.10 で仮説が棄却される場合、すなわち寄与がある場合に相当する。また、ステップワイズ法により選択された変数の中に、共線性の影響が懸念された x_{IACS} と x_{CS} が含まれるが、両者に対して得られた係数の符号には不整合が見られなかった。

4.2 判別関数による判別用データの分類結果

得られた判別関数における分類結果のクロス集計表を表 3 に示す。なお、観測データにおける拘留船舶群と非拘留船舶群の分散共分散行列の同等性は、Box の M 検定においても 0.05 で棄却される。しかし、ここでは前述の理由により線形判別関数を作成してその判別結果を検証する。

Fig. 2 に得られた判別関数を判別用データに適用した場合の判別得点の度数分布グラフを示す。拘留なしのグループでは、 $z = -0.6$ の度数がもっとも多く、平均値が -0.12 であるのに対し、拘留ありのグループでは、 $z = -0.2$ 付近と $z = 2.6$ 付近の 2 箇所度数が多くなる他、平均値も 1.36 と高くなっている。

Table 3 Classification matrix of training data. ($z = 2.3$)

		detention (yes/no)	estimated group		Total
			no	yes	
original data	frequency	no	28,305	1,308	29,613
	yes	yes	1,702	806	2,508
%		no	95.6	4.4	100.0
	yes	yes	67.9	32.1	100.0

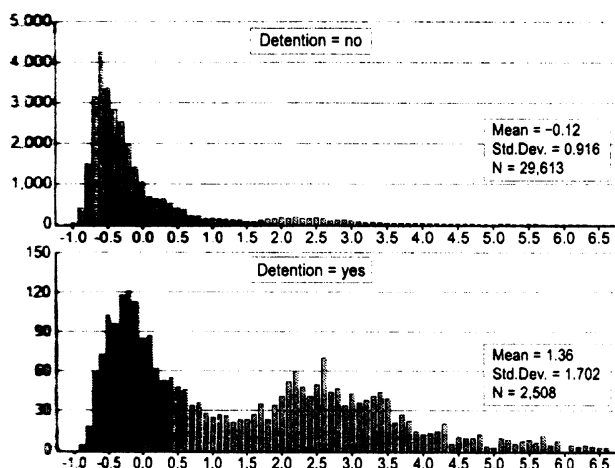


Fig. 2 Distribution of z values.

5. 検査率と捕捉率の推定

5.1 目標捕捉率を実現するための検査率の推定

線形判別関数を用いた判別の分点は、各群の平均の中心を通るように定めたり、全体の誤判率を最小化するように定めることが多い。しかし、ここで目的としているのは、拘留による改善の必要な船舶をできるだけ検査できるように、限られた検査資源を割り当てることである。そこで、Fig. 3のように判別の分点を移動させながら、目標とする捕捉率を探索することを考える。z値が与えられれば対応する検査率と推定される捕捉率が計算できることから、推定される捕捉率が目標とする値となるようにz値を変動させればよい。z値に対して捕捉率は単調減少することから、変動させたz値に対して与えられたデータを一通り判別して捕捉率を計算し、2分法によりz値を変動させながら再帰的に目標とする捕捉率を探索して、変動の幅 $\epsilon = 10^{-5}$ をもって収束とみなすことで対応するz値を得ることとした。この方法は、目標とする検査率を満足するz値を得るためにも適用することができる。探索の手順をFig. 4にまとめた。

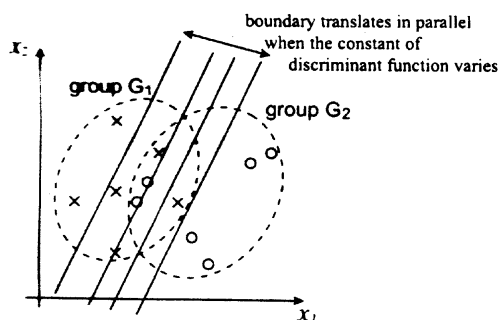


Fig. 3 Shifting the boundary of z value.

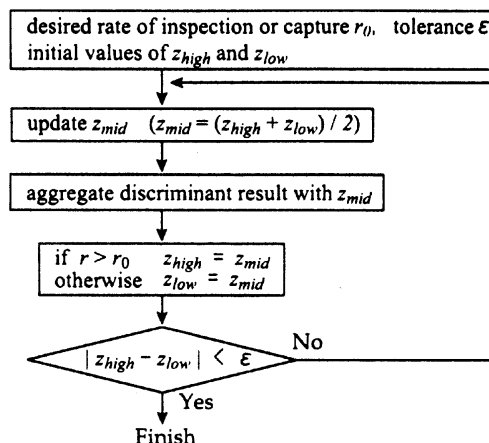


Fig. 4 A method to estimate z value corresponds to a given capture rate.

判別用データ 32,121 件に関して 2 分法を適用しながら、十分位数に対応する評価値を推定したところ、Table 4 のようになり、対応して検査率と捕捉率関係が Fig. 5 のように得られた。

Table 4 Relations of capture rate, z value, and inspection rate (discrimination data).

capture rate (%)	z	inspection rate (%)
0	6.50	0
10.0	3.59	1.4
20.0	2.92	3.2
30.0	2.40	5.9
40.0	1.90	9.2
50.0	0.94	13.5
60.0	0.31	20.1
70.0	-0.05	28.6
80.0	-0.28	40.2
90.0	-0.51	60.1
100.0	-1.00	99.9

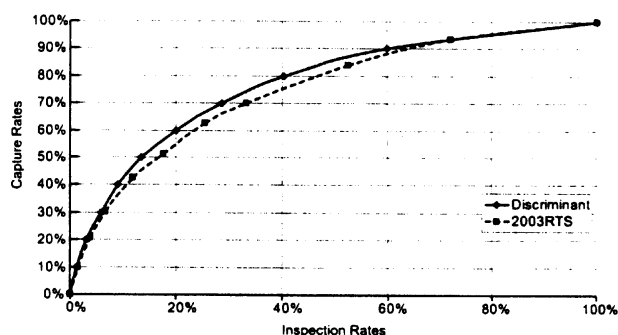


Fig. 5 Result of discriminant analysis compared with 2003 RTS (discrimination data).

6. 検証

検証用データを用いて判別式で評価し、検査率と捕捉率の関係をもとめた。その結果を Fig. 6 に示す。捕捉

率の20%、40%、60%、80%の各点における判別用データによる結果との誤差が0.2~1.6%の間であることから、判別時とほぼ同様の結果を得られたと考えられる。比較のため2003年修正案によって同様の判別を行い、合わせてグラフ化した。

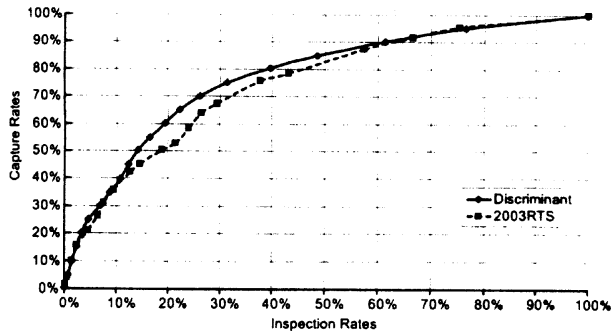


Fig. 6 Result of discriminant analysis compared with 2003 RTS(verification data).

Fig. 6より分かるように、検査率75%付近では、2003年修正案の方が1%程度良い捕捉を行ったものの、検査率10%から60%にかけて、提案する手法が2003年修正案による判別よりも数%から10%高い捕捉率となり、検査率10%以下の範囲でも提案する手法は2003年修正案と同等以上の捕捉を行っている。以上から、次のように考察する。

- 2003年修正案では、検査率を十分確保できる場合の領域では良い結果を出している反面、低い検査率の領域に問題があった。これに対し、提案する手法では、近似的に線形判別関数を用いたものでありながら低い検査率に対しても良好な結果が得られ、現実的に全数検査の難しい港等での適用に期待できると考えられる。
- 提案する手法では、捕捉したい割合を与えると、判断の指標となる判別得点 (z 値) が得られることから、検査における意思決定に有用と考えられる。
- 判別、検証に用いたデータは、実際に行われたPSC検査の結果であるため、検査対象となる寄航の約8割でしかない点に注意する必要がある。ここに入っていないデータには、既存のターゲット方式等により優良と推測された船舶や、6ヶ月以内に検査した船舶は除外するという東京MOUの取り決めに従って除外

された船舶が含まれるが、それらがどのようなものであるかについて、今後さらに調査すべきと考える。

7. 結論と今後の課題

本論文では、PSC業務の効率化のために、統計的手法に基づいた判断基準を提供する新たなターゲットシステムを提案した。具体的には、まず、判別分析により与えられた船舶データから拘留による改善を必要とするかどうかを推定する判別関数を導出した。さらに、この判別関数に対して2分法を適用しながら目的とする捕捉率に対応する検査率を導出する方法を提案した。これらにより、実施可能な検査率に応じて検査対象を選出したり、所望する捕捉率を達成するために必要な検査率を推定することができるようになった。また、提案手法の推定結果は、最新のターゲットシステムである2003年修正案と比較して良好な結果が得られることが分かった。

パリMOUなどでは、PSCの影響により、寄港する船舶の質が改善し、それに伴い拘留率が下がるという現象も見られる。提案した手法を用いれば、判別用データを最新のものとして随時判別式を更新していくことで、このような経時変化へ対応していくことも考えられるので、今後さらに検討したい。また、今回使用したデータの中には、会社の情報が含まれているものの、現時点では判別分析用に十分に良い質のデータではなかった。今後収集方法が整備されれば、会社ごとの履歴を変数として組み入れて分析することでより有効な判別式が得られると期待できる。

謝辞

本研究は、日本財団の2003年度助成事業として実施された「PSCターゲット方式に関する調査研究」の一部として行われました。

参考文献

- (1) Tokyo MOU Secretariat: Annual Report on Port State Control in the Asia-Pacific Region 2002.
- (2) 有馬哲, 石村貞夫: 多変量解析のはなし. 東京図書 1987.
- (3) 岡田光豊: 東京MOUの活動. 海運 平成14年8月号, pp.68-. (社)日本海運集会所 2002.