

# Mahalanobis · Taguchi System 法による健康診断の 予測的評価と効率化

ナカジマ ヒサト カタダ ケイ ヤノ ヒロシ タカギ イチロウ\*  
中島 尚登\* 高田 圭<sup>2\*</sup> 矢野 宏<sup>3\*</sup> 高木 一郎\*  
シバモト ユカ ヤマウチ マサヨシ トダコウタロウ\*  
柴本 由香\* 山内 眞義\* 戸田剛太郎\*

**目的** 多数項目の所見を統合化して判断するパターン認識に対して共通の評価を与える品質工学の評価技術である, Mahalanobis · Taguchi System (MTS) 法を用い, 健康診断の判定への応用, 項目選択によるコスト削減を含めた効率化, および将来の健康状態の予測の可能性を検討した。

**方法** ①某企業の2年間にわたる健康診断結果を用い, 2年共「異常無し」のA判定者で基準空間を作成し, 「要治療・治療中」のC判定者のMahalanobisの距離を計算して, A判定とC判定の識別が可能か検討した。

②MTS法による項目選択を行い, 健康診断の判定に意味のある項目を選び, 効率化を図った。

③2年間のデータを用い, 2年目A判定者の1年目のデータで基準空間を作成し, 1年目どのようなMahalanobisの距離なら2年目A判定になるか検討した。

**結果** ①2年間にわたる健康診断結果を用い, 2年共A判定159人, C判定37人で距離を検討したところ, しきい値2.0ではAとの正診率98.1%, Cとの正診率91.9%と高率であった。

②項目選択による特に有効な項目は, 血圧, GOT, GPT,  $\gamma$ GTP, T. Chol, T. G, UA, Cr, BS, HbA1c, Hb, MCHであった。さらに項目選択で検査項目を約6割に減らすことが可能であり, これによって効率化がはかれる。

③距離のしきい値を1.5とすると, 次年度A判定と予測される割合は90%, C判定でありながらA判定と予測される割合は15%であり, 90%の割合で次年度の健康状態の予測が可能であった。

**結論** マハラノビスの距離を用いて健康診断の判定が可能であり, また, 次年度の健康状態の予測も行い, 完全ではないが, 比較的良い結果が得られた。したがって, 今年度の結果から次年度健康と予測される場合は, 健康診断が必要でなくなり, 効率化がはかれる。さらに, 項目選択によって有効と判定される検査項目は, 日常の診療上重要と思われる項目を網羅しており, 健康診断において, 医学的に必要と思われる項目の判別とコストを含めた効率化がはかれる。本研究の結果, MTS法は健康診断のシステムの改善に有用であり, さらに発展させると, 疾患の診断と治療への応用が可能であると思われる。

**Key words** : Mahalanobis · Taguchi System, Mahalanobisの距離, 健康診断, 品質工学, 予測

## I 緒 言

健康診断においては, 一般に, 検査対象者の問

診, 既往歴, 家族歴, 数十項目にもおよぶ血液検査データ, 尿検査, 画像診断, 心電図や耳鼻科, 眼科領域など, 多くの検査項目を統合して, 検査データの関係をもとに, 医師の判断により, 健康の度合いがランク分けされる。このような判断を別の角度から見れば, 多数項目の所見を統合化して判断するパターン認識と同じである。このような多数項目のパターン認識に対して, 共通の評価尺

\* 東京慈恵会医科大学内科学第1

<sup>2\*</sup> (株)セイコー・エプソン

<sup>3\*</sup> (財)日本規格協会

連絡先: 〒105-8461 東京都港区西新橋 3-25-8

東京慈恵会医科大学内科学第1 中島尚登

度を与えようというのが, Mahalanobis・Taguchi System (MTS法)<sup>1)</sup>である。本来, 判別関数の Mahalanobis の距離は, 多数項目のデータの関係の強さを多次元空間における評価尺度として数値化したものであるが, さらに一步進めて, Signal-to-Noise (SN) 比 (デシベル値) という評価尺度としてあらし, 診断の信頼性の判断に利用しようとしたのが MTS法である。

MTS法の医学への応用は, 病気の診断の確かさを評価するところにある。この方法の特徴は, 健康人のデータより基準空間を作成して, 患者のデータからの Mahalanobis の距離を求めるところにある。すなわち, 健康人と診断された人のデータから Mahalanobis の距離を用いた基準空間を作成し, この基準空間に対する患者の Mahalanobis の距離を求めることに特徴がある。肝臓病の診断に MTS法を用いる研究は, 田口<sup>1)</sup>, 兼高<sup>2)</sup>により, 診断における数値データを Mahalanobis の距離で表す事によって, 医師の総合診断と極めてよく一致するような判定を下すのが可能である事が明らかにされている。さらに MTS法を発展させると, 従来の統計学的手法では困難な, 将来の結果の予測を行える可能性<sup>3,4)</sup>がある。つまり, 2年分の健康診断データを用い, 2年目健康人の1年目のデータによる基準空間を求める事により, 今年度 Mahalanobis の距離がどの程度なら次年度が健康かどうか予測することに応用できる。さらに MTS法では, Mahalanobis の距離から基準空間作成の際に, 予測・判定に必要な項目と必要ではない項目の評価を行うことにより, 医学的に健康診断に必須な項目を選択することが可能である。その結果, 健康診断の効率化, 検査コストの削減を行うことが可能である。

今回我々は, MTS法を用いて健康診断の結果より基準空間を作成し, MTS法による診断精度を検討した。さらに項目選択による健康診断の効率化および, 本年度の健康診断から次年度の健康診断を予測する事の可能性を検討し, かなりの確かさをもって予測可能であること, および項目選択と健康診断の予測を組み合わせることにより, 従来の健康診断の方法の大幅な改善が可能であることを明らかにしたので報告する。

## II 研究方法

MTS法は, Mahalanobis の距離という既存の統計量を利用することによって, 正常状態と異常状態を合理的に計測しようとする方法である。特に, 正常状態と異常状態を単に区別するのではなく, 正常状態という基準空間を作成することにより, 両者の距離を量的に計測する点においても, 新しい考え方である。

MTS法の手順としては次のとおりである。

1. 計測項目を設定する。
2. 正常と考えられるデータを収集し, 基準空間を定義する。
3. 未知状態のデータを収集し, Mahalanobis の距離を求める。
4. 正常状態と異常状態の識別力を SN 比によって計測する。
5. 正常と異常の識別力があると判断されれば, 計測コストを考慮するため識別力を考慮し, 計測項目を減らす。

以上の手順を通して, 正常状態と異常状態を合理的に計測しようとするのが MTS法である。

対象として, 2年間にわたる健康診断結果を用いた。健康診断項目は表1に示すとおりで, 100項目あり, 健康診断結果の判定は表2に示す区分で行った。受診者5,902人中, A判定の「異常無し」および, C判定の慢性疾患に「要治療・治療中」の判定結果で, 2年間連続して表1の項目すべて受診したA判定者とC判定者を対象とし, 健康診断結果を解析に使用する事の同意を得た。

はじめに, 現在健康かどうかの判定が, MTS法で可能か明らかにする目的で, 2年間連続してA判定159人を健康人として, 各々の1年間のデータを基準空間に使用して, A判定とC判定37人との分離の可能性を検討した。

次に, 健康であるかの予測のために, 2年目A判定である336人の, 1年目のデータで基準空間を作成し, A判定とC判定137人との分離の程度から, 予測の可能性を検討した。

### 1. Mahalanobis の距離による健康診断データの数値化

性別のような文字列は「男性=1」, 「女性=2」などとしてカテゴリーデータとして数値に変換する(表3)。年齢や生化学検査値などはそのまま

表1 健康診断項目

項目	健康診断項目
加療病名	
既往歴	
家族歴	脳卒中・心臓病・高血圧・糖尿病・痛風・がんなど
嗜好	タバコ [本数, 期間], アルコール [量, 期間]
自覚症状	1. 動悸・息切れ, 2. 不整脈, 3. 狭心症状, 4. 欠番, 5. めまい, 6. 頭痛・耳鳴り, 7. 舌のもつれ, 8. 胃症状, 9. 便潜血, 10. 易疲労感, 11. 食欲不振, 12. 排尿困難, 13. 顔・手足のむくみ, 14. せき・タン, 15. 目疲れ, 16. 声のかすれ, 17. 難聴, 18. 嚥下困難, 19. 肩・首のこり, 20. 空腹時痛, 21. 食後に痛む
体側	身長・体重・体脂肪率
視力	右・左 (裸眼・矯正)
血圧	1回目・2回目 (最高・最低)
検尿	蛋白・潜血・糖
他覚症状	
聴力	会話・オーディオ (1K・4K)
胸部X線	
心電図	
血液検査	GOT・GPT・ALP・γGTP・LDH・ T. Bil・ZTT・HBs抗原・TG・T. Chol ・HDL・Cr・UA・BS・HbA1c・WBC ・RBC・Hb・Ht・MCV・MCH・ MCHC
眼底写真	
性別	
年齢	

使用する。「身長」、「体重」、「体脂肪率」であるが、「体脂肪率」が「身長」、「体重」から算出されているため、どれか1つを外して計算しないと、多重共線性となり逆行列を求めることができなくなる。よって今回は、「体脂肪率」を計算から外した。また、項目内の全データが共通して同じ値の場合はその項目を使用することはできな

表2 健康診断の総合判定の分類と判定別分布

2年目 1年目	A判定 異常なし	B判定 要経過 観察	C判定 要治療 治療中	G判定 要精密 検査	不明	合計
	A判定 異常なし	159	129	2	20	10
B判定 要経過 観察	67	242	18	37	0	364
C判定 要治療 治療中	12	262	37	54	85	450
G判定 要精密 検査	4	38	0	40	0	82
不明	94	262	80	67		
合計	336	933	137	218		

い。例えば、聴力の判定は「異常無し=1」、「異常あり=2」であり、健康人を対象にしているので、基準空間作成に使用した全データにおいて全員「異常無し=1」のことが多々存在する。よってこのような項目は除外する。

## 2. Mahalanobis の基準空間の作成と健康診断結果の識別

兼高<sup>2)</sup>の研究では、健康診断の対象者を別途診断により厳密に健康者と病人に分類し、適用の妥当性と MTS 法の信頼性の研究をしているが、今回は、複数の医師が健康診断結果を検討して、対象者が ABCG 判定 (表2) の4つのどれに分類されるか判断している。A判定とB判定の判断は難しい場合があり、同じ対象者でも、医師によりA判定に分類されたり、B判定に分類されたりする可能性がある。よって、本研究では2年間続けてA判定であった被験者を健康人と定義した。データとしては、2年間Aの人の、各年の1年分の健康診断のデータを用いた。ただし、健康人と判断しても、欠測値がある健康診断データは基準空間作成に使用すると判定精度が落ちるので、欠測値があるデータは基準空間作成に使用しなかった。また、基準空間データの中で、全項目が同じ値である項目 (たとえば健康人なので全員聴力は「異常無し」の1の値) や、多重共線性の可能性のある項目も省いた。データ数は159個、項目数は67個で基準空間を作成した。

表3より得られたデータを  $y_{ij}$  とし、項目数を  $i=1, 2, \dots, k$ , データ数を  $j=1, 2, \dots, n$  とする。まず、 $y_{ij}$  より平均値  $m_i$  と標準偏差  $\sigma_i$  を計算し、さらにこの値を用いて基準化値  $z_{ij}$  を次式(1)により求める (表4)。

表3 健康診断のデータ

No.	実の家族の病気							...	心電図		
	祖父母				父母				3	7	9
	脳卒中	心臓病	高血圧	がん	脳卒中	心臓病	高血圧				
1	0	0	0	1	0	0	0	...	0	0	0
2	0	0	0	0	0	0	0	...	0	0	0
3	0	0	0	1	0	0	0	...	0	0	0
4	0	0	0	0	0	0	0	...	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...
159	0	0	0	0	0	0	0	...	1	0	1
平均	0.075	0.050	0.044	0.126	0.038	0.038	0.182	...	0.075	0.038	0.352
標準偏差	0.264	0.219	0.205	0.332	0.191	0.191	0.386	...	0.264	0.191	0.478

表4 健康診断データの基準化値

No.	実の家族の病気							...	心電図		
	祖父母				父母				3	7	9
	脳卒中	心臓病	高血圧	がん	脳卒中	心臓病	高血圧				
1	-0.286	-0.230	-0.215	2.636	-0.198	-0.198	-0.472	...	-0.286	-0.198	-0.737
2	-0.286	-0.230	-0.215	-0.379	-0.198	-0.198	-0.472	...	-0.286	-0.198	-0.737
3	-0.286	-0.230	-0.215	2.636	-0.198	-0.198	-0.472	...	-0.286	-0.198	-0.737
4	-0.286	-0.230	-0.215	-0.379	-0.198	-0.198	-0.472	...	-0.286	-0.198	-0.737
...	...	...	...	...	...	...	...	...	...	...	...
159	-0.286	-0.230	-0.215	-0.379	-0.198	-0.198	-0.472	...	3.500	-0.198	1.356

表5 基準化値の相関行列

	実の家族の病気							...	心電図				
	祖父母				父母				3	7	9		
	脳卒中	心臓病	高血圧	がん	脳卒中	心臓病	高血圧						
実の家族の病気	祖父母	脳卒中	1	0.152	0.055	-0.037	-0.057	0.068	0.173	...	0.009	0.068	-0.111
		心臓病	0.152	1	-0.049	0.086	-0.046	0.105	-0.034	...	-0.066	-0.046	-0.170
		高血圧	0.055	-0.05	1	0.103	-0.042	0.440	0.137	...	0.055	-0.042	0.034
		がん	-0.037	0.09	0.103497	1	-0.075	0.223	0.017	...	-0.108	-0.075	0.038
	...	...	...	...	...	...	...	...	...	...	...	...	...
心電図		9	-0.111	-0.170	0.034	0.038	-0.008	-0.008	-0.007	...	-0.011	-0.077	1

表6 相関行列の逆行列

2.413	-0.097	-0.173	-0.129	0.467	-0.419	-0.718	...	0.203	-0.318	0.394
-0.097	1.869	0.017	-0.078	0.218	0.025	0.005	...	0.262	0.166	0.596
-0.173	0.017	2.131	0.045	0.173	-0.668	0.057	...	-0.325	-0.199	-0.843
-0.129	-0.078	0.045	2.155	0.266	-0.378	-0.050	...	0.326	0.159	0.967
...	...	...	...	...	...	...	...	...	...	...
0.394	0.596	-0.843	0.967	0.286	0.086	-0.713	...	3.771	3.402	15.323

$$z_{ij} = \frac{y_{ij} - m_i}{\sigma_i} \quad (i=1, 2, \dots, k \quad j=1, 2, \dots, n) \quad (1)$$

次に基準化値  $z_{ij}$  の相関行列  $R$  を次式(2)で求める(表5)。

$$R = \begin{bmatrix} 1 & r_{12} & \dots & r_{1k} \\ r_{21} & 1 & \dots & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & \dots & 1 \end{bmatrix} \quad (2)$$

ここに、項目  $i$  と項目  $j$  (この  $j$  は式(1)の  $j$  とは異なる) の相関係数  $r_{ij}$  は次式(3)となる。

$$r_{ij} = r_{ji} = \frac{1}{n} \sum_{l=1}^n z_{il} z_{jl} \quad (3)$$

相関行列  $R$  の逆行列  $R^{-1}$  の成分を  $a_{ij}$  とし、逆行列を  $A$  とすれば、

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \dots & a_{kk} \end{bmatrix} = \begin{bmatrix} 1 & r_{2r} & \dots & r_{1k} \\ r_{2r} & 1 & \dots & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & \dots & 1 \end{bmatrix}^{-1} \quad (4)$$

この(4)式  $A$  が基準空間である(表6)。

この基準空間の要素を使い、Mahalanobis の距離  $D^2$  を次式(5)により求める。

$$D_l^2 = \frac{1}{k} \sum_{i=1}^k \sum_{j=1}^k a_{ij} z_{il} z_{jl} \quad (l=1, 2, \dots, n) \quad (5)$$

この逆行列に対し、「健康」か「健康ではない」か不明な被験者のデータを式(5)にあてはめ、基準空間からの Mahalanobis の距離を求める。この基準空間に  $C$  判定(要治療・治療中)のデータをあてはめ、距離を計算し、距離から  $A$  判定が可能か検討する。

Mahalanobis の距離をイメージ化すれば、図1のようになる。すなわち、(a)と(b)では、ユークリッドの距離は  $A$  と  $B$  では同じでも、(b)のように相関性の高い場合には、 $A$  と  $B$  の分布の中心からの距離の意味は異なり、これを表したのが Mahalanobis の距離である。なお、MTS 法では、図1のような正規分布の仮定は必要とはしていない。

### 3. Mahalanobis の距離を用いた健康の予測のための項目選択

項目選択の方法を示す。第一水準を「基準空間作成にこの項目を用いる」、第二水準を「基準空

図1 Mahalanobis・Taguchi System 法とは

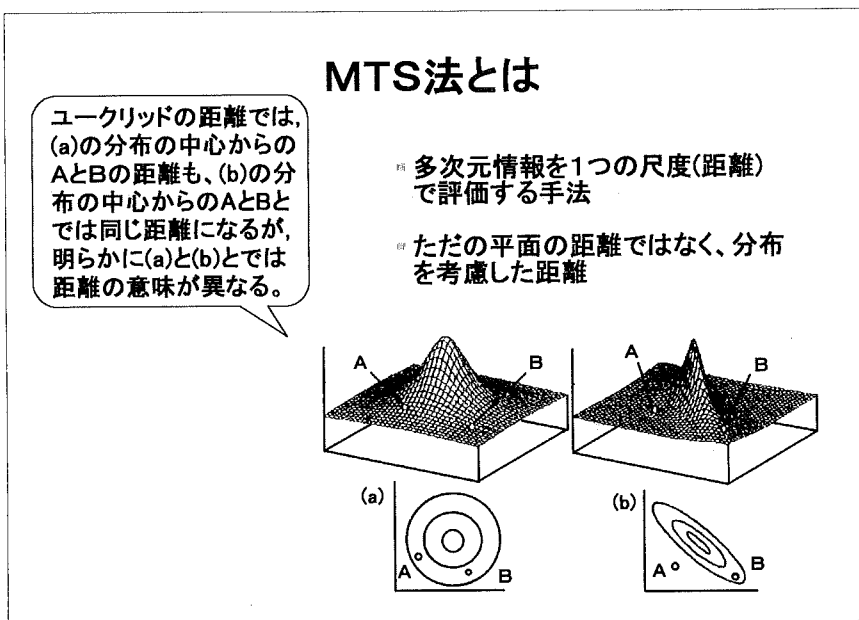
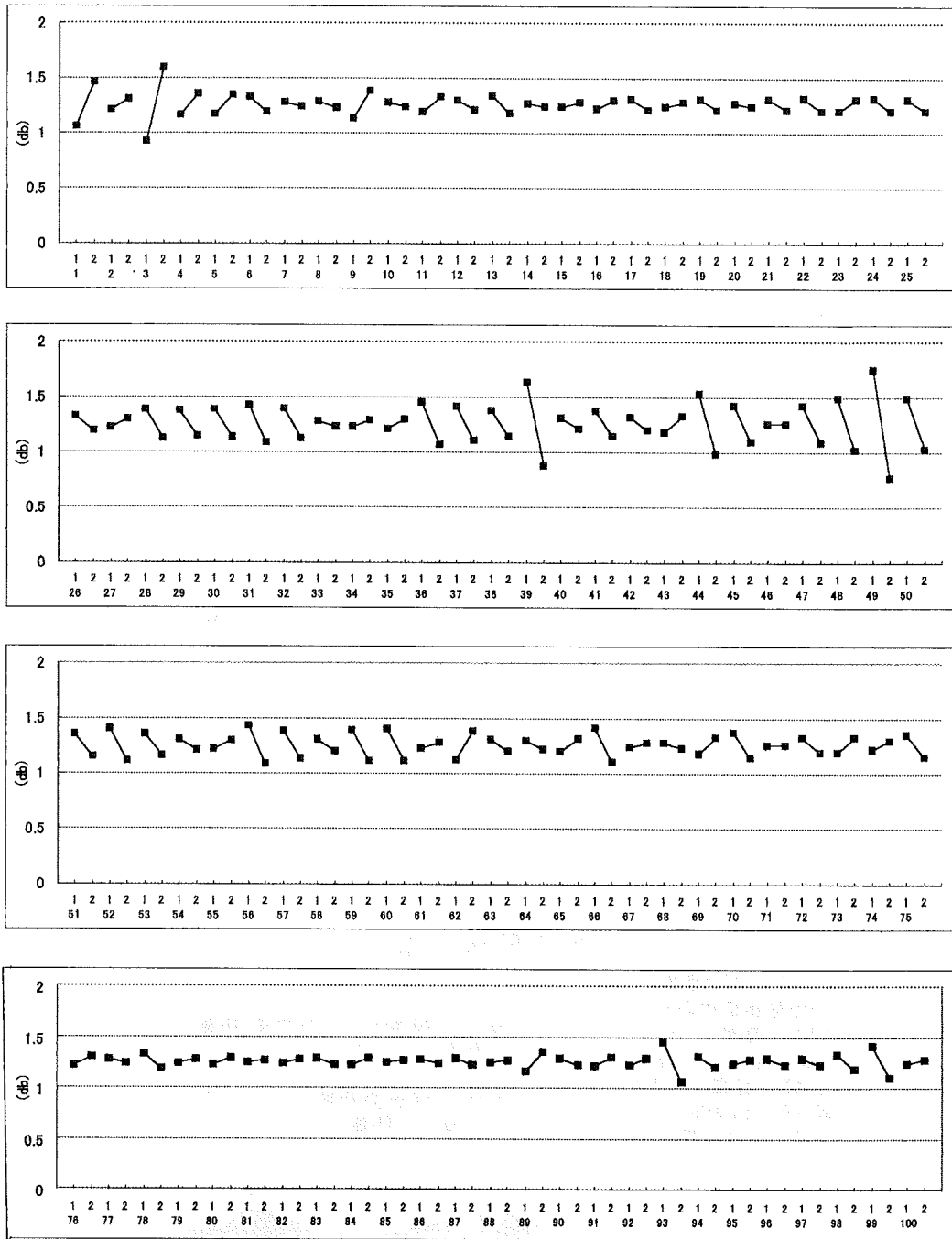


図2 項目選択の望大特性のSN比の要因効果図



縦軸はSN比を示す

間作成にこの項目を用いない」として各水準のSN比の平均値を求め、図2の要因効果図を作成する。左上がりの項目は健康診断の判定に重要な

項目（項目6, 7, …）であり、右上がりの項目は判定に関係しない項目（項目1, 2, …）となる。まず、項目の数により2水準系の直交表を選択

表7 L128直交表

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	...	115	116	117	118	119	120	121	122	123	124	125	126	127	SN比	
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	...	1	1	1	1	1	1	1	1	1	1	1	1	1	1	$\eta^1$
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	...	2	2	2	2	2	2	2	2	2	2	2	2	2	2	$\eta^2$
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	...	2	2	2	2	2	2	2	2	2	2	2	2	2	2	$\eta^3$
4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	...	1	1	1	1	1	1	1	1	1	1	1	1	1	1	$\eta^4$
5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	...	2	2	2	2	2	2	2	2	2	2	2	2	2	2	$\eta^5$
6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	...	1	1	1	1	1	1	1	1	1	1	1	1	1	1	$\eta^6$
7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	...	1	1	1	1	1	1	1	1	1	1	1	1	1	1	$\eta^7$
8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	...	2	2	2	2	2	2	2	2	2	2	2	2	2	2	$\eta^8$
9	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	...	1	1	1	1	1	2	2	2	2	2	2	2	2	2	$\eta^9$
10	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	...	2	2	2	2	2	1	1	1	1	1	1	1	1	1	$\eta^{10}$
11	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	...	2	2	2	2	2	1	1	1	1	1	1	1	1	1	$\eta^{11}$
12	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	...	1	1	1	1	1	2	2	2	2	2	2	2	2	2	$\eta^{12}$
13	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	...	2	2	2	2	2	1	1	1	1	1	1	1	1	1	$\eta^{13}$
14	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	...	1	1	1	1	1	2	2	2	2	2	2	2	2	2	$\eta^{14}$
15	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	...	1	1	1	1	1	2	2	2	2	2	2	2	2	2	$\eta^{15}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
115	2	2	1	2	1	1	2	1	2	2	1	2	1	1	2	...	2	1	2	2	1	2	1	1	2	1	2	1	2	1	$\eta^{115}$
116	2	2	1	2	1	1	2	1	2	2	1	2	1	1	2	...	1	2	1	1	2	1	2	2	1	2	1	2	1	1	$\eta^{116}$
117	2	2	1	2	1	1	2	1	2	2	1	2	1	1	2	...	2	1	2	2	1	2	1	1	2	1	2	1	2	2	$\eta^{117}$
118	2	2	1	2	1	1	2	1	2	2	1	2	1	1	2	...	1	2	1	1	2	1	2	2	1	2	1	2	1	1	$\eta^{118}$
119	2	2	1	2	1	1	2	1	2	2	1	2	1	1	2	...	1	2	1	1	2	1	2	2	1	2	1	2	1	1	$\eta^{119}$
120	2	2	1	2	1	1	2	1	2	2	1	2	1	1	2	...	2	1	2	2	1	2	1	1	2	1	2	2	2	1	$\eta^{120}$
121	2	2	1	2	1	1	2	2	1	1	2	1	2	2	1	...	1	2	1	1	2	2	1	1	2	1	2	2	2	1	$\eta^{121}$
122	2	2	1	2	1	1	2	2	1	1	2	1	2	2	1	...	2	1	2	2	1	1	2	2	1	2	1	1	1	2	$\eta^{122}$
123	2	2	1	2	1	1	2	2	1	1	2	1	2	2	1	...	2	1	2	2	1	1	2	2	1	2	1	2	1	1	$\eta^{123}$
124	2	2	1	2	1	1	2	2	1	1	2	1	2	2	1	...	1	2	1	1	2	2	1	1	2	1	2	2	2	1	$\eta^{124}$
125	2	2	1	2	1	1	2	2	1	1	2	1	2	2	1	...	2	1	2	2	1	1	2	2	1	2	1	1	1	2	$\eta^{125}$
126	2	2	1	2	1	1	2	2	1	1	2	1	2	2	1	...	1	2	1	1	2	2	1	1	2	1	2	2	2	1	$\eta^{126}$
127	2	2	1	2	1	1	2	2	1	1	2	1	2	2	1	...	1	2	1	1	2	2	1	1	2	1	2	2	2	1	$\eta^{127}$
128	2	2	1	2	1	1	2	2	1	1	2	1	2	2	1	...	2	1	2	2	1	1	2	2	1	2	1	1	1	2	$\eta^{128}$

する。11項目以下ならばL<sub>12</sub>直交表, 15項目以下ならばL<sub>16</sub>直交表, 31項目以下ならばL<sub>32</sub>直交表……, となる。選択した直交表の因子に項目を割り振る。全因子において, 第一水準は「基準空間作成にこの項目を用いる」, 第二水準は「基準空間作成にこの項目を用いない」という項目のパラメーター設計を行う。

直交表の列は空欄でも支障ない。1行目はすべて1なので, 全項目を使用して基準空間を作成する。2行目は約半分が1なので, 約半分の項目のみで基準空間を作成する。3行目……, 以下同様である。今回の健康診断の場合は, 項目は100個あるので, L<sub>128</sub>直交表を使用し, 全部で128個の基準空間を作成する。次に, 作成した128個の基準空間に対して距離が離れて欲しいデータ(異常データ)の個々の距離を求める。異常データの数は

は多ければ多いほどよい。しかし, 条件によっては異常データの数を集めることができない場合がある。その場合は2~3個でも可能である。1行目の基準空間から求めた異常データ距離をD<sub>1</sub>, D<sub>2</sub>, …, D<sub>n</sub>とする。異常データで正常空間からの距離が離れるのが望ましいので, 望大特性のSN比ηで評価する。

$$\eta = -10 \cdot \log \frac{1}{n} \left[ \sum_{i=1}^n \frac{1}{D_i^2} \right] \quad (6)$$

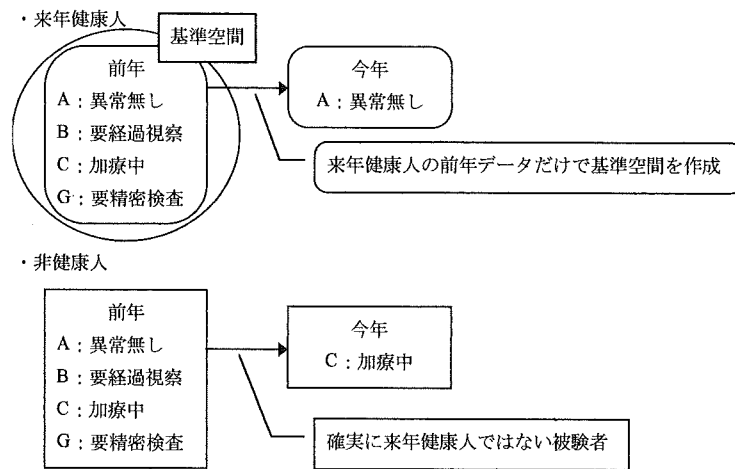
表7に, L<sub>128</sub>直交表の一部を示した。L<sub>128</sub>直交表のAの場合,

$$A_1 = (\eta_1 + \eta_2 + \eta_3 + \dots + \eta_{62} + \eta_{63} + \eta_{64}) / 64$$

$$A_2 = (\eta_{65} + \eta_{66} + \eta_{67} + \dots + \eta_{126} + \eta_{127} + \eta_{128}) / 64$$

で各水準のSN比ηの平均を求め, 次にSN比の差(A<sub>1</sub>-A<sub>2</sub>)を求める。この差が大きい項目が

図3 健康状態の予測の場合の基準空間のイメージ



重要となる。望大特性のSN比とは、品質工学においては大きいほうが望ましく、かつ個々のデータのばらつきが小さいということを評価するために作られた尺度であり、デシベル値(対数)をとることにより、データが $-\infty$ から $+\infty$ に分布するように変換して、加法性を持たせたものである。このSN比の差の大きい項目が今回の健康診断において重視される項目である。

#### 4. Mahalanobis の距離を用いた項目選択の結果と次年度の健康の予測

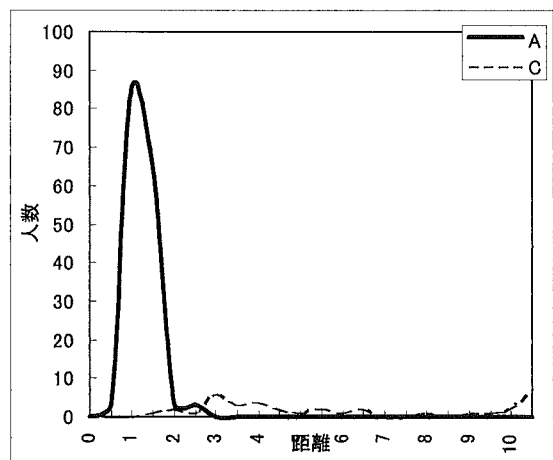
次年度A判定かどうかの健康の予測に用いる基準空間のとりかたを図3に示す。項目選択後、2年目A判定を対象として選び、その対象者の1年目のデータから作成した基準空間を用いて、1年目の基準空間と、A判定とC判定の対象者のMahalanobisの距離を求め次年度A判定の予測が可能か検討する。したがって2年目A判定なら、1年目の判定はどの判定でもかまわない。つまり、その1年目のデータで基準空間を作成し、2年目A判定を次年度の予測と見立てて検討する。

### III 研究結果

#### 1. Mahalanobis の基準空間と距離によるA判定とC判定の分離の可能性

A判定とC判定の分離の可能性を検討するために、2年間にわたる健康診断結果を用い、2年間A判定者で作成した基準空間に対して、健康

図4 健康診断C判定の、基準空間(A判定)からの距離のヒストグラムのグラフ



人(A判定)159人、現在療養中(C判定)37人の距離を求めた。結果を図4および表8に示す。A判定は、Mahalanobisの距離が2.5以下に全例入り、またC判定では距離は1.5以上になった。しきい値として距離を2.0で区切ると、A判定では、159人中3人(1.9%)がC判定と誤診され、逆にC判定では、37人中3人(8.1%)がA判定とされるという結果になり、また距離を1.5で区切ると、A判定では159人中6人(3.8%)がC判定とされるが、C判定は1人のみ(2.7%)と



表8 基準空間からの Mahalanobis の距離の分布

距離 D	A 判定 (人)	C 判定 (人)	A 判定 (%)	C 判定 (%)
0	0	0	0	0
0.5	3	0	2	0
1	85	0	55	0
1.5	65	1	96	3
2	3	2	98	8
2.5	3	1	100	11
3	0	6	100	27
3.5	0	3	100	35
4	0	4	100	46
4.5	0	2	100	51
5	0	1	100	54
5.5	0	2	100	59
6	0	1	100	62
6.5	0	2	100	68
7	0	0	100	68
7.5	0	0	100	68
8	0	1	100	70
8.5	0	0	100	70
9	0	1	100	73
9.5	0	1	100	76
10	0	2	100	81
11以上	0	7	100	100
合計	159	37		

なる。つまり、しきい値が2.0では、Aであるとの判定率は98.1%、Cであるとの正診率は91.9%、しきい値1.5では、Aとの判定率は96.2%、Cとの正診率97.3%であった。

2. Mahalanobis の距離を用いた健康の予測のための項目選択

次年度の健康状態の予測のために、2年目 A 判定の人の、1年目のデータで新たに基準空間を作成する。

基準空間に使用したデータは、2年目が健康な人 (A 判定) の1年目の健康診断のデータで、数は336人、項目数は100、全員欠側値なしである。さらに非健康人としては、2年目に慢性疾患に治療中の判定 (C 判定) を受けている137人の1年目のデータを使用した。これらの人の1年目の健康状態の内訳は表2に示すとおりであり、2年目 A 判定でも1年目はC判定のものが3.5%、また2年目 C 判定でも1年目 A 判定が1.4%含まれて

表9 項目選択で有効と判断された項目

	有効と判断された項目
実の家族の病気	脳卒中, 心臓病, 高血圧, がん
嗜好品	タバコ
体側	身長, 体重
血圧	最高, 最低
検尿	蛋白
血液検査	GOT, GPT, ALP, $\gamma$ GTP, LDH, T. Bil, ZTT, TG, T. Chol, HDL, Cr, UA, BS, HbA1c, WBC, RBC, Hb, Ht, MCH, MCHC
	性別, 年齢, 自覚症状, 既往歴, 胸部 X-P, 心電図, 視力

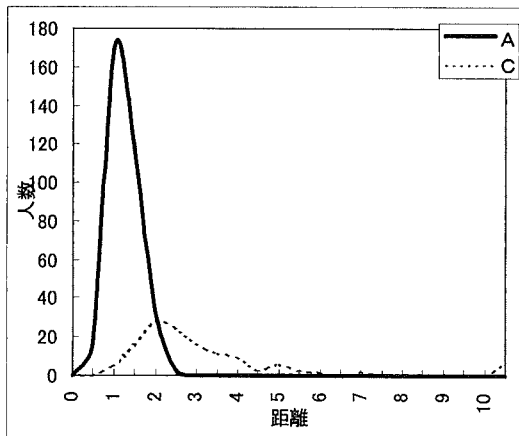
いる。

項目選択作業を行い、予測に関係の薄い項目を除外して望大特性の SN 比の高い項目で再度基準空間を作り直した。データ数は前回同様336個であるが、項目数は項目選択後100から66へ減らすことが可能であった。有効な検査項目を表9に示した。さらにこのなかでも特に SN 比の差の大きかったものは、性、年齢、既往歴、血圧などのほかに、GOT, GPT,  $\gamma$ GTP, T. Chol, T. G, Cr, UA, BS, HbA1c, Hb, MCH など表9に示したとおりであり、これらの項目を見ると通常健康診断において重視される血液検査の項目である事がわかる。

3. 項目選択後の今年度のデータから次年度の健康の予測

項目選択後の基準空間からの距離のデータは、表10と図5に示すようになった。すなわち、2年目 A 判定の人を選んで、その人の1年目のデータから作成した基準空間を用いて、1年目の基準空間と、A 判定の人、C 判定の人の距離を求めたものである。A 判定かどうかの予測のために、2年目 A 判定の対象者の1年目のデータで基準空間を作成するため、1年目はさまざまな判定でかまわない。ここでしきい値を1.5として、しきい値とその距離 D を分類すれば、A 判定と C 判定の割合は表11のようになった。つまり、次年度 A 判定と予測される割合が90%であり、C 判定でありながら A 判定と予測される割合が15%であ

図5 項目選択後の基準空間からの、2年目のA判定とC判定の、1年目のデータで作成した基準空間からの、Mahalanobisの距離のグラフ



った。

#### IV 考 案

健康診断<sup>2,5)</sup>、病気の治療<sup>6)</sup>、薬効の評価<sup>7)</sup>などに品質工学の立場での提案と研究が行われている。品質工学は田口が提唱した研究開発の合理化のための評価技術で、世界的にはTaguchi Method<sup>1)</sup>として知られており、いわゆる統計的方法、実験計画法などの不十分さを克服した新しい評価技術である。

医学の方面においては、兼高<sup>8)</sup>により、健康診断の信頼性の評価、治療効果の評価などについて検討されており、これを体系化したものがMahalanobis・Taguchi System (MTS) 法<sup>1)</sup>である。

本研究においては、MTS法を用いて、健康診断における予測的な評価について検討することを目的とした。MTS法による兼高<sup>8)</sup>、長谷川<sup>9)</sup>らの検討をさらに発展させて、健康診断の効率化を図るために項目選択を行い、また今年度の健康診断データより、次年度の健康状態の予測が可能であるかを検討した。

MTS法により、健康人の基準空間から健康診断の確かさを評価するためには、基準空間の作成の段階で、一人の医師により同一の基準で健康であるかの判定を行うことが必要である。しかし現実には数千人の健康診断の判定は複数の医師で行っている。よって、基準空間の作成には数百人の

表10 項目選択後の基準空間からの距離とパーセントの分布

距離 D	A 判定 (人)	C 判定 (人)	A 判定 (%)	C 判定 (%)
0	0	0	0	0
0.5	15	0	4	0
1	170	5	55	4
1.5	117	15	90	15
2	31	28	99	35
2.5	3	25	100	53
3	0	17	100	66
3.5	0	12	100	74
4	0	10	100	82
4.5	0	3	100	84
5	0	26	100	88
5.5	0	3	100	91
6	0	2	100	92
6.5	0	0	100	92
7	0	2	100	93
7.5	0	1	100	94
8	0	0	100	94
8.5	0	0	100	94
9	0	1	100	94
9.5	0	1	100	95
10	0	0	100	95
10.5	0	7	100	100
合計	336	137		

データを必要とすることから、本研究では、2年間の健康診断のデータより、2年間A判定者を健康人として基準空間を作成し、今年度のデータから次年度の健康を予測した。

はじめに、基準空間をどのように作成するかが重要な問題であり、なるべく安定したデータで基準空間を作成する。今回は健康人で基準空間を作成するが、疾患患者のデータで基準空間を作成しようとした場合は、疾患別の項目の検討も必要であるが、疾患による異常値は値の幅が広く、その結果データのばらつきが大きくなるために基準空間の作成には不適當である。また、対象とする健康人でも、対象者に偏りが認められると不適當な場合がある。例えば、体育大学の学生なのか、老人ホームのお年寄りなのか、サラリーマンなのか、それぞれに合ったデータで基準空間を作る必要がある。体育大学の学生で作った基準空間で

表11 しきい値を1.5にした場合の健康診断の分類の予測

しきい値	<1.5	>1.5	計
予想される A 判定	90%	10%	100%
来年度の判定 C	15%	85%	100%

は、いくら健康な老人ホームのお年寄りでも、Mahalanobis の距離が大きくなり、病人扱いされてしまう。逆に、老人ホームのお年寄りで基準空間を作った場合は、体育大学の学生の Mahalanobis の距離が大きくなると考えられる。ここでは、20歳から50歳代の、一般に健康診断の対象となるサラリーマンの健康を基にして基準空間を作成した。その結果、健康人で作成した基準空間からの Mahalanobis の距離が2.0以下であれば、98.1%の割合では健康人であり、2.0より大きな距離になると、健康人ではないという判定、つまり何か病気を持っている可能性があるという医師の判断なく MTS 法のみで診断が可能になる。

健康診断の判定のほかに、Mahalanobis の距離の特筆すべき点は、予測に用いることが可能である。つまり、ある人の現在の健康診断のデータから、次年度健康かどうかの予測の可能性を検討した。さらに予測の精度をあげるために、100項目より有効な項目の選択を行った。まず、2年分の健康診断のデータを用い、2年目健康と判断された対象者の1年目の健康診断のデータで基準空間を作成する。1年目は健康でも健康でなくてもかまわない。Mahalanobis の距離による予測、判定の後に、基準空間作成の際に予測、判定に必要な項目と必要ではない項目の評価(項目選択)を行う。つまり、Mahalanobis の距離による予測、判定能力を下げることなく、項目を減らすことが可能なため、健康診断に意味のある項目が選べられ、健康診断のコストの削減と効率化が可能である。

項目選択によって、100項目が66項目に減らせるが、項目選択することにより、MTS 法の精度も同時に上がる。今回の検討では、100項目を66項目に減らせるわけで、単純に言えば、検査コストや時間などを三分の二に削減でき、健康診断の効率化が可能である。また、項目選択で残った項

目は、判定精度に大きな影響を与える項目であり、年齢や性も判定へ影響すると考えられるが、年齢の場合は、どこで区切るかが問題であり、また、データ数が少ないため、今後データ数を増やして検討したい。

項目選択後、Mahalanobis の距離のしきい値を1.5とすると、A と予想されるのが90%であり、C でありながら A のグループに入るのが15%であった。つまり、しきい値を1.5とすれば、90%の割合で、次年度 A 判定を予測できた。今年度の健康診断の結果から、精度よく次年度の健康状態が予測できるなら、少なくとも毎年健康診断を行う必要は無くなり、次年度 A 判定の予測がつかない場合は、次年度は必要ないという根拠になり、項目選択とあわせて健康診断の効率化がはかれる。

従来から使われている判別関数分析による識別結果と MTS 法の比較では、長谷川ら<sup>4)</sup>によると、従来の判別関数分析の方が識別力が47.8%に減少し、MTS 法の方が優れていると報告しており、将来の結果の予測については、高精度の応用が可能であると思われる。

今後は、健康状態の予測の精度を上げるために、3年以上の時系列データを用いて計算をし検討すれば、より高い比率で予測できると思われる。さらに、一般健康診断項目だけではなく、疾患別の病態を把握できる検査項目で正常空間を作成することにより、各々の病状の重みや進行度、治療効果の判定に応用できると思われる。

## V 結 語

健康診断の判定に、Mahalanobis の距離が利用できるかを検討し、98.1%の正診率で十分利用できる事が判明した。さらに来年度の健康状態の予測も行い、まだ完全に予測はできないが、90%の割合で可能であるという比較的良好な結果が得られた。また、項目選択により、従来の検査項目を約三分の二に削減でき、予測と組み合わせると、次年度健康と予測できる場合は健康診断が不用になり、かなりの効率化がはかれる。今回は2年間のデータのうちの1年だけを使い予測を行ったが、時系列データを入れた健康状態の予測を行うと精度が向上すると言われているので、今後は3年間のデータを集め、そのうちの前2年間を使い、来

年の健康状態の予測を行う事が必要になると思われる。

(財)日本規格協会における実験計画法研究会 (DERG) における田口玄一博士, 兼高遠貳博士などの御指導に対して深く感謝の意を表します。

(受付 '98. 6. 8)  
(採用 '99. 3.18)

## 文 献

- 1) 田口玄一, 品質工学タグチメソッド入門, 標準化と品質管理 1993; 46(5): 88-94
- 2) 兼高遠貳, マハラノビスの距離を応用した健康診断の判定, 品質工学 1997; 5(3): 35-44
- 3) 小島通代, 長谷川良子, マハラノビスの距離の応用事例—パターン認識による脳疾患患者の排尿自立達成の予測, 標準化と品質管理 1994; 47(3): 38-44
- 4) 長谷川良子, 小島通代, マハラノビスの距離を応用した脳疾患患者の排尿自立達成の予測, 品質工学 1997; 5(5): 55-63
- 5) 長谷川良子, マハラノビスの距離を用いた健康診断と欠損値対策, 品質工学 1997; 5(5): 46-54
- 6) 兼高遠貳, マハラノビスの距離の応用事例—胃潰瘍と十二指腸潰瘍の心理学的特性の解析, 標準化と品質管理 1994; 47(11): 41-46
- 7) 兼高遠貳, マハラノビスの距離を応用した薬剤の効果判定, 標準化と品質管理 1992; 45(10): 40-44
- 8) 兼高遠貳, 医療のための実験計画[1], 標準化と品質管理 1987; 40(10): 57-64

## PREDICTIVE EVALUATION AND EFFICIENT MANAGEMENT OF MEDICAL EXAMINATIONS USING MAHALANOBIS TAGUCHI SYSTEM METHOD.

Hisato MAKAJIMA\*, Kei TAKADA<sup>2\*</sup>, Hiroshi YANO<sup>3\*</sup>, Ichiro TAKAGI\*,  
Yuka SHIBAMOTO\*, Masayoshi YAMAUCHI\*, Gotaro TODA\*

**Key words:** Mahalanobis · Taguchi System, Prediction, Mahalanobis distance, Medical examination, Quality engineering

The Mahalanobis-Taguchi System (MTS), an evaluative technique used in quality engineering, was utilized in a system analysis of medical examinations to determine practicability, to differentiate cost reduction benefits of specific components for increasing efficiency, and to determine possibility of utility on forecasting of future health condition of individuals.

### Methods:

1. Medical examination date over a two year period were utilized to differentiate subject into two groups: Group A—Healthy or normal; and Group C—Patient group requiring treatment or currently undergoing treatment. The Mahalanobis distance for Group C from the Mahalanobis space of Group A was calculated for each subject to determine differentiability.

2. Utilizing the MTS method specific components of the examination system were selected and analyzed for effectiveness in diagnosis and screening to enhance efficiency of the examination.

3. Group A subjects (determined from the two year data), were analyzed using the data at the end of 1 year to develop the Mahalanobis space and the Mahalanobis distance at the 2<sup>nd</sup> year point was calculated and analyzed.

### Results:

1. Subjects who were classified as group A at both 1<sup>st</sup> and 2<sup>nd</sup> year points numbered 159. With a cutoff point of 2.0 for Mahalanobis distance 98.1% of the healthy subjects (Group A) were classified as belonging to the healthy group, and 91.9% of C group as belonging to the patient group.

2. Blood pressure, GOT, GPT,  $\gamma$ -GTP, T Chol, TG, UA, Cr, BS, HbA1c, Hb, and MCH were found to be effective components in the examination. Based on this result the items included in the examination can be reduced by approximately 60%.

3. With cutoff value of 1.5, the results of 90% of subjects 1 year later could be predicted. The prediction that 15% of Group C subjects would improve to join Group A could be made. Overall, the results 1 year later could be predicted at a level of 90% accuracy.

### Conclusion:

These results suggest that the Mahalanobis-Taguchi System technique can accurately assess medical examination quality and can also be used, with reasonable accuracy, to predict the condition of individuals one year later. Therefore, the MTS can be used to help determine which patients should be receiving the examination 1 year later. In addition, components of the examination can be assessed using the MTS for benefits and costs to determine levels of essentiality of components. As a result, it can be used an effective tool for development and improvements in the examination system. With further development in the technique, it may become possible for use as a tool in disease diagnosis and treatment.

\* Department of Internal Medicine I, The Jikei University School of Medicine.

<sup>2\*</sup> Seiko Epson Co. LTD.

<sup>3\*</sup> Japanese Standards Association.