R. A. Fisher以後の判別分析の新理論（2）

－Alon他のマイクロアレイデータの130個の癌の基本遺伝子(BGSs)の検証(1)－

新　村　秀　一

1. はじめに

30年以上かけて（Golub et al., 1999），世界中の統計や医学の研究者が，「遺伝子情報の高次元空間データの統計解析」と称して癌の遺伝子を特定することを研究し，多くの論文を出してきたが，はっきりした結果を提案していなかった。彼らが研究に用いている手法[1]は，癌患者と正常者の2群の分布の平均に差があるはずだという希望的観測から次のようにアプローチしている。ここで「癌遺伝子」という定義を明確に示した論文は，狭い文献調査では探せなかったが，一応正常と癌患者の違いを示す遺伝子を「(統計的) 癌遺伝子」と呼ぶことにする。少なくとも統計的に見つけた癌遺伝子の中に，医学的に意味のある真の癌遺伝子が含まれることを期待している。一般的なアプローチは，次の通りである。

1） 癌遺伝子は，「一元配置の分散分析のt検定」で大きなt値をもつと考えることは不自然ではない。しかし，今までの検証では少なくとも1個の遺伝子で，2群を完全に分けるものは存在しない。さらに，2群に差のない小さなt値をもつ遺伝子が，Basic Gene Set or Subspace（BGS）[2]に多く含まれている。

2） クラスター分析をすれば，2群を分ける変数とケースの関係が分かるはずであるという期待から，主としてヒートマップなどを用いてアプローチすることが試みられている。しかし，BGSのクラスター分析でも，明確に分けることは難しい。

3） 主成分分析（Principal Component Analysis, PCA）のスコアプロットで正常群と癌患者が分かれるのではないかという思い込みで，PCAの分析が試みられているが一般的に2群の重なり具合は大きい。

4） 統計的な判別分析で2群が判別できるのではないかという期待に基づくアプローチが行

---

[1] 一元配置の分散分析とt検定，クラスター分析，PCA，統計的判別分析などが一般的。その他，統計で一般的でない手法も多い。

[2] 分析を行った６種の遺伝子情報はLSDである。LSDのデータには，最小のMNM=0の部分空間（変数の組）がありこれをBGSと呼ぶ。このBGSを含む全ての部分空間はMNM=0であることが，すでにスイス銀行紙幣データで発見し，MNMの単調減少性と呼んできた。しかし，遺伝子解析ではp次元のデータ全体をMatroskaと考え，その中に入れ子状でBGSまでのMatroskaが含まれているという構造に着目した。手法2では，残念ながらBGSを直接求めることができず，20~30次元以下の小さなMatroska（SM）を求めることができる。今回，Program4と手作業でSMを足がかりにAlon他のMicroarrayデータで全てのBGSを求めることができ，それを通常の統計手法で検証する。

われてきた。しかし，分散共分散行列に基づく判別関数は線形分離可能（Linearly Separable Data, LSD[3]）なデータの判別を理論的にできないし，誤分類数（Number of Misclassification, NM）や誤分類確率が高い（Miyake & Shinmura, 1976）。

筆者自身これらの研究の雰囲気は承知していたが，2010年に整数計画法（Integer Programming, IP）による最小誤分類数基準（Minimum NM, MNM[4]）による最適線形判別関数（Optimal Linear Discriminant Function, OLDF）の改定IP-OLDF（Shinmura, 2007b ; 2011），改定LP-OLDFと改定IPLP-OLDF（Shinmura, 2009 ; 2014b ; 新村, 2010b）と3個のSVM（Vapnik, 1995）とFisher（1936 ; 1956）のLDFとロジスティック回帰（Cox, 1958）を用いた実証研究による研究成果を2010年に出版した（新村, 2010a）。これまでFisherが確立した統計的な判別分析理論の常識とは隔絶した多くの問題を実証研究で示したので，驚きと称賛をもって受け入れられると期待したがほどんと反応がなかった。その中で，竹内啓先生（竹内, 2011）に書評していただいたことと，故牧野都治東京理科大学元教授から，「すぐに理解されないが，後世評価されるでしょう」というハガキをいただいた。私自身，判別分析の4つの問題（Shinmura, 2014a ; 2015b）のうち，学生データ（新村, 2004）を用いた実証研究で問題1[5]は完全に解決した（新村, 2007a ; 2007b）。問題2は，ハードマージン最大化SVM（H-SVM）と改定IP-OLDF以外の判別関数はLSDの判別を理論的にできないことである。この重要性に気づき2010年以降の応用研究のテーマとし，分散共分散に基づくLDFは，NMあるいは誤分類確率が非常に高いことが分かった（新村, 2011a）。そして，判別分析は回帰分析（新村訳著, 1986；新村, 1996）のように推測統計学でないという問題4を，小標本による100重交差検証法（100-fold cross validation for small sample method, Method1）を開発し解決した（Shinmura, 2013 ; 2014c ; 2015a ; 2016d）。これによって学習標本と検証標本の平均誤分類確率をM1とM2とし，M2が最小値になるモデルを選べば良いという単純明快な「モデル（Featureあるいは変数）選択法」を開発し，多くのデータで改訂IP-OLDFが他の7種のLDF[6]より良いことを実証研究で示した。しかし，判別係数の95%信頼区間の有意義な意味を見つけられなかった（新村, 2010a）。2010年以降の応用研究では，成績の得点を用いた合否判定（新村, 2011a）と日本車44車種の判別とスイス銀行紙幣データ（Flury & Riedwyl, 1988）を用いて，LSD判別の

---

[3] LSDはデータ全体が線形分離可能である用語であり，その部分空間が線形分離可能なものと区別しているが，データ全体と部分空間が線形分離可能なものをLSDと拡張解釈した方が簡単かもしれない。

[4] MNMは筆者の提案した新しい統計量である。MNM=0でもって初めて，データがLSDであることが定義でき，MNMが1以上であれば2群がオーバーラップしていることが定義できる。

[5] 改定IP-OLDF以外の判別関数は，判別超平面上のケースを正しく判別できない。これは，IP-OLDF（Shinmura, 2000a ; 2000b ; 2003-2005 ; 新村, 1980 ; 1998 ; 新村，垂水，2000）で判別係数とNMの関係と，MNMの単調減少性の事実から解明できた。

[6] 改定IP-OLDF，改定LP-OLDF，改定IPLP-OLDF，H-SVM，SVM4（Penalty c=10000），SVM1（Penalty c=1），Fisher's LDF，ロジスティック回帰

研究を行った。また，2014年に判別係数の95%信頼区間に関して，「試験の合否判定データでFisherのLDF以外は自明なLDFになることが分かった（Shinmura, 2015c）。この事実は「自明な判別関数というアイデアが実態にあった提案であり，正規分布を仮定するFisherのLDFだけがこれを求めることができない」ことが分かった。またスイス銀行紙幣データで，改定IP-OLDFだけが判別係数の95%信頼区間で有意なモデル系列を見つけた（Shinmura, 2016c）。これで4個の問題を全て解決しFisher以後の新しい判別分析の理論を確立したと確信した（Shinmura, 2016f）。しかしResearch Gate（RG）では世界の多くの研究者に受け入れられているが（新村, 2015a），分散共分散行列[7]の問題を指摘しているので，多くの日本の統計研究者はそれを前提に研究を行っており受け入れないと思っていた。それが2015年10月に富山の統計シンポジュームで研究論文に用いたMicroarrayデータが公開されていることを知り（Jeffery et al., 2006），問題5の解決を行っていないことを気づいた。しかし，比較研究を行ってきた8種のLDFで，筆者の開発した3種の改定OLDFだけが，判別分析するだけで，ケース数のn個以下の判別係数が0でなく，残りの全ての判別係数が自然に0になることが分かった[8]。そして，Matroska Feature Selection Method（Method2[9]）を僅か54日で開発し，「6種のデータセットは複数のSmall Matroska（SM）と呼ぶ遺伝子の部分空間の排他的な和集合であること」を示した（Shinmura, 2015d-2015r ; 2016a ; 新村, 2016a-c）。このようなデータ構造を持つことは，誰も想像していなかったし，統計分析としては未体験のデータ構造である。これらの事実は，スイス銀行紙幣データ，日本車データといった普通のデータでも確認できることが重要である（新村，2016b ; 2017a）。

これまで多くの研究者が高次元の遺伝子情報の統計分析を試みても何も結果は出なかったが，各SMは小標本であり統計分析が容易に行えて，癌の遺伝子解析に革新をもたらすことを期待した。2016年8月に，6種のデータセットの中で2000個の遺伝子という最も小さいAlon et al.（1999）で，130個の排他的な癌のBGSを見つけることに成功した。当初これをLINGO（Schrage, 2006 ; 新村, 2011b）でプログラム化し残りの5個のデータセットのBGSを見つけることを目標にして研究を行なってきたが，すでに得られているAlon他のBGSの詳細な統計分析を先行すべきと考えなおし，JMP（Sall et al., 2004）で評価する方法を模索した。

---

[7] 問題3は，分散共分散行列の問題の一つである。変数が一定値をとる場合に逆行列が計算できないが，一般化逆行列の技術が開発された。それがQDFに用いた場合に瑕疵がある点である。一定値に乱数を加えることで解決できる。

[8] 現在注目を集めているLASSO（Simon et al., 2013）は分散共分散行列をベースに難しい理論を展開し，判別係数を0にしようとする理論である。はたして，LSDを正しく判別できるか，そして筆者の得た結論に迫れるか，疑問である。

[9] Shinmura（2016f）では，学習標本を解析するLINGOプログラムをProgram1とし，小標本のための100重交差検証法をMethod1として，それを実行するプログラムをProgram2として公開している。Method２を実行するプログラムのProgram3は，Shinmura（2017a）で解説した。

これから一連の研究で，SMやBGSの標準的な解析法を確立し，多くの研究者のガイドラインを示すことを目的としている。

2. Method2によるSMからMethod3によるBGSへ

2.1 Method2によるSMの検証

新村（2016c）では，Method2を，Golub et al.（1999）とAlon他を用いて検証した。検証方法は次の点である。

1） LSDであるスイス銀行紙幣データと日本車データでLINGO Program2とProgram3が正しく稼働するかをまず検証し，高次元データでは検証できない点を検討して問題のないことや，これまで不明であった点を明らかにした。

2） 2つのMicroarrayデータの分析結果を，JMPを用いて検討した。即ち得られたSMが本当にMNM = 0であるかをロジスティック回帰のNMが0であることで確認した。

しかし，8月下旬にSMに含まれるBGSの選択方法を考えてLINGO Program4の作成を試みた。これまで提案してきた全ての判別モデルで探索する統計手法の利用（Shinmura, 2016a）なしで，LINGOだけで行える方法がAlon他のデータをProgram4を手作業で分析して分かった（Method3）。昨年10月以来，Method2とProgram3で6種類のデータの膨大な分析を行ってきた。そこで得られたSMからBGSを決定し，BGSに含まれない遺伝子を集めて再度BGSを決定する方法1と，直接データからBGSを決定し，それを省いた部分空間から次のBGSの決定を繰り返し行う方法2のうち，どちらが計算時間などを含めて有利なのかは決めかねている。そこで，本稿ではとりあえず方法2で得た手作業を含む分析結果を以下に示す。

2.2 Method3によるAlon他の130個のBGSの検討

Method2では，「Alon他のデータは64個のSM（合計1152個の遺伝子）と848個のMNMが1以上の遺伝子の背反集合である」ことが分かった。Method3では，MNM=0になる130個のBGS（1995個の遺伝子）と残り5個の遺伝子（MNM≧1）の131個の排反な和集合であることが分かった。Method2で64個のSMを求めたが，各SMの中には少なくとも1個以上のBGSが含まれる。その上で，SMの中でBGSに選ばれなかった遺伝子を集めて再度BGSを探せば，高々66個の別のBGSが選ばれたことを示している。引き続きMethod2で直接BGSを見つけることは検討するが，現時点ではProgram3の中に，Program4を組み込んで，

1） SMを探し，

2） 得られたSMの中でBGSを決定し，

3） このBGSをBig Matroskaから省いて，探索を繰り返す方法を，方法1とする。

一方，Method2を捨てて，Method3だけでBGSを直接探す方法2をとるかは未検討である。当

面は，方法2を手作業で行うことにした。

この130個のBGSを統計的に得られる癌遺伝子として，医学的な立場からの検証と，統計的な検証を行う必要がある。

2.2.1　医学的な検証

得られた130個のBGSが医学的に癌診断に利用できるか否かの検証は，次のことが考えられる。

(1) 遺伝子学的な考察

得られた130組の遺伝子の組は，遺伝子学の専門家がその組み合わせを見れば，何か遺伝学的に意味があるのではないかと考えている。そのために専門家と共同研究の必要がある。長野県の大病院の創業者と連絡を取り，専門家に打診していたが，長年の癌の病理標本の資料室が放火され研究協力どころではないと断りの連絡を10月27日に受け希望が閉ざされた。Research Gate（RG）に共同プロジェクトの募集を公開し，Alon本人にも11月8日に共同研究を申し込んだが返事は貰えていない。

130個のBGSに含まれる遺伝子をRGに掲載することも考えた。本来であれば論集に掲載すべきであるが，スペースを取るので省く。恐らく遺伝子名の組み合わせが分かれば，なぜそれらがBGSとして一括りにされるかを遺伝子学的に分かり新しい知見が得られるかもしれない。あるいは，個々のBGSは容易に普通の統計手法で分析し，有効な知見が得られるかもしれない。これまでおよそ30年以上, 高次元の遺伝子情報を直接統計手法でアプローチし, 種々の方法が提案されたが決め手に欠いていた。もともと遺伝子空間は，BGSあるいはSMの背反集合であるので，それを前提としない分析は，全く意味がないと考える。

(2) 130個の判別関数による癌診断

Alon他のデータは，他と比べて僅か2000個と遺伝子数が少なく，それが130組のBGSと5個の遺伝子の組に分かれた。彼らの論文には，クラスター分析による分析結果しかなく，またなぜ2000個に絞ったかの理由は読み取れなかった。

しかし，130個の判別関数を作成し，同じ方法で集めた正常と癌患者と，治療によって予後が良い患者のデータがあれば提供してもらい，それを判別すれば著しい癌診断の結果が得られるのではないかと考えている。すなわち,「正常と癌患者は正しくそれらの群に判別され，予後の良い癌患者は130個の判別結果のうち何個かは正常に誤判別されるはずである。130個全てで誤判別されれば，5年生存率を待たないで完治したと診断できるのではないか」と考えている。とりあえずは，3つのグループの5 〜 6例程度の検証データが得られれば簡単に筆者の作業仮説は検証できる。

2.2.2　統計学的な検証

本論文では，これまでの研究で使われてきた，一元配置の分散分析とt検定，クラスター

分析，PCA，統計的判別分析でBGSの検証を行う。高次元遺伝子空間を直接これらの手法で分析しても何も良い結果が得られていない事実がある。しかし，SMやより次元数の小さなBGS[10]が分かったので，これらを分析すれば癌診断に有効な情報が得られると考えた。現在，模索中であるが，どうも筆者の開発した3種のOLDFに加えてロジスティック回帰が，これらの部分空間を正しく線形分離可能な最小の部分空間であることを確認した。Method3を改良したLINGOのProgram4で自動的にBGSを求めることには成功していないが，手作業を介してAlon他で全BGSを求めることに成功した。ここで冗長なSMを分析することと，BGSを分析することのどちらが医学的に有意義かは分からないが，全BGSが分かった場合は全SMの検証より優先した方が良いと考えている。

3. 130組のBGSの統計的判別分析による検証

3.1 ロジスティック回帰，QDFとFisherのLDFによる検討

130組のBGSを，OLDF以外のロジスティック回帰，2次判別関数（Quadratic Discriminant Function, QDF）とFisherのLDFをJMP（Sall et al. 2004）で分析し表1のようにNMを求めた。

SN=0は，5個のBGSでない遺伝子の組（B0）である。SNが1から130はBGSである。BGS列は，手作業で決めたBGS名であるが，分析作業で同じものを2度間違って分析したため連番になっていない。そのため，BGSにSNを添え字に付けたものが130個のBGSになる。Gene列は各BGSに含まれる遺伝子数である。LogiとQDFとLDF2とLDF1列は，ロジスティック回帰とQDFとFisherのLDFのNMすなわち誤分類数である。LDF2は事前確率をケース数の22:40に比例させたものであり，LDF1は1:1にしたものである。（LDF2-QDF）と（LDF2-LDF1）は各NMの差である。前者の値はすべて正で，QDFのNMがLDF2より少ないことを示す。後者の差が0にならないのは，事前確率の設定で大きくNMが異なるためである[11]。ロジスティック回帰のNMが全て0であるので，130個のBGSはLSDであることが再確認できた。これに反してFisherのLDFのNMは全て0ではない。新村（2016b）では，0にならないことを逆手にとってSMの抽出順に癌診断の優先度が認められると予見したが，BGSの選択順は手作業で恣意的であるので，その議論はできない。

代わって得られた改定IP-OLDFの定数項をc列に示し，2つのSupport Vector（SV）の距離が2に固定されているので“200/c”を計算した値が大きいほど，2群を明確に分けると考えて他の変数と比較検討したが有効な情報は得られなかったので表から省く。この値がB1では

[10] BGSとは，それに含まれる遺伝子の一つを取り除くと，もうMNM=0にならない最小の部分空間である。
[11] 事前確率をケース数に比例させた方が，他のSVMの研究などとNMを統一して比較できる。しかし，2群のケース数がどうであれ，そこから計算された正規分布でFisherのLDFが定義されているので，事前確率を1:1にしたものが，多くの統計ソフトのデフォルトであることに注意がいるがデフォルトを変えるべきであろう。

0.01であり，B90では2.27である。恐らくB90の遺伝子でできる判別関数は，正常と癌の判別が容易であることを示すかもしれない。B21は定数項が0であり，この値が計算できない。表の最後の4行に合計，最大値，平均値，最小値が求めてある。Cの最大値は861,273で最小値は-824,980と異常に大きい。SV間の距離が2であることを考えれば異常に大きく，他のデータでは恐らく見られない現象であろう。“C”に代わってSRange列は，改定IP-OLDFの判別スコアの範囲の大きさである。“RatioSV=200/SR”は200をSRangeで割ったものであり，tは11月27日に求めた130個の改定IP-OLDFの判別スコアのt値である。SRangeの最大値と最小値は，300470と222であり，RatioSVは0.901と0.001である。すなわち判別スコアの範囲に対して，SV間の距離は僅か0.001％から0.901％と狭い幅であることが分かる。またt値の範囲は，［1.2, 9.18］である。表5で130組のBGSに含まれる遺伝子のt値の最大値と最小値を求め，2000個の遺伝子のt値の範囲は［2.49, 7.93］である。個々の遺伝子のt値の範囲は，判別スコアのt値の範囲に含まれる。判別スコアのt値は，［7.93, 9.18］と大きなものがある半面，［1.2,2.49］と個々の遺伝子のt値より小さいものもある。これらの指標が医学的に何を表すかは，専門医に検討してもらう必要がある。C列以降は4章で詳しく検討する。

表1　130個のBGSとBGS0

SN	BGS	Gene	Logi	QDF	LDF2-QDF	LDF2	LDF1	LDF2-LDF1	C	SRange	200/SR	t1127
0	B0	5	8	10	1	11	11	0	-32433	~~23762~~	~~0.008~~	~~7.41~~
1	B1	20	0	0	9	9	9	0	16568	395	0.507	6.96
2	B2	13	0	3	0	3	3	0	-537	1303	0.144	7.27
3	B3	17	0	0	6	6	6	0	362	666	0.3	7.19
4	B4	14	0	0	9	9	9	0	-1038	~~7675~~	~~0.026~~	~~7.52~~
5	B5	16	0	0	7	7	9	-2	4852	2888	0.069	6.91
6	B6	18	0	0	7	7	7	0	-1798	~~14400~~	~~0.014~~	~~6.87~~
7	B7	15	0	0	5	5	8	-3	18376	1415	0.141	8.74
8	B8	13	0	2	4	6	6	0	10421	592	0.338	7.83
9	B9	16	0	1	8	9	8	1	922	1159	0.173	6.81
10	B10	10	0	1	3	4	7	-3	272	3171	0.063	8.2
11	B11	12	0	0	4	4	6	-2	9574	~~7926~~	~~0.025~~	~~6.82~~
12	B12	19	0	0	7	7	6	1	693	~~20008~~	~~0.01~~	~~7.51~~
13	B13	16	0	0	6	6	6	0	57689	~~5351~~	~~0.037~~	~~8.71~~
14	B14	9	0	4	2	6	7	-1	9759	~~13706~~	~~0.015~~	~~2.24~~
15	B15	12	0	4	3	7	8	-1	7330	~~89578~~	~~0.002~~	~~6.2~~
16	B16	19	0	0	9	9	9	0	333300	1535	0.13	8.18
17	B17	13	0	1	8	9	7	2	4164	~~6718~~	~~0.03~~	~~6.59~~
18	B18	18	0	0	5	5	6	-1	13450	3434	0.058	3.88

19	B19	12	0	3	3	6	5	1	956	~~7087~~	~~0.028~~	~~8.2~~
20	B20	15	0	1	6	7	9	-2	3433	~~5080~~	~~0.039~~	~~2.49~~
21	B21	13	0	3	5	8	8	0	0	~~20746~~	~~0.01~~	~~5.93~~
22	B22	19	0	0	8	8	9	-1	-23269	~~42356~~	~~0.005~~	~~3~~
23	B23	14	0	2	5	7	6	1	708	~~84975~~	~~0.002~~	~~6.89~~
24	B24	18	0	0	6	6	6	0	-824980	3106	0.064	6.33
25	B25	12	0	1	7	8	5	3	-8920	~~4529~~	~~0.044~~	~~8.15~~
26	B26	13	0	1	4	5	4	1	1811	2076	0.096	5.22
27	B27	10	0	2	2	4	4	0	1550	~~9542~~	~~0.021~~	~~7.84~~
28	B28	16	0	2	3	5	6	-1	-98657	652	0.307	5.1
29	B29	16	0	1	8	9	4	5	-362	~~10260~~	~~0.019~~	~~3.37~~
30	B30	15	0	2	4	6	6	0	12022	~~61569~~	~~0.003~~	~~6.49~~
31	B31	21	0	0	6	6	8	-2	-113641	~~16959~~	~~0.012~~	~~6.17~~
32	B32	15	0	0	7	7	10	-3	-3482	~~82982~~	~~0.002~~	~~6.2~~
33	B33	18	0	1	5	6	11	-5	-80354	404	0.495	6.5
34	B34	15	0	2	4	6	7	-1	-1147	2711	0.074	4.04
35	B35	11	0	2	4	6	7	-1	-1147	~~5806~~	~~0.034~~	~~8.3~~
36	B36	11	0	2	3	5	6	-1	10249	2599	0.077	6.58
37	B37	20	0	0	7	7	8	-1	-9450	~~9028~~	~~0.022~~	~~2.96~~
38	B39	14	0	0	6	6	9	-3	1864	~~22011~~	~~0.009~~	~~7.28~~
39	B40	18	0	1	6	7	7	0	22880	~~7182~~	~~0.028~~	~~8.65~~
40	B41	12	0	2	2	4	5	-1	-34524	~~11647~~	~~0.017~~	~~7.39~~
41	B42	15	0	0	6	6	5	1	-19325	2969	0.067	6.18
42	B43	16	0	1	5	6	4	2	-3292	~~6563~~	~~0.03~~	~~8.05~~
43	B51	14	0	1	6	7	6	1	3713	1706	0.117	7.94
44	B52	13	0	1	9	10	12	-2	-7278	~~15193~~	~~0.013~~	~~4.49~~
45	B53	21	0	0	8	8	6	2	-21304	~~42535~~	~~0.005~~	~~6.41~~
46	B54	13	0	0	5	5	7	-2	25164	1044	0.192	6.36
47	B55	16	0	2	8	10	8	2	-1134	2377	0.084	6.66
48	B56	15	0	1	3	4	7	-3	1205	~~16621~~	~~0.012~~	~~6.38~~
49	B57	21	0	0	7	7	8	-1	-68788	~~6700~~	~~0.03~~	~~6.12~~
50	B58	9	0	4	1	5	7	-2	4001	~~72060~~	~~0.003~~	~~7.25~~
51	B59	25	0	0	9	9	9	0	-529699	~~4611~~	~~0.043~~	~~8.67~~
52	B60	12	0	0	5	5	5	0	14961	505	0.396	8.03
53	B61	14	0	2	5	7	8	-1	-2453	3359	0.06	4.24
54	B62	12	0	2	1	3	3	0	4146	~~11231~~	~~0.018~~	~~7.45~~
55	B63	16	0	1	8	9	11	-2	56028	1657	0.121	8.3
56	B64	9	0	2	6	8	7	1	-945	295	0.679	7.31
57	B65	18	0	0	7	7	11	-4	-1092	2714	0.074	3.97
58	B66	16	0	0	8	8	8	0	2321	~~10722~~	~~0.019~~	~~6.73~~

59	B67	15	0	0	10	10	9	1	51644	~~7526~~	~~0.027~~	~~6.17~~
60	B68	16	0	1	7	8	6	2	-3323	2645	0.076	7.88
61	B69	14	0	0	8	8	7	1	-906	~~12589~~	~~0.016~~	~~5.97~~
62	B70	19	0	0	11	11	8	3	12437	~~11921~~	~~0.017~~	~~7.63~~
63	B71	20	0	0	7	7	8	-1	62169	~~24301~~	~~0.008~~	~~1.2~~
64	B72	12	0	4	3	7	8	-1	-1491	~~300470~~	~~0.001~~	~~7.18~~
65	B73	17	0	1	8	9	8	1	278310	~~14167~~	~~0.014~~	~~5.42~~
66	B74	21	0	0	3	3	5	-2	-39553	615	0.325	7.49
67	B75	15	0	3	3	6	6	0	2651	423	0.473	6.94
68	B76	13	0	0	8	8	10	-2	1215	1351	0.148	6.7
69	B77	11	0	1	5	6	6	0	-1783	~~5685~~	~~0.035~~	~~7.35~~
70	B78	16	0	0	5	5	7	-2	-24894	3618	0.055	5.82
71	B79	16	0	0	8	8	7	1	-6594	~~4667~~	~~0.043~~	~~7.73~~
72	B80	18	0	0	4	4	8	-4	-11983	~~8442~~	~~0.024~~	~~7.2~~
73	B81	16	0	0	6	6	6	0	-361	734	0.273	7.35
74	B82	16	0	0	7	7	7	0	-2103	2727	0.073	7.87
75	B83	15	0	1	7	8	9	-1	12042	~~5134~~	~~0.039~~	~~8.08~~
76	B84	18	0	0	5	5	7	-2	14480	619	0.323	8.06
77	B85	15	0	2	7	9	10	-1	936	1379	0.145	7.06
78	B86	15	0	0	7	7	10	-3	-1950	711	0.281	6.58
79	B87	16	0	3	6	9	10	-1	481	1255	0.159	6.66
80	B88	18	0	0	7	7	8	-1	-2452	~~29656~~	~~0.007~~	~~7.37~~
81	B89	19	0	0	6	6	7	-1	-67777	1330	0.15	9.18
82	B90	11	0	3	2	5	9	-4	88	~~45623~~	~~0.004~~	~~1.46~~
83	B92	18	0	0	11	11	9	2	861274	~~176727~~	~~0.001~~	~~7.54~~
84	B93	19	0	0	5	5	6	-1	-20023	~~27342~~	~~0.007~~	~~6.58~~
85	B94	12	0	0	7	7	5	2	925	508	0.394	5.8
86	B95	9	0	7	-1	6	7	-1	1721	1407	0.142	8.9
87	B96	17	0	0	7	7	8	-1	-799	~~17146~~	~~0.012~~	~~1.9~~
88	B97	13	0	0	7	7	7	0	58216	~~72420~~	~~0.003~~	~~8.06~~
89	B98	18	0	0	5	5	5	0	800	~~4497~~	~~0.044~~	~~2.71~~
90	B99	17	0	0	13	13	11	2	-1177	~~5902~~	~~0.034~~	~~5.85~~
91	B100	12	0	1	6	7	9	-2	970	1659	0.121	6.48
92	B101	17	0	2	7	9	10	-1	6621	~~5994~~	~~0.033~~	~~6.59~~
93	B102	12	0	3	2	5	7	-2	-252	258	0.774	6.32
94	B103	11	0	3	2	5	5	0	-1376	~~10969~~	~~0.018~~	~~1.96~~
95	B104	15	0	1	6	7	8	-1	12278	~~35391~~	~~0.006~~	~~6.93~~
96	B105	15	0	2	5	7	8	-1	-2037	2059	0.097	7.6
97	B106	19	0	0	6	6	6	0	3645	2181	0.092	7.81
98	B107	14	0	0	6	6	5	1	3988	~~17637~~	~~0.011~~	~~2.19~~

99	B108	17	0	0	5	5	4	1	436472	~~60450~~	~~0.003~~	~~7.82~~
100	B109	16	0	2	3	5	5	0	21	869	0.23	4.72
101	B110	19	0	1	6	7	7	0	-5398	3691	0.054	6.2
102	B111	17	0	0	6	6	8	-2	-674	1405	0.142	7.05
103	B112	11	0	2	2	4	5	-1	183	1704	0.117	5.72
104	B113	16	0	1	8	9	9	0	-10924	~~16946~~	~~0.012~~	~~3.55~~
105	B114	16	0	1	5	6	8	-2	-23370	~~26419~~	~~0.008~~	~~7.46~~
106	B115	17	0	0	6	6	10	-4	205	~~4940~~	~~0.04~~	~~4.98~~
107	B116	18	0	0	5	5	10	-5	39694	~~17345~~	~~0.012~~	~~6.85~~
108	B117	18	0	1	5	6	5	1	2558	2525	0.079	8.1
109	B118	16	0	1	5	6	7	-1	-7071	~~12209~~	~~0.016~~	~~6.8~~
110	B119	14	0	0	5	5	6	-1	1263	3842	0.052	3.9
111	B120	17	0	1	7	8	8	0	-5739	~~5570~~	~~0.036~~	~~7.22~~
112	B121	16	0	0	11	11	10	1	3902	3110	0.064	7.53
113	B122	21	0	0	8	8	8	0	-8361	~~19071~~	~~0.01~~	~~3.04~~
114	B123	17	0	0	8	8	9	-1	-5482	~~39965~~	~~0.005~~	~~6.99~~
115	B124	19	0	0	7	7	6	1	-85	1702	0.118	5.68
116	B125	16	0	0	4	4	5	-1	885	~~4200~~	~~0.048~~	~~6.81~~
117	B126	14	0	1	5	6	5	1	-1400	1681	0.119	7.42
118	B127	13	0	3	5	8	10	-2	-1203	552	0.363	6.48
119	B128	17	0	1	7	8	10	-2	900	1934	0.103	7.35
120	B129	17	0	0	7	7	7	0	430	712	0.281	6.99
121	B130	16	0	0	6	6	10	-4	-1504	~~9491~~	~~0.021~~	~~2.31~~
122	B131	17	0	1	7	8	8	0	-65889	~~19968~~	~~0.01~~	~~6.57~~
123	B132	13	0	3	7	10	7	3	1836	~~5506~~	~~0.036~~	~~6.69~~
124	B133	15	0	1	6	7	8	-1	-6915	~~5950~~	~~0.034~~	~~6.67~~
125	B134	15	0	2	3	5	9	-4	-6372	~~8825~~	~~0.023~~	~~6.98~~
126	B135	15	0	0	8	8	8	0	5420	~~29652~~	~~0.007~~	~~2.11~~
127	B136	14	0	2	2	4	14	-10	-190925	~~113129~~	~~0.002~~	~~8.13~~
128	B137	11	0	0	7	7	7	0	143	222	0.901	8.49
129	B138	12	0	3	3	6	5	1	465	380	0.526	8.38
130	B139	12	0	5	4	9	8	1	-287	401	0.499	7.38
	Total	2000	0	141	748	889	968	-79				
	Max	25	8	10	13	13	14	5	861274	300470	0.901	9.18
	Mean	15.27	0.06	1.08	5.71	6.79	7.39	-0.60	1115.10	16152	0.110	6.42
	MIN	5	0	0	-1	3	3	-10	-824980	222	0.001	1.20

3.2 NMの分析

図1はB0を省いた130個のBGSに含まれるGeneの数と，QDFとLDF2のNMの分布である。130個のBGSに9個から25個の遺伝子が含まれている。正常者は22人，悪性の腫瘍患者が40

人の計62人で最大25個の変数の分析は，130組の分析の繰り返しになるが統計手法にとって容易である。QDFのNMの範囲は［0, 7］であるが，0が63，1から7までの度数は30，20，11，4，1，1である。一方，LDF2のNMの範囲は［3, 13］で誤分類確率は［5%, 21%］である。成績の合否判定でもこれぐらいの悪い誤分類確率であることを示したが，多くの研究者はあまり興味を示さなかった。筆者としては，判別超平面上に多くのケースのある医学診断と同じ構造を持つデータとして合否判定をLSD判別研究に用いた（新村, 1984）。しかし，医学論文誌の中には結果を第三者が検証できるよう公開を義務づけている論文誌もあるので，信頼性の高い遺伝子情報でも「分散共分散行列に基づく判別関数の誤分類確率が高い」ことを示せた。即ち，3種のOLDFでLSDであることが簡単に分ったが，FisherのLDFはLSDであることが分からない。QDFは130個の5割弱の2群がLSDなことが分かる。恐らく，正常を癌群が包み込むような状態のためQDFでNM=0になるものがあるが，MNM=0であるので包み込む範囲は狭いと考えられる。

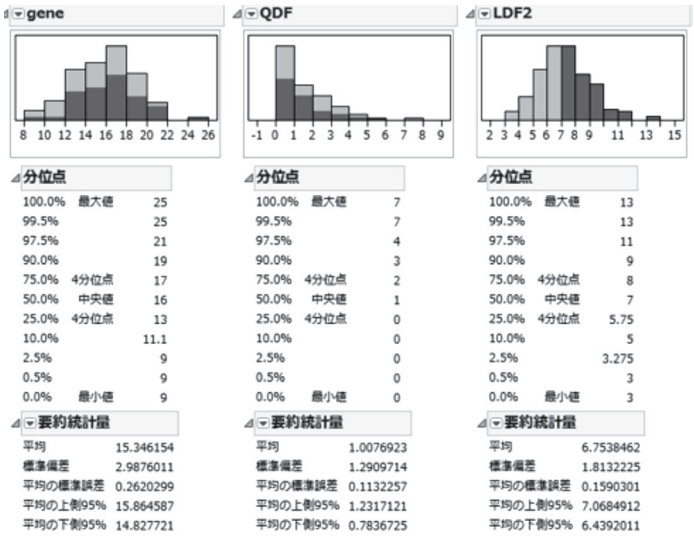


図1　Geneの数とQDFとLDF2のNMの分布

図2は，GENEとQDFとLDF2の行列散布図である。遺伝子数が増えると，QDFのNMが減少し，LDF2のNMが増加する傾向があり，相関係数は-0.5996と0.2291である。QDFは遺伝子数pが増えると実際に用いられる説明変数は（$p^2/2 + 1.5p$）個になるので，MNMは少なくなると考えられるがその影響は良く分らない。LDFは，遺伝子数の多いBGSのNMが大きくなるのは，遺伝子数が少ないBGSの方が一般的に癌の悪性度が強いのかもしれない。一方，QDFとLDF2の相関は-0.0952と無相関に近い。この他，RDAやLASSOのNMも調べたいが，

作業過多になるため，これらの手法の有効性を信じる研究者が実証研究すべきであろう[12]。筆者は，2群は正規分布でないと考えられるのでLSDの判別はできないと考えている。

	gene	QDF	LDF2
gene	1.0000	-0.5996	0.2291
QDF	-0.5996	1.0000	-0.0952
LDF2	0.2291	-0.0952	1.0000

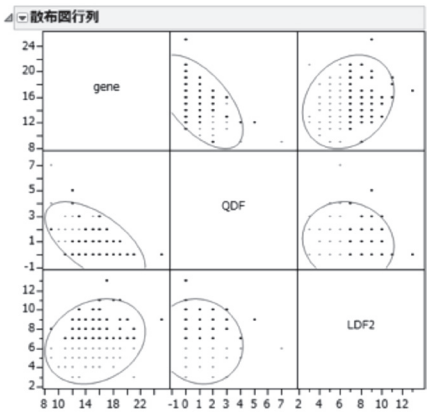


図2 GENE,QDFとLDF2の行列散布図

3.3 LDFとQDFの正準相関図の検討

BGS1のLDFの誤分類数は9でQDFは0である。図3のようにLDFとQDFの判別で正準プロットが表示される。上図がLDFで，下図がQDFである。これの解釈に不慣れなので，一応正常を表す右に付置する□の境界に線を引き，左に布置する癌を表す×の重なりを調べたが，両方とも重なりが大きい。すなわち，統計はばらつきの大きなことが情報を表すと考えているので，LSDか否かの情報を検出することは期待できない。11月18日にJMPのユーザー会でSAS社の創業者の一人でJMPの開発者であるSall博士（新村訳著（Sall著），1986 ; 2012）を捕まえ，癌診断に数理計画法LINGOの結果をJMPで分析し，この分野の多くの研究者に標準的な解析法を提示したい意思を伝えて，この図を含む本稿の主要な結果を渡した。何か進展があるかもしれないことを期待している。またPCAの各主成分軸上で，2群のt値を計算するオプションを付け加えることを要望した。

[12] これまでの実証研究は，数個のモデルで検証し，全てのモデルで考えられる全てのLDFで検証していないことが大きな問題である。このため，筆者のように多くの問題点が指摘できなかった。

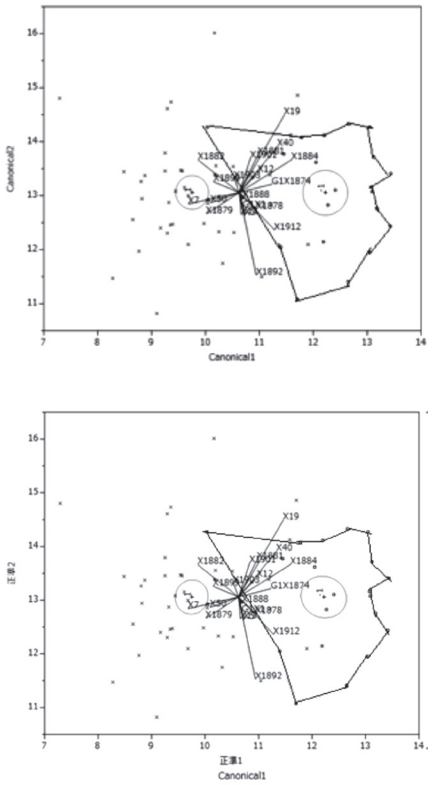


図3　BGS1のLDF（上）とQDF（下）の正準プロット

4. 改定IP-OLDFの結果をJMPで検証

4.1　改定IP-OLDFの判別スコアの検証

表2は，130個の改定IP-OLDFの判別スコアの分析結果である。最初のMinとMaxは正常の範囲で-1以下にあり，次のMINとMAXは癌で1以上にある。このことは，130個のBGSがLSDであることを示す。表1のC列は改定IP-OLDFの定数項である。SV間の距離は2に固定されているので2を定数項で割った逆数を遺伝子数やLDFとQDFの誤分類数と比較したが良い情報は得られなかった。それに代わって，判別スコアの範囲（SRange）とSV間の距離の範囲に対する比RatioSV（＝200/SR，％表示）と2群の判別スコアのt値である。「t異」値は，分散が等しいという仮定と等しくないという仮定で結果が異なるが，一般的には「異なるを選ぶべき」と考えている。R1とR2はRatioSVとt値による順位である。RatioSVはMNMと同じく筆者が導入した統計量であるが，SVを取り入れたMP判別関数の評価に有用な役割を果たすと考えている。主成分1は4.3で行った判別スコアを転置して，130個の判別スコアをケ

ースにしてPCAを行った第1主成分である。R3が1から23位までが第1主成分が正で，残りの106個の判別スコアが負の値である。正になる23個は癌を識別し，106個は正常を識別しているようだ。「Omit」列は，RatioSVが0.05以下を0に，それ以上を1で示す。

表2 130個の改定IP-OLDFの判別スコア

ID	Min	Max	MIN	MAX	Srange	RatioSV	R1	t異	R2	主成分1	R3	Omit
1	-9481	-1	1	14281	23762	0.008	111	7.41	41	1.587	17	0
2	-149	-1	1	246	395	0.507	5	6.96	60	-2.132	123	1
3	-720	-1	1	672	1393	0.144	27	7.27	49	-1.962	87	1
4	-257	-1	1	410	666	0.300	16	7.19	53	-2.082	111	1
5	-1993	-1	1	5682	7675	0.026	81	7.52	35	-1.070	42	0
6	-543	-1	1	2345	2888	0.069	49	6.91	63	-1.893	78	1
7	-6610	-1	1	7790	14400	0.014	98	6.87	65	-0.048	25	0
8	-511	-1	1	904	1415	0.141	30	8.74	3	-1.875	74	1
9	-221	-1	1	371	592	0.338	12	7.83	26	-2.072	106	1
10	-253	-1	1	907	1159	0.173	22	6.81	68	-2.070	105	1
11	-1283	-1	1	1889	3171	0.063	53	8.20	11	-1.572	58	1
12	-2408	-1	1	5518	7926	0.025	82	6.82	67	-0.975	38	0
13	-8280	-1	1	11728	20008	0.010	108	7.51	36	1.141	20	0
14	-2455	-1	1	2897	5351	0.037	67	8.71	4	-1.205	46	0
15	-398	-1	1	13308	13706	0.015	96	2.24	125	-2.014	93	0
16	-42194	-1	1	47383	89578	0.002	127	6.20	92	8.590	8	0
17	-685	-1	1	851	1535	0.130	31	8.18	13	-1.910	81	1
18	-2611	-1	1	4107	6718	0.030	77	6.59	77	-1.377	52	0
19	-384	-1	1	3050	3434	0.058	55	3.88	116	-2.050	101	1
20	-2483	-1	1	4604	7087	0.028	78	8.20	12	-0.906	37	0
21	-241	-1	1	4839	5080	0.039	65	2.49	123	-2.100	116	0
22	-6303	-1	1	14443	20746	0.010	109	5.93	100	0.011	24	0
23	-3252	-1	1	39104	42356	0.005	119	3.00	120	-1.037	40	0
24	-46017	-1	1	38958	84975	0.002	126	6.89	64	11.657	4	0
25	-762	-1	1	2344	3106	0.064	51	6.33	90	-1.895	79	1
26	-1563	-1	1	2966	4529	0.044	61	8.15	14	-1.488	57	0
27	-312	-1	1	1764	2076	0.096	40	5.22	107	-2.066	103	1
28	-3717	-1	1	5825	9542	0.021	87	7.84	25	-0.174	26	0
29	-114	-1	1	538	652	0.307	15	5.10	108	-2.136	127	1
30	-793	-1	1	9467	10260	0.019	88	3.37	118	-1.852	71	0
31	-31915	-1	1	29654	61569	0.003	122	6.49	84	6.369	10	0
32	-4374	-1	1	12585	16959	0.012	102	6.17	96	-0.737	33	0
33	-35120	-1	1	47862	82982	0.002	125	6.20	94	7.170	9	0

34	-158	-1	1	246	404	0.495	7	6.50	83	-2.134	125	1
35	-165	-1	1	2546	2711	0.074	46	4.04	113	-2.061	102	1
36	-1813	-1	1	3993	5806	0.034	71	8.30	9	-1.121	43	0
37	-1266	-1	1	1332	2599	0.077	44	6.58	81	-1.860	73	1
38	-493	-1	1	8535	9028	0.022	85	2.96	121	-1.974	88	0
39	-8323	-1	1	13689	22011	0.009	110	7.28	48	1.586	18	0
40	-3294	-1	1	3889	7182	0.028	79	8.65	6	-0.745	34	0
41	-5849	-1	1	5799	11647	0.017	92	7.39	42	-0.331	29	0
42	-634	-1	1	2336	2969	0.067	50	6.18	95	-1.944	85	1
43	-2302	-1	1	4262	6563	0.030	75	8.05	20	-1.010	39	0
44	-637	-1	1	1069	1706	0.117	37	7.94	22	-1.878	76	1
45	-2248	-1	1	12946	15193	0.013	99	4.49	111	-1.434	55	0
46	-23573	-1	1	18962	42535	0.005	120	6.41	87	3.281	12	0
47	-250	-1	1	794	1044	0.192	21	6.36	89	-2.078	110	1
48	-640	-1	1	1737	2377	0.084	42	6.66	75	-1.937	84	1
49	-7355	-1	1	9266	16621	0.012	100	6.38	88	0.147	22	0
50	-1103	-1	1	5597	6700	0.030	76	6.12	98	-1.697	61	0
51	-26694	-1	1	45366	72060	0.003	123	7.25	50	9.483	6	0
52	-2107	-1	1	2503	4611	0.043	62	8.67	5	-1.187	45	0
53	-254	-1	1	251	505	0.396	9	8.03	21	-2.106	119	1
54	-244	-1	1	3115	3359	0.060	54	4.24	112	-2.036	97	1
55	-4228	-1	1	7003	11231	0.018	91	7.45	39	-0.286	28	0
56	-547	-1	1	1110	1657	0.121	32	8.30	10	-1.876	75	1
57	-84	-1	1	211	295	0.679	3	7.31	47	-2.126	122	1
58	-182	-1	1	2532	2714	0.074	47	3.97	114	-2.073	107	1
59	-4498	-1	1	6223	10722	0.019	89	6.73	71	-0.885	36	0
60	-3852	-1	1	3674	7526	0.027	80	6.17	97	-1.050	41	0
61	-862	-1	1	1783	2645	0.076	45	7.88	23	-1.769	66	1
62	-7323	-1	1	5266	12589	0.016	95	5.97	99	-0.691	31	0
63	-3822	-1	1	8099	11921	0.017	93	7.63	31	0.090	23	0
64	-188	-1	1	24113	24301	0.008	112	1.20	130	-2.140	128	0
65	-118569	-1	1	181901	300470	0.001	130	7.18	54	42.434	1	0
66	-9210	-1	1	4957	14167	0.014	97	5.42	106	-0.445	30	0
67	-251	-1	1	364	615	0.325	13	7.49	37	-2.084	112	1
68	-117	-1	1	306	423	0.473	8	6.94	61	-2.123	121	1
69	-524	-1	1	827	1351	0.148	25	6.70	72	-2.048	100	1
70	-2202	-1	1	3483	5685	0.035	70	7.35	46	-1.278	47	0
71	-692	-1	1	2925	3618	0.055	56	5.82	102	-1.931	83	1
72	-1615	-1	1	3052	4667	0.043	63	7.73	29	-1.359	51	0
73	-4103	-1	1	4339	8442	0.024	83	7.20	52	-0.826	35	0

74	-431	-1	1	302	734	0.273	19	7.35	44	-2.066	104	1
75	-910	-1	1	1818	2727	0.073	48	7.87	24	-1.700	62	1
76	-1981	-1	1	3153	5134	0.039	66	8.08	17	-1.129	44	0
77	-210	-1	1	409	619	0.323	14	8.06	19	-2.073	108	1
78	-648	-1	1	731	1379	0.145	26	7.06	55	-1.983	90	1
79	-278	-1	1	433	711	0.281	17	6.58	79	-2.089	113	1
80	-658	-1	1	596	1255	0.159	23	6.66	76	-2.003	92	1
81	-13596	-1	1	16060	29656	0.007	116	7.37	43	2.785	13	0
82	-592	-1	1	738	1330	0.150	24	9.18	1	-1.884	77	1
83	-72581	-1	1	104146	176727	0.001	129	7.54	33	27.189	2	0
84	-12618	-1	1	14724	27342	0.007	114	6.58	80	1.618	16	0
85	-96	-1	1	412	508	0.394	10	5.80	103	-2.143	129	1
86	-588	-1	1	819	1407	0.142	29	8.90	2	-1.904	80	1
87	-817	-1	1	16329	17146	0.012	103	1.90	129	-1.961	86	0
88	-32813	-1	1	39607	72420	0.003	124	8.06	18	10.589	5	0
89	-207	-1	1	4290	4497	0.044	60	2.71	122	-2.090	114	0
90	-3099	-1	1	2802	5902	0.034	72	5.85	101	-1.343	49	0
91	-241	-1	1	1417	1659	0.121	33	6.48	85	-2.043	98	1
92	-1946	-1	1	4048	5994	0.033	74	6.59	78	-1.404	54	0
93	-69	-1	1	189	258	0.774	2	6.32	91	-2.152	130	1
94	-224	-1	1	10745	10969	0.018	90	1.96	128	-2.093	115	0
95	-14020	-1	1	21372	35391	0.006	117	6.93	62	2.103	15	0
96	-818	-1	1	1241	2059	0.097	39	7.60	32	-1.835	70	1
97	-1019	-1	1	1162	2181	0.092	41	7.81	28	-1.808	68	1
98	-636	-1	1	17001	17637	0.011	105	2.19	126	-1.976	89	0
99	-21724	-1	1	38726	60450	0.003	121	7.82	27	9.022	7	0
100	-74	-1	1	795	869	0.230	20	4.72	110	-2.133	124	1
101	-1728	-1	1	1963	3691	0.054	57	6.20	93	-1.751	65	1
102	-694	-1	1	711	1405	0.142	28	7.05	56	-1.991	91	1
103	-374	-1	1	1330	1704	0.117	36	5.72	104	-2.032	95	1
104	-1653	-1	1	15293	16946	0.012	101	3.55	117	-1.676	60	0
105	-10543	-1	1	15876	26419	0.008	113	7.46	38	2.485	14	0
106	-950	-1	1	3990	4940	0.040	64	4.98	109	-1.911	82	0
107	-8171	-1	1	9175	17345	0.012	104	6.85	66	1.259	19	0
108	-797	-1	1	1728	2525	0.079	43	8.10	16	-1.727	64	1
109	-6396	-1	1	5813	12209	0.016	94	6.80	70	-0.177	27	0
110	-440	-1	1	3401	3842	0.052	58	3.90	115	-2.034	96	1
111	-1461	-1	1	4110	5570	0.036	69	7.22	51	-1.279	48	0
112	-1154	-1	1	1956	3110	0.064	52	7.53	34	-1.588	59	1
113	-1329	-1	1	17742	19071	0.010	106	3.04	119	-1.709	63	0

114	-15244	-1	1	24721	39965	0.005	118	6.99	57	4.461	11	0
115	-575	-1	1	1126	1702	0.118	35	5.68	105	-2.047	99	1
116	-2346	-1	1	1854	4200	0.048	59	6.81	69	-1.404	53	0
117	-699	-1	1	982	1681	0.119	34	7.42	40	-1.828	69	1
118	-331	-1	1	221	552	0.363	11	6.48	86	-2.105	118	1
119	-596	-1	1	1338	1934	0.103	38	7.35	45	-1.797	67	1
120	-243	-1	1	468	712	0.281	18	6.99	58	-2.077	109	1
121	-261	-1	1	9230	9491	0.021	86	2.31	124	-2.023	94	0
122	-9827	-1	1	10141	19968	0.010	107	6.57	82	0.788	21	0
123	-2961	-1	1	2546	5506	0.036	68	6.69	73	-1.347	50	0
124	-3035	-1	1	2915	5950	0.034	73	6.67	74	-1.456	56	0
125	-3727	-1	1	5098	8825	0.023	84	6.98	59	-0.714	32	0
126	-867	-1	1	28785	29652	0.007	115	2.11	127	-1.854	72	0
127	-46881	-1	1	66247	113129	0.002	128	8.13	15	20.068	3	0
128	-74	-1	1	148	222	0.901	1	8.49	7	-2.135	126	1
129	-167	-1	1	213	380	0.526	4	8.38	8	-2.105	117	1
130	-151	-1	1	249	401	0.499	6	7.70	30	-2.115	120	1
Max	-69	-1.0	1.0	181901	300470	0.901	130.0	9.18	130.0	42.434	SUM	58
Mean	-5979	-1.0	1.0	9947	15926	0.107	65.5	6.46	65.5	0		
Min	-118569	-1.0	1.0	148	222	0.001	1.0	1.20	1.0	-2.152		
Range	5910	0.0	0.0	171954	284544	0.794	64.5	2.72	64.5	42.434		

最後の4行は, 10項目の最大値, 平均値, 最小値, 範囲を示す。BGS65の2つの範囲は[-118569, -1]と［1, 181901］で, Minの最小値であり, MAXの最大値でもある。全体の範囲は［-118569, 181901］である。RatioSVは，SV間の距離2はこの範囲の何%に相当するかで，判別の容易さと信頼性が評価できると考えている。BGS65は0.001%で130個の中で一番小さいことがR1で分かる。しかしt値は7.18で，R2では54位で大きな値をとっている。主成分1は図7で説明するが，BGS65, 83, 127はR3で第1, 2, 3順位であり，いずれもOmit=0である。SRangeと，RatioSVの範囲は, [222, 300470]と[1.20, 9.18]である。主成分1の判別スコアの範囲は[-2.152, 42.434］であり，断定はできないが癌に対応する23個の少ない判別関数も値が正でばらついているようだ。

図4は，4変数の相関係数と行列散布図である。大きな相関係数は，SRangeとRatioSVが-0.2676, tとRatioSVが0.2113である。行列散布図は，横軸に各変数の異なったスケールが表示されている。2行3列のSRangeとRatioSVの散布図が，ほぼ直角に布置している。RatioSVが小さいものでRANGEが大きく異なっているものと，RANGEの小さなものでRatioSVが大きく異なっているものに分裂しているが，医学的な意味は分からない。このような規則性を持つ特徴は，何か重要な意味を持っているかと思われる。SRange列とRatioSV列を見ると，y

軸方向にこの傾向が認められる。一方, SRange行を見ると, x軸方向にこの傾向が認められる。×印は，図5で触れるがRatioSVが0.05以下のBGSを表していて，Omit=0の58個に対応している。

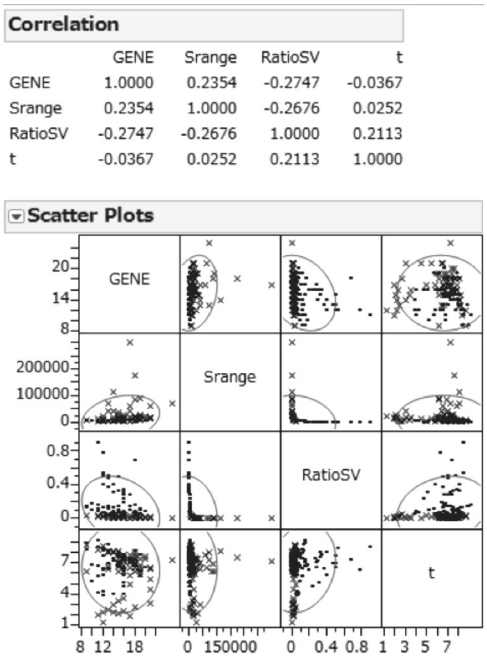


図4　遺伝子数と3変数の相関係数と行列散布図

図5は，図4に用いた4変数の基礎統計量である。“RatioSV≦0.05”の区間を選んで，濃いグリーンにすると他の変数の対応するケースも濃いグリーンに変わる。これらのケースに×の記号を付けて，図4に示した。SRangeの区間幅は1万に設定してあり，1万以下の区間の約3割と1万以上のほぼ全てが濃いグリーンになっている。これらが図4の2行3列のY軸方向に布置したBGSで，SV間の距離が判別スコアの0.05%以下であり，検証標本で判別した場合，信頼性が乏しいと判定される可能性が高い。この値が大きくなるほど，癌診断が確定的に行われることが期待されるが, その閾値をMethod1で検証する必要がある[13]。ただし図12ではこれを否定する結果になっている。R1はRatioSVの値が大きなものから降順に振ったランクである。表2で求めた最大のRangeの値をもつBGS65は130個の中で130位，即ちRatioSVの値

[13] 当初，遺伝子情報は計測値として信頼性が高く，癌も病気の中で一番明確に正常と区別できるので，130個のBGSによる改定IP-OLDFはMethod1で検証する必要がないと考えていた。

が一番小さいことを示す。SV間の距離が2に固定されていることに注意されていないが，判別スコアの範囲で正規化して評価すべきである。t値の濃い緑の割合は，明確な傾向は分からない。遺伝子数が10から20の間では，濃い緑は50%を超えている。

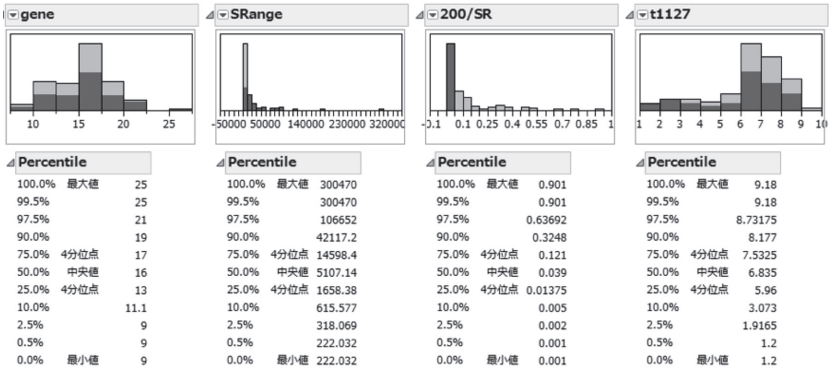


図5　SRangeRatioSVの基礎統計量

図6は，t値が最大値の9.18になるBGS82とt値が1.20で130番目[14]のBGS64の一元配置の分散分析である。正常群（-1）は値の小さい方に裾を引いた分布である。癌群（1）は25%と50%の区間幅が大きいが中央値はほぼ分布の真ん中にある。一方，BGS64のt値は1.20と小さく箱ひげ図が異常な表示になっているのは，癌患者の90%が173.56で，最大値が24113と異常に大きな値があるためである。即ち癌患者の判別スコアで異常に大きなものがあり，これがt値の考察を困難にしているようだ。これをどう対応するかは今のところ未検討である。

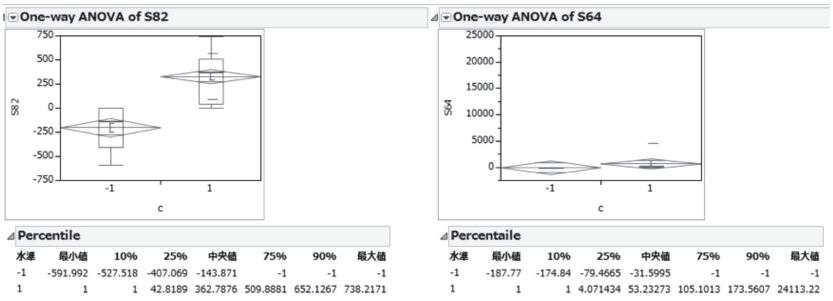


図6　t値が最大値をとるBGS81と130番目のBGS82の一元配置の分散分析

[14] BGS63のt値が1.2と一番小さい。

4.2 判別スコアのPCAによる検証

判別スコアを一元配置の分散分析で検討してt値が小さなものもあったが，全て2群が分かれていることを確認した。図7は，130個の判別スコアの主成分分析の結果である。因子負荷プロットから判別スコアは，1象限と4象限に布置し，スコアプロットは第1主成分軸におよそ-30度の傾きはあるが，ほぼ直線上に布置し第1主成分軸にほぼ並行している。ただし□印の正常群は2象限に，×印の癌群は3象限から4象限の布置しその長さは正常のほぼ2倍である。元のデータを用いた130個のBGSのPCA分析では，第1と2主成分のスコアプロットは2群の重なりが大きく，因子負荷プロットが4象限に広く布置していて，なぜ改定IP-OLDFが簡単に線形分離になるか理由付けできない。しかし130個の判別スコアのPCA分析は，スコアプロットがほぼ直線上に布置し，正常の第1主成分の値が表3から-4.52以下に，癌が-2.86以上に長く布置している。この直線に直交する判別境界を設定すれば，簡単に2群が分かれるが，判別超平面上では2群が小さな幅に重なってしまう。癌の方のばらつきが大きいのでQDFで分析すれば，正常群を2次曲線の内側に，癌群を外側に分けるようになり，LDF2のNMは0にならないのにQDFのNMの半数が0になることが説明できる。しかし，判別スコアの解釈を元のデータの分析結果のように解釈することは差し控えるべきかもしれない。

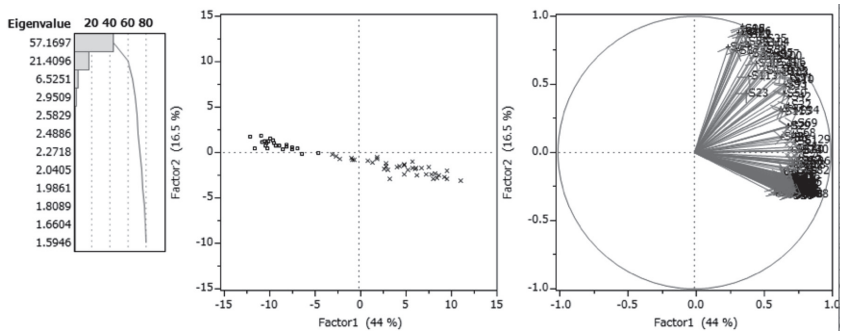


図7 130個の判別スコアの主成分分析の結果

図8は，横軸を第1主成分，縦軸を第2から5主成分にとったスコアプロットである。SN=61と62は図7で範囲外にあり表示されていない。この2個のBGSは他と異なっているようだ。99%の信頼区間は重なっているが，実際はLSDであることが目視で確認できる。そして当然であるが，正常群より癌群のばらつきが大きい。このため，LDFではNMが全て0でないが，QDFでは正常を2次判別境界が包むことで60個がNM=0であることも理解できる。

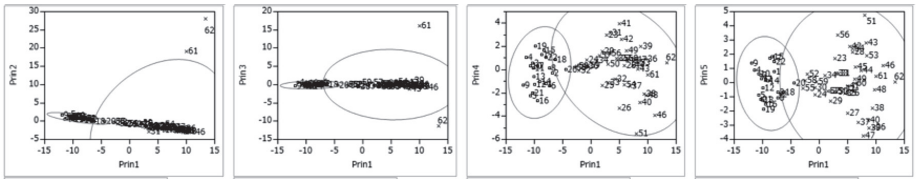


図8　横軸が第1主成分で縦軸は第2主成分から第5主成分にとったスコアプロット

表3は第1主成分の値を降順にランキングすると，癌群のR4の順位が1から34までが正で，35から40までは-2.86以上の負の値で，正常の22件は41から62に第1主成分の値が-4.52以下の値できれいに順位づけられる。癌は他の疾病と異なり人類最後の難病である。これが，遺伝子という究極の計測値で明確に峻別されるのは当然といえる。CTが開発され，訓練を受けた医師が読影すれば，統計の助けを借りる必要がなくなったという事実の認識が，数式を厳密に展開して新しい統計理論を開発するよりも優先されるべきであろう。

表3　第1主成分による癌と正常の判別

ID	主成分1	R4	主成分2	主成分3	主成分4	主成分5	主成分6
62	13.39	1	28.00	-11.31	0.62	0.07	-0.05
46	11.19	2	-3.11	-0.86	-3.92	1.21	-2.03
61	10.00	3	19.19	16.10	-0.42	0.46	-0.27
48	9.74	4	-2.89	-0.33	-2.24	-0.48	-3.14
36	9.67	5	-2.00	0.04	0.91	-3.17	-0.74
38	9.39	6	-2.72	-0.55	-2.06	-1.80	0.55
39	8.77	7	-2.53	1.15	2.01	-3.18	0.10
40	8.65	8	-2.80	-0.34	-2.81	-2.63	-0.59
53	8.37	9	-2.35	-0.13	0.63	1.89	-2.24
43	8.19	10	-2.94	-1.14	0.06	2.79	-0.95
51	8.03	11	-2.42	-0.18	-5.49	4.76	1.67
47	7.71	12	-1.62	0.57	0.86	-3.76	3.78
44	7.35	13	-2.40	-0.44	0.37	0.86	-1.29
37	6.89	14	-1.72	-0.35	-1.40	-2.81	-0.83
45	6.43	15	-2.54	-0.59	0.80	1.12	-0.92
49	6.31	16	-1.72	0.15	1.66	0.24	0.68
60	6.25	17	-1.84	-0.06	0.99	-0.04	2.18
54	5.86	18	-1.01	0.75	-1.28	2.40	-0.07
28	5.85	19	-1.71	0.13	0.37	2.14	0.42
42	5.42	20	-1.91	0.07	2.64	2.56	0.24
27	5.01	21	-1.42	-0.28	0.96	-2.13	4.48

41	5.00	22	-2.40	0.30	3.98	-0.24	-2.46
26	4.96	23	-1.30	0.14	-3.30	-0.74	0.66
35	4.51	24	-1.55	-0.62	0.98	-0.77	-1.55
32	4.22	25	-1.58	0.39	-0.80	-0.46	-0.14
31	3.44	26	-2.89	-0.17	3.15	0.67	2.12
56	3.19	27	-1.88	-0.16	1.38	3.35	1.69
33	3.01	28	-1.55	-0.65	-1.13	0.66	1.00
50	2.97	29	-1.24	0.14	0.49	-0.64	2.87
23	2.79	30	-1.76	0.14	2.97	1.70	-0.08
25	2.05	31	-0.69	-0.20	-1.34	-0.61	-0.75
29	2.01	32	-0.54	-0.13	1.58	-1.31	-0.08
57	1.49	33	-1.15	0.62	1.19	-0.60	0.15
34	1.01	34	-0.93	-0.48	0.78	0.50	0.22
59	-0.55	35	-0.82	0.58	0.37	0.07	0.17
30	-0.71	36	-0.83	-0.56	0.32	-0.34	-0.73
24	-0.80	37	-0.63	-0.19	0.83	-0.86	-0.75
52	-2.09	38	-0.74	-0.04	-0.09	0.59	-0.42
58	-2.46	39	-0.36	-0.28	0.30	0.13	-0.17
55	-2.86	40	-0.17	-0.12	0.16	-0.38	-0.38
20	-4.52	41	-0.06	-0.17	0.00	-0.06	-0.05
18	-6.35	42	-0.16	-0.11	0.84	-0.71	0.39
2	-6.82	43	0.35	0.15	-0.33	-0.85	0.12
6	-7.38	44	0.45	0.59	-1.43	-1.12	-1.05
8	-7.41	45	0.33	0.13	0.09	-0.72	-1.01
22	-7.97	46	0.64	0.15	1.02	1.44	0.26
1	-8.00	47	0.77	-0.12	-1.21	0.74	0.82
15	-8.38	48	0.33	0.14	1.63	1.81	0.60
17	-8.79	49	0.69	0.02	1.30	1.68	-0.78
14	-9.16	50	0.72	-0.55	-1.04	0.09	0.32
16	-9.44	51	1.03	0.21	-2.70	-1.56	0.29
7	-9.47	52	1.26	-0.05	0.31	0.19	-1.23
19	-9.84	53	1.45	0.42	2.00	-1.91	-1.26
12	-10.04	54	1.21	-0.64	-1.30	-0.41	1.62
13	-10.13	55	0.43	-0.70	-0.61	-1.25	-2.26
11	-10.20	56	1.04	0.12	0.12	0.23	1.26
3	-10.35	57	0.72	-0.21	0.56	0.51	-2.14
21	-10.36	58	1.22	-0.64	-2.00	-1.21	0.20
10	-10.74	59	1.05	0.10	0.28	0.62	-2.27
5	-10.77	60	1.84	-0.19	-2.28	-0.90	-0.78
4	-11.50	61	0.45	0.53	1.02	0.83	-0.85
9	-12.04	62	1.70	-0.33	-1.34	1.34	5.43

図9から固有値が1以上の10個の累積固有値は92%である。第1主成分と第2主成分は62.1%のデータのばらつきを表し，図7と図8のようにスコアプロットは2象限から4象限にほぼ直線上に×印の正常から腫瘍が散布している。図7のように因子負荷プロットは，全て4象限から1象限の第1主成分と正の相関がある。

SN	Eigenvalue	Contri.	20 40 60 80	Cum. Contri.	χ2	DF	p値(Prob>ChiSq)
1	30.3098	48.887		48.887	24576.8	1887.39	<.0001*
2	8.1905	13.211		62.097	20486.1	1884.69	<.0001*
3	4.6596	7.515		69.613	18838.6	1838.90	<.0001*
4	3.2501	5.242		74.855	17702.1	1786.69	<.0001*
5	3.0825	4.972		79.827	16779.4	1732.51	<.0001*
6	1.9764	3.188		83.014	15713.9	1678.98	<.0001*
7	1.7759	2.864		85.879	14941.0	1624.27	<.0001*
8	1.3849	2.234		88.113	14127.0	1570.17	<.0001*
9	1.2774	2.060		90.173	13402.4	1516.23	<.0001*
10	1.1785	1.901		92.074	12612.5	1463.10	<.0001*
11	0.8345	1.346		93.420	11730.1	1410.78	<.0001*
12	0.7116	1.148		94.568	11013.7	1358.74	<.0001*

図9　固有値

図10はWard法のDendrogramである。筆者は，多くの遺伝子解析の研究者の失敗は，高次元空間の分析を統計手法に頼りすぎた点にあると考えている。130個の判別スコアは，改定IP-OLDFで正常は-1以下に，腫瘍は1以上の値をとりMNM=0で判別できる。この情報を分析すれば，Ward法で2クラスターに分けた場合，図10の上のように正常の22件が最初のクラスターになり，図10の下のように癌の40件のクラスターに分かれた。さらに癌のクラスターは，最後のBGS61とBGS62の2件と残りの38件に分かれて図7と8のPCAで認められたことがWard法でも確認できた。

右に，ケースの樹状図が表示されている。最初にBGS1とBGS2がクラスターになり，そこにBGS6とBGS14がクラスターになる。BGS15とBGS22にBGS18とBGS8が次のクラスターになり，最初の4個でできたクラスターに融合される。このようなクラスタリングは，筆者は最短距離法の良くない特徴と理解していたが，Ward法でも確認できた。ヒートマップでは，全て青色になっている。一方，BGS55からBGS62までが癌の40件である。ヒートマップは，赤と青に分かれてて，変数と対応している。

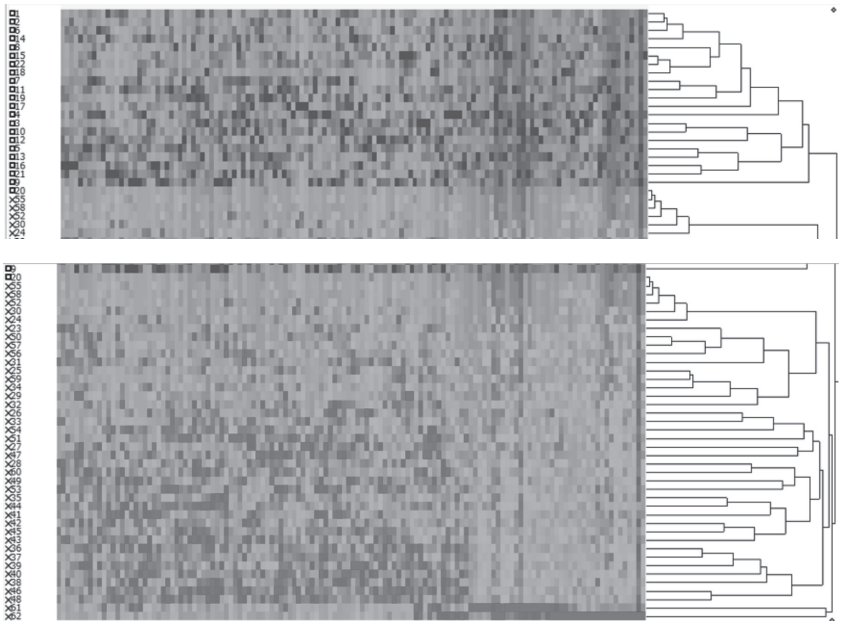


図10　Ward法のDendrogram

図11は，変数のクラスターの樹状図である。最初に2個の遺伝子がクラスターになるのは50組あり，その多くが小さな距離でクラスター化されている特徴がある。医学的な裏づけはないが，癌遺伝子が目的達成のためにお互い補完する組を準備しているのかもしれない。

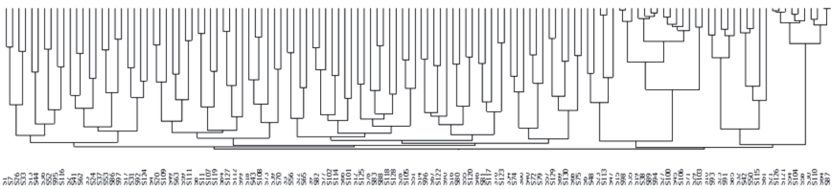


図11　変数のクラスター

表4は変数のクラスターである。図11の確認に用いればよい。N_clusterは，最初に右側から1番目と2番目のS64とS87が結合の距離が0.683と最短で一つのクラスターになり，130個が129個のクラスターになることから始まる。次に，cluster列の12でS30とS110が距離が1.966でクラスターになる。この2つのクラスターが，clusterの13で距離が2.345で一つのクラスターになる。一方，clusterの21でS21とS126がクラスターになり，その後clusterの22でS126が結合し3個がクラスターになる。そして同じようにclusterの23と24でS54とS104にS58が結合し一つのクラスターになる。これらの3個のBGSがclusterの25で6個のBGSが一つ

のクラスターになっている。以上の右側から6個の一番距離が近いクラスターを特定するにも，表4は大変である。クラスター数を指定すると，それに含まれる各クラスターに属するケースを表示する機能が必要である。しかし，がん診断の専門家であれば，20個前後のクラスターに何か医学的な意味を見つけれるかもしれない。多くの研究者が，全体の遺伝子をクラスター分析で分析している例も見られるが，この表の分析の方が有意義な情報が得られると考えられる。

表4　変数のクラスター

N_cluster	Dist.	Join1	Join2	cluster
129	0.683	S64	S87	11
128	0.751	S15	S98	
127	0.787	S21	S126	21
126	0.893	S21	S121	22
125	1.100	S89	S94	
124	1.548	S54	S104	23
123	1.602	S23	S113	
122	1.626	S38	S89	
121	1.892	S27	S100	
120	1.917	S19	S38	
119	1.966	S30	S110	12
118	1.995	S54	S58	24
117	2.032	S15	S35	
116	2.080	S21	S54	25
115	2.186	S27	S45	
114	2.345	S30	S64	13
113	2.541	S27	S106	
112	2.660	S29	S103	
111	3.183	S42	S50	
110	3.210	S25	S91	
109	3.482	S27	S71	
108	3.550	S84	S127	
107	3.631	S32	S42	
106	3.680	S8	S11	
105	3.684	S10	S93	
104	3.724	S25	S85	
103	3.735	S43	S108	
102	3.829	S57	S129	
101	3.870	S86	S97	
100	3.898	S59	S111	
99	3.911	S19	S27	
98	3.911	S76	S123	
97	3.928	S21	S30	
96	3.997	S4	S20	
95	4.021	S83	S88	
94	4.041	S1	S7	
93	4.062	S9	S82	
92	4.107	S36	S52	
91	4.108	S37	S53	
90	4.161	S26	S33	
89	4.190	S5	S56	
88	4.224	S41	S62	
87	4.239	S40	S72	
86	4.255	S68	S75	
85	4.257	S55	S120	
84	4.281	S10	S25	
83	4.294	S14	S96	
82	4.307	S28	S105	
81	4.337	S69	S130	
80	4.360	S15	S19	
79	4.366	S46	S81	
78	4.390	S16	S80	
77	4.399	S67	S122	
76	4.409	S95	S116	
75	4.409	S17	S31	
74	4.448	S39	S70	
73	4.453	S9	S77	
72	4.455	S107	S119	
71	4.469	S13	S44	
70	4.496	S6	S48	
69	4.521	S49	S63	

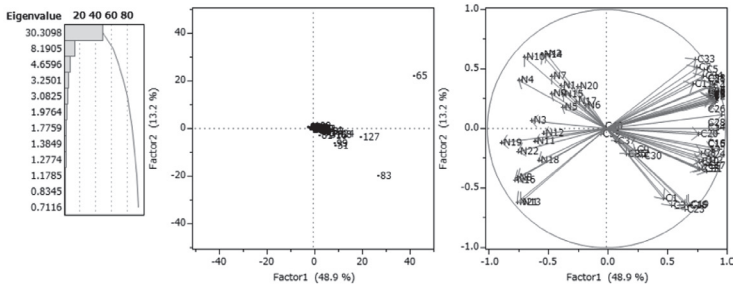
68	4.522	S102	S114	
67	4.540	S34	S74	
66	4.551	S23	S47	
65	4.554	S3	S24	
64	4.607	S18	S43	
63	4.608	S92	S124	
62	4.634	S4	S109	
61	4.662	S40	S79	
60	4.664	S12	S125	
59	4.664	S84	S112	
58	4.671	S32	S115	
57	4.693	S22	S65	
56	4.738	S28	S51	
55	4.742	S15	S29	
54	4.762	S78	S83	
53	4.811	S34	S60	
52	4.830	S37	S86	
51	4.857	S118	S128	
50	4.883	S2	S41	
49	4.912	S57	S69	
48	4.957	S17	S92	
47	4.961	S66	S101	
46	4.985	S49	S59	
45	4.996	S32	S61	
44	5.053	S46	S117	
43	5.070	S6	S23	
42	5.073	S18	S73	
41	5.205	S1	S26	
40	5.274	S16	S55	
39	5.282	S8	S107	
38	5.299	S2	S3	
37	5.357	S36	S95	
36	5.357	S57	S68	
35	5.420	S46	S76	

34	5.434	S14	S67	
33	5.440	S78	S118	
32	5.460	S9	S102	
31	5.532	S14	S90	
30	5.573	S84	S99	
29	5.630	S5	S22	
28	5.828	S12	S78	
27	5.861	S9	S66	
26	5.884	S4	S49	
25	6.023	S18	S39	
24	6.140	S13	S36	
23	6.150	S14	S16	
22	6.199	S12	S28	
21	6.232	S2	S37	
20	6.262	S34	S40	
19	6.340	S10	S32	
18	6.440	S8	S84	
17	6.558	S2	S17	
16	6.660	S4	S8	
15	6.688	S9	S12	
14	6.905	S14	S46	
13	7.042	S1	S13	
12	7.117	S5	S9	
11	7.446	S4	S18	
10	7.926	S34	S57	
9	8.408	S5	S14	
8	8.526	S6	S15	
7	8.609	S1	S2	
6	8.648	S5	S34	
5	8.699	S10	S21	
4	8.861	S4	S5	
3	9.664	S1	S4	
2	16.646	S6	S10	
1	33.558	S1	S6	

4.3 転置行列による検証

62人をケースとして2000個の遺伝子を変数とするデータから，改定IP-OLDFで130個のBGSとB0に分け，それらの判別係数とする130個の判別スコアのデータに縮約した62行130列のデータの分析を行ってきた。これを転置し，130行62変数のデータを作成しPCAで分析すると図12の結果が得られた。N1からN22は正常を表す変数で，C1からC40は癌患者を表

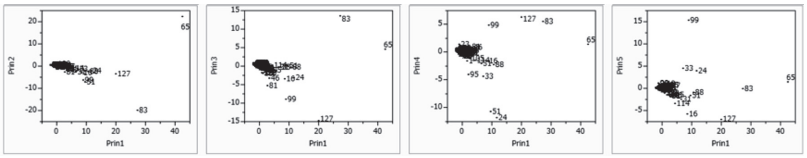
す変数である。これらは当然であるが，2象限と4象限，そして1象限と4象限にきれいに分れて布置しているようだが，矢印に隠れてC39とC40は-0.04と-0.03で3象限と2象限に布置している。一方，真ん中のスコアプロットは，130個のBGSの判別スコアである。統計的に65番のBGSが1象限に，127と83番と重なっているが99と51番のBGSによる判別スコアが4象限に他と離れて布置している。他の多くのBGSは，第1主成分の正の軸と重なっている。統計的には，これらの5個のBGSが他の125個より癌診断で特徴的と考えられるが，医学的に何を意味するか専門家の検討が必要である。



1：固有値，スコア図と因子負荷図

SN	Eigenvalue	Contri.	20 40 60 80	Cum. Contri.	χ2	DF	p値(Prob>ChiSq)
1	30.3098	48.887		48.887	24576.8	1887.39	<.0001*
2	8.1905	13.211		62.097	20486.1	1884.69	<.0001*
3	4.6596	7.515		69.613	18838.6	1838.90	<.0001*
4	3.2501	5.242		74.855	17702.1	1786.69	<.0001*
5	3.0825	4.972		79.827	16779.4	1732.51	<.0001*
6	1.9764	3.188		83.014	15713.9	1678.98	<.0001*
7	1.7759	2.864		85.879	14941.0	1624.27	<.0001*
8	1.3849	2.234		88.113	14127.0	1570.17	<.0001*
9	1.2774	2.060		90.173	13402.4	1516.23	<.0001*
10	1.1785	1.901		92.074	12612.5	1463.10	<.0001*
11	0.8345	1.346		93.420	11730.1	1410.78	<.0001*

2：固有値の詳細



3：横軸が第1主成分，縦軸が第2から第5主成分のスコア図

図12　転置行列のPCA

第1主成分と第2主成分の左端のスコア図では5個のBGSの判別スコアが他と異なっているようだが，第3主成分から第5主成分を縦軸とするプロット図では少なくとも10個以上が他と異なっているようである。これらは医学的な知見と合わせて検討すべきであろう。

5. 2000個の遺伝子の一元配置の分析

表5は2000個の遺伝子の一元配置の分散分析を130個のBGSに分けて分析したt値である。SN列は，表1のSNに対応している。Gene列は各BGSに含まれる遺伝子数である。次の4列は得られたt値の最大値，最小値，範囲と平均値である。次の4列はt値の絶対値をとった最大値，最小値，範囲と平均値である。この最小値を見ると全て1未満であり，ほどんとは2群の平均値に差のない遺伝子が多いことを示す。これが，多くの研究者がt検定を頼りにしても有意義な結論を得られない理由の一つである。即ち，BGSに選ばれた遺伝子の組み合わせ全てを考えることで，2群をMNM=0で判別できることを示している。130個のBGSの最小値と最大値を判別スコアのt値と比較しても決して値が小さいわけではない。多くの研究者が一元配置の分散分析でアプローチしているが，各遺伝子のt値が大きいか小さいかは，2群が線形分離可能か否かの情報を検出するのに適していないことを示す。2群が完全に分かれているかいないかは，平均値の小さい正常群の最大値が，平均の大きな癌群の最小値よりも小さいことを意味するが，これを判別スコアであってもt検定でとらえることはできない。また，癌群に異常値があれば，t検定はその外れ値に大きく影響を受けることも今回の分析で分かった。結局30年以上かけて思い込みで統計分析してきたが，それが的を得なかっただけである。

表5 2000個の遺伝子の一元配置の分散分析を130個のBGSに分けたt値

SN	Gene	Max (tvalue)	Min (tvalue)	Range (tvalue)	Mean (tvalue)	Max (AbsT)	Min (AbsT)	Range (AbsT)	Mean (AbsT)
0	5	0.4	-3.15	3.56	-1.3	3.15	0.01	3.14	1.47
1	20	3.4	-5.6	9	-0.28	5.6	0.09	5.51	1.91
2	13	3.4	-5.24	8.64	0.15	5.24	0.01	5.24	1.51
3	17	4.95	-4.06	9.01	0.87	4.95	0.05	4.91	1.92
4	14	4.35	-3.44	7.79	0.73	4.35	0.89	3.46	2.55
5	16	4.49	-5.58	10.07	-0.69	5.58	0.5	5.08	2.45
6	18	4.23	-3.99	8.22	0.36	4.23	0.05	4.18	1.81
7	15	3.54	-2.98	6.52	-0.25	3.54	0.03	3.51	1.56
8	13	4.04	-4.42	8.46	0.87	4.42	0.03	4.39	2.33
9	16	2.42	-3.46	5.88	-0.84	3.46	0	3.46	1.34
10	10	3.7	-5.38	9.09	-0.47	5.38	0.14	5.24	1.7
11	12	5.83	-3.89	9.72	0.61	5.83	0.16	5.67	1.92
12	19	2.62	-2.87	5.5	-0.11	2.87	0.02	2.86	1.1

13	16	1.98	-3.42	5.4	-0.7	3.42	0.18	3.24	1.57
14	9	4.69	-3.06	7.75	1.02	4.69	0.02	4.67	2.25
15	12	3.14	-4.34	7.48	-0.5	4.34	0.2	4.14	1.73
16	19	2.25	-3.99	6.24	-0.15	3.99	0	3.99	1.39
17	13	2.48	-2.53	5.01	-0.43	2.53	0.13	2.4	1.52
18	18	4.44	-5.33	9.77	-0.04	5.33	0.18	5.16	1.44
19	12	2.75	-6.87	9.62	-1.01	6.87	0.81	6.06	2.58
20	15	3.83	-2.83	6.66	1.1	3.83	0.25	3.59	1.91
21	13	2.25	-3.98	6.23	-0.4	3.98	0.25	3.73	1.56
22	19	2.76	-2.34	5.1	0.3	2.76	0.08	2.68	1.29
23	14	2.78	-4.39	7.17	-0.21	4.39	0.05	4.34	1.72
24	18	3.12	-3.39	6.52	0.16	3.39	0	3.39	1.39
25	12	3.17	-2.65	5.82	0.4	3.17	0.42	2.75	1.83
26	13	4.97	-1.59	6.57	0.09	4.97	0.21	4.77	1.1
27	10	1.93	-7.24	9.17	-0.4	7.24	0.12	7.12	1.53
28	16	5	-3.47	8.47	-0.04	5	0.03	4.97	1.36
29	16	4.33	-2.52	6.84	0.09	4.33	0.05	4.28	1.13
30	15	2.14	-4.8	6.95	-0.76	4.8	0.12	4.68	1.62
31	21	3.46	-3.95	7.41	-0.28	3.95	0.16	3.78	1.55
32	15	2.61	-4.03	6.65	-0.18	4.03	0.07	3.96	1.58
33	18	2.09	-2.82	4.91	-0.31	2.82	0.05	2.77	1.14
34	15	4.29	-1.84	6.13	0.26	4.29	0.22	4.07	1.48
35	11	2.56	-7.93	10.48	-0.8	7.93	0.23	7.69	2.02
36	11	6.24	-2.76	9	1.12	6.24	0.26	5.98	2.49
37	20	3.89	-2.98	6.87	0.24	3.89	0.05	3.84	1.31
38	14	3.45	-4.03	7.48	-0.44	4.03	0.72	3.3	2.06
39	18	3.09	-4.28	7.37	-1.37	4.28	0.23	4.06	1.79
40	12	4.42	-2.51	6.93	1.12	4.42	0.13	4.29	1.8
41	15	2.58	-1.42	4	0.24	2.58	0.05	2.53	0.94
42	16	3.65	-3.98	7.63	-0.55	3.98	0.01	3.97	1.55
43	14	6.41	-3.75	10.15	0.12	6.41	0.07	6.34	2.49
44	13	4.07	-3.38	7.45	0.53	4.07	0.09	3.98	1.48
45	21	3.82	-2.34	6.16	0.01	3.82	0.08	3.74	1.42
46	13	1.7	-4.02	5.72	-0.74	4.02	0.06	3.97	1.36
47	16	3.18	-3.44	6.63	-0.38	3.44	0.44	3	1.58
48	15	3.95	-2.38	6.33	0.19	3.95	0.02	3.93	1.43
49	21	2.55	-3.91	6.46	-0.1	3.91	0.19	3.73	1.41
50	9	2.99	-5.2	8.19	-1.15	5.2	0.59	4.61	2.36
51	25	2.59	-3.66	6.25	0.14	3.66	0.01	3.65	1.09
52	12	1.77	-5.13	6.9	-1.12	5.13	0.02	5.11	1.61
53	14	5.09	-4.16	9.24	0.32	5.09	0.12	4.97	2.1
54	12	2.18	-3.69	5.87	-1.24	3.69	0.02	3.67	1.8

55	16	4.74	-3.62	8.36	-0.81	4.74	0.04	4.71	1.73
56	9	1.95	-5.82	7.77	-2.12	5.82	0.15	5.66	2.67
57	18	4.3	-2.56	6.87	0.05	4.3	0.02	4.28	1.69
58	16	2.25	-2.91	5.16	-0.12	2.91	0.06	2.85	0.94
59	15	1.34	-3.94	5.27	-0.76	3.94	0.05	3.88	1.3
60	16	3.68	-4.28	7.96	-0.02	4.28	0.05	4.23	1.31
61	14	2.65	-6.32	8.98	-0.62	6.32	0.1	6.22	1.46
62	19	2.11	-3.59	5.7	-0.5	3.59	0.2	3.38	1.25
63	20	2.47	-4.24	6.71	-0.33	4.24	0.04	4.2	1.64
64	12	3.19	-2.3	5.48	0.51	3.19	0.09	3.09	1.2
65	17	5.6	-2.34	7.94	0.49	5.6	0.1	5.5	1.47
66	21	4.49	-3.28	7.76	-0.05	4.49	0.01	4.48	1.52
67	15	4.65	-3.27	7.92	-0.02	4.65	0.03	4.62	1.7
68	13	4.18	-2.76	6.94	-0.46	4.18	0.06	4.12	1.49
69	11	2.24	-3.19	5.43	-0.72	3.19	0.15	3.04	1.64
70	16	6.36	-1.76	8.12	0.63	6.36	0.12	6.24	1.8
71	16	5.37	-3.19	8.57	-0.05	5.37	0.12	5.25	1.81
72	18	3.55	-1.86	5.42	0.49	3.55	0.17	3.39	1.37
73	16	2.84	-3.21	6.06	-0.1	3.21	0.01	3.2	1.36
74	16	3.95	-4.32	8.28	-0.19	4.32	0.02	4.3	1.59
75	15	3.93	-2.49	6.42	0	3.93	0.11	3.82	1.56
76	18	2.93	-2.33	5.26	0.39	2.93	0.06	2.86	1.46
77	15	4.58	-2.26	6.84	0.17	4.58	0.01	4.56	1.33
78	15	4.32	-2.4	6.72	0.71	4.32	0.26	4.06	1.79
79	16	3.78	-1.54	5.32	-0.13	3.78	0.24	3.55	1.13
80	18	3.95	-3.64	7.58	0.81	3.95	0.01	3.93	1.74
81	19	3.3	-2.1	5.4	0.46	3.3	0.12	3.18	1.3
82	11	3.99	-4.5	8.49	-0.61	4.5	0.12	4.39	1.82
83	18	4.39	-4.47	8.86	-0.13	4.47	0.32	4.15	1.98
84	19	4.74	-4.04	8.78	-0.11	4.74	0.02	4.72	1.59
85	12	3.88	-4.2	8.08	-0.28	4.2	0.09	4.11	1.69
86	9	4.32	-3.18	7.5	0.07	4.32	0.01	4.31	1.34
87	17	2.89	-2.68	5.57	0.36	2.89	0.17	2.72	1.33
88	13	4.21	-3.37	7.58	0.43	4.21	0.06	4.16	2.03
89	18	4.33	-2.94	7.28	0.14	4.33	0.08	4.25	1.49
90	17	3.05	-2.65	5.7	0.2	3.05	0.13	2.93	1.15
91	12	4.07	-3.25	7.32	0.52	4.07	0.07	4	1.74
92	17	2.06	-4.25	6.31	-0.68	4.25	0.02	4.22	1.66
93	12	3.49	-4.33	7.82	0.28	4.33	0.25	4.08	1.88
94	11	5.48	-3.25	8.73	0.92	5.48	0.41	5.07	2.2
95	15	1.12	-6.68	7.8	-0.64	6.68	0.11	6.57	1.39
96	15	3.43	-2.75	6.18	0.53	3.43	0.03	3.39	1.27

97	19	2.76	-4.32	7.08	-0.01	4.32	0.06	4.26	1.24
98	14	4.2	-2.41	6.61	0.59	4.2	0.05	4.15	1.58
99	17	3.62	-2.77	6.38	-0.26	3.62	0.33	3.29	1.73
100	16	4.85	-5.89	10.73	-0.31	5.89	0.16	5.73	1.76
101	19	3.13	-3.12	6.25	-0.2	3.13	0.03	3.11	1.35
102	17	4.01	-3.32	7.33	0.24	4.01	0.28	3.73	1.67
103	11	4.57	-2.7	7.28	0.5	4.57	0.05	4.53	2.03
104	16	3.48	-3.41	6.88	-0.23	3.48	0.17	3.31	1.93
105	16	5.61	-3.4	9.02	0.35	5.61	0.01	5.6	1.57
106	17	2.9	-4.1	7	0.48	4.1	0.09	4.01	1.34
107	18	3.48	-3.72	7.2	0.17	3.72	0.05	3.67	1.61
108	18	4.81	-7.86	12.67	0.41	7.86	0.22	7.64	2.16
109	16	5.03	-3.72	8.74	0.22	5.03	0.04	4.98	1.71
110	14	4.48	-5.26	9.74	-0.55	5.26	0.25	5	2.47
111	17	4.8	-2.3	7.1	0.59	4.8	0.09	4.71	1.23
112	16	2.49	-2.48	4.97	-0.42	2.49	0.25	2.24	1.11
113	21	3.43	-2.43	5.86	-0.15	3.43	0.02	3.41	1.19
114	17	5.55	-2.02	7.57	0.69	5.55	0	5.55	1.24
115	19	3.31	-4.06	7.36	-0.09	4.06	0.02	4.03	1.56
116	16	6.11	-2.72	8.83	0.88	6.11	0.06	6.04	1.72
117	14	6.08	-3.16	9.24	0.16	6.08	0.06	6.02	1.68
118	13	3.21	-0.13	3.34	1.36	3.21	0.13	3.08	1.38
119	17	3.22	-2.64	5.86	-0.19	3.22	0.11	3.12	1.32
120	17	4.3	-4.33	8.64	0.24	4.33	0.15	4.18	1.42
121	16	1.9	-6.65	8.55	-0.46	6.65	0.09	6.56	1.15
122	17	5.01	-3.22	8.23	0.45	5.01	0.02	4.99	1.71
123	13	4.45	-4.67	9.12	0.13	4.67	0.07	4.6	2.09
124	15	2.89	-3.16	6.05	0.41	3.16	0.39	2.77	1.53
125	15	3.51	-4.25	7.75	-0.02	4.25	0.1	4.15	1.52
126	15	3.74	-3.5	7.24	-0.81	3.74	0.34	3.4	1.79
127	14	4.43	-2.38	6.81	1.12	4.43	0	4.43	1.7
128	11	3.46	-4.79	8.26	0.17	4.79	0.23	4.57	1.87
129	12	3.98	-3.89	7.87	-0.55	3.98	0.08	3.9	1.68
130	12	2.68	-4.02	6.7	0.34	4.02	0.19	3.83	1.65
	MAX	6.41	-0.13	12.67	1.36	7.93	0.89	7.69	2.67
	MIN	0.4	-7.93	3.34	-2.12	2.49	0	2.24	0.94
	MEAN	3.65	-3.63	7.27	-0.02	4.39	0.13	4.26	1.62

6. まとめ

これまで30年以上かけて，高次元の遺伝子空間を通常の統計手法で分析し数多くの論文が発表されてきたが大きな成果が得られなかった。しかし筆者の開発した3種のOLDFだけが，

LSDである高次元の遺伝子空間から，数10個以下のLSDである部分空間を見つけることができた。SVMはLSDであることは正しく指摘するが，なぜか癌遺伝子を特定できなかった。FisherのLDFは全てNMが0にならず誤分類確率も大きかった。しかし改定IP-OLDFは，6種のMicroarrayデータはLSDである小さな遺伝子の部分空間すなわちSMの排他的な和集合であることを示した。本研究では，その中のAlon他のデータが，130個のBGSの排他的和集合であることが分かった。本研究では，130個のBGSをロジスティック回帰で分析すると全てNM=0であるが，QDFは63個だけがNM=0であり，FisherのLDFは全て0でなかった。

そこで普通の統計手法でBGSを検証したが，2000個の遺伝子（変数）を持つデータを通常の統計手法で分析してもはっきりした結果を得られなかったので，130個の改定IP-OLDFの判別スコアを変数とするデータを分析することにした。

判別スコアをPCAで分析すると，図7の第1主成分軸上で2群がほぼ直線上になり，2群が目視でLSDであることが分かった。130個の判別スコアのRatioSVの範囲は［0.01%, 0.901%］であるのに対して，130個の判別スコアを総合特性値化した第1主成分で，判別スコアの範囲は［-2.152, 42.434］でRatioSV=200/44.586=4.48%である。130個のRatioSVの最大値の0.901%に比べて約5倍であり，目視で簡単に2群が判別できる。一方，転置行列で分析すると5個から10個以上のBGSが他と異なった特徴を持っていることが分かった。Ward法で，正常と癌が完全に2クラスターに分かれた。以上から，癌診断に判別スコアを用いると有意義であることが分かった。また，今のところ分析技術が未熟なためかもしれないが一元配置の分散分析とt値では，2群がLSDであることを明確に示せなかった。さすがに人類の究極の疾病である癌であり，一筋縄ではいかないことが分かった。

Fisherらは，統計学に推測統計と考え方を導入し，統計を科学的な学問に棚上げした。田邉（2011）は，その考え方を次のように紹介している。

「有意性検定において検証される仮説は「帰無仮説」と呼ばれます。この仮説の評価は統計量の観測データに基づいて行います。その手続きは，まずこの仮説が真であると仮定したときに観測されるべき統計量の分布を数学的に割り出し，この仮説の下で稀にしか起こらない事象群を定め，観測データがこの稀な事象群に入るなら，観測データから見てこの仮説は考慮に値しないものとして「棄却」します。棄却の論理的な意味は「仮説が真であるがその下で非常に稀な事象が置きたか，あるいは仮説自体が真でないかのどちらかである」とFisher自身が述べています。

すなわち，Fisher自身はFisherの仮説を現実のデータが満たさないとき，推測の結果は保証できないと明確に述べている。またFisherのIrisデータと呼ばれる現実のデータで謙虚にFisherのLDFを評価している。またFisherの仮説を満たさないで，2群が等分散でないときQDFが提案されている。しかし，現実のデータを出発点にしないで，正規分布と分散共分散

行列から数学的な展開で，その後に一般化逆行列，RDA，LASSOなどが研究されている。しかしCoxはCox回帰やロジスティック回帰を医学診断に開発したが，彼こそFisherの正当な第2世代の後継者と考えている。VapnikはMPによるH-SVMでLSD判別を定義した。現実のデータはLSDであることは稀であるので，オーバーラップするデータにQPで解けるS-SVMを定式化し，その後でKernel-SVMを提案した。多くの研究者が，魅力的なKernel-SVMの研究に注意が払われ，LSD判別が行われなかったと考える。彼がFisherの第3世代の後継者とすれば，筆者はこれらの成果を踏まえ，LSDであろうがなかろうが，これまでの判別関数をMethod1で評価検証した。そしてLSD研究が，遺伝子解析の解析に最も適していたため，筆者の研究テーマの中で一般的に一番難しい問題5が僅か54日で確立できたといえる。

（成蹊大学名誉教授）

REFERENCES

1. Alon, U.et al.(1999)“Broad Patterns of Gene Expression Revealed by Clustering Analysis of cancwe and Normal Colon Tissues Probed by Oligonucleotide Arrays.”*Proc. Natl. Acad. Sci*. USA, 96, 6745-6750.
2. Cox, D.R. (1958)“The regression analysis of binary sequences (with discussion).”*J Roy Stat Soc* B 20: 215-242.
3. Fisher, R. A. (1936)“The Use of Multiple Measurements in Taxonomic problems.”*Annals of Eugenics*, 7, 179–188.
4. Fisher, R. A. (1956) *Statistical methods and statistical inference*. Hafner Publishing Co.
5. Flury, B., Riedwyl,H. (1988) Multivariate Statistics: A Practical Approach. Cambridge University Press.
6. Golub, T.R. et al. (1999)“Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring.”*Science*. 1999 Oct 15; 286(5439): pp. 531-537.
7. Jeffery, IB. Higgins, DG. Culhane, AC. (2006)“Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data.”*BMC Bioinformatics*. Jul 26; pp. 7:359. http://www.bioinf.ucd.ie/people/ian/
8. Miyake, A., Shinmura, S. (1976) *Error rate of linear discriminant function*, F.T. de Dombal & F. Gremy editors 435 - 445, North-Holland Publishing Company.
9. Sall, J. P., Creighton, L., Lehman, A. (2004) *JMP Start Statistics, Third Edition*. SAS Institute Inc. (Shinmura, S. edited Japanese version)
10. Schrage, L. (2006) *Optimization Modeling with LINGO.* LINDO Systems Inc. (Shinmura, S. translated Japanese version)

11. *Shinmura, S. (2000a) "A new algorithm of the linear discriminant function using integer programming." *New Trends in Probability and Statistics*, 5, 133-142.
12. *_____ (2000b) *Optimal Linear Discriminant Function using Mathematical Programming*. Dissertation, March 200, 1-101, Okayama Univ.
13. *_____ (2003) "Enhanced Algorithm of IP-OLDF." *ISI2003 CD-ROM*, 428-429.
14. *_____ (2004) "New Algorithm of Discriminant Analysis using Integer Programming." *IPSI 2004 Pescara VIP Conference CD-ROM*, 1-18.
15. *_____ (2005) "New Age of Discriminant Analysis by IP-OLDF –Beyond Fisher's Linear Discriminant Functions." *ISI2005*, 1-2.
16. *_____ (2007b) "Comparison of Revised IP-OLDF and SVM." *ISI2009*, 1-4.
17. *_____ (2009) "Practical discriminant analysis by IP-OLDF and IPLP-OLDF." *IPSI 2009 Belgrade VIPSI Conference CD-ROM*, 1-17.
18. *_____ (2011) "Beyond Fisher' s Linear Discriminant Analysis - New World of Discriminant Analysis -." *ISI2011 CD-ROM*, 1-6.
19. *_____ (2013) "Evaluation of Optimal Linear Discriminant Function by 100-fold Cross Validation." *ISI2013 CD-ROM*, 1-6.
20. *_____ (2014a) "End of Discriminant Functions based on Variance-Covariance Matrices." *ICORES*, 5-14.
21. *_____ (2014b) "Improvement of CPU time of Linear Discriminant Function. Statistics." *Optimization and Information Computing*, vol. 2, 114-129.
22. *_____ (2014c) "Comparison of Linear Discriminant Functions by K-fold Cross Validation." *Data Analytics 2014*, 1-6.
23. *_____ (2015a) "The 95% confidence intervals of error rates and discriminant coefficients." *Statistics, Optimization and Information Computing*, vol. 3, 66-78.
24. *_____ (2015b) "Four Serious problems and New Facts of the Discriminant Analysis." E. Pinson et al. (Eds.) *ICORES 2014 Revised and Selected Papers*, CCIS 509, 15-30, Springer. ISSN: 1865-0929, ISBN: 978-3-319-17508-9, DOI: 10.1007/978-3-319-17509-6.
25. *_____ (2015c) "A Trivial Linear Discriminant Function." *Statistics, Optimization, and Information Computing*, Vol.3, December 2015, 322-335. DOI: 10.19139/soic.20151202.
26. *_____ (2015d) "The Discrimination of the microarray data (Ver. 1)." *Research Gate (1)*, Oct. 28, 2015, 1-4.
27. *_____ (2015e) "Feature Selection of three Microarray data." *Research Gate (2)*, Nov.1, 2015, 1-7.

28. *_____ (2015f) "Feature Selection of Microarray Data (3) – Ship et al. Microarray Data." *Research Gate (3)*, 2015, 1-11.

29. *_____ (2015g) "Validation of Feature Selection (4) – Alon et al. Microarray Data." *Research Gate (4)*, 2015, 1-11.

30. *_____ (2015h) "Repeated Feature Selection Method for Microarray Data (5)." *Research Gate (5)*, Nov. 9, 2015, 1-12.

31. *_____ (2015i) "Comparison Fisher's LDF by JMP and Revised IP-OLDF by LINGO for Microarray Data (6)" *Research Gate (6)*, Nov. 11, 2015, 1-10.

32. *_____ (2015j) "Matroska Trap of Feature Selection Method (7) –Golub et al. Microarray Data." *Research Gate (7)*, Nov. 18, 2015, 1-14.

33. *_____ (2015k) "Minimum Sets of Genes of Golub et al. Microarray Data (8)." *Research Gate (8)*, Nov. 22, 2015, 1-12.

34. *_____ (2015l) "Complete Lists of Small Matroska in Shipp et al. Microarray Data (9)." *Research Gate (9)*, Dec. 4, 2015,1-81.

35. *_____ (2015m) "Sixty-nine Small Matroska in Golub et al. Microarray Data (10)." *Research Gate (10)*, Dec. 4, 1-58.

36. *_____ (2015n) "Simple Structure of Alon et al. et al. Microarray Data (11)." *Research Gate (11)*, Dec. 4, 2015, 1-34.

37. *_____ (2015o) "Feature Selection of Singh et al. Microarray Data (12)." *Research Gate (12)*, Dec. 6, 2015, 1-89.

38. *_____ (2015p) "Final List of Small Matroska in Tian et al. Microarray Data." *Research Gate (13)*, Dec. 7, 1-160.

39. *_____ (2015q) "Final List of Small Matroska in Chiaretti et al. Microarray Data." *Research Gate (14)*, Dec. 20, 2015, 1-16.

40. *_____ (2015r) "Matroska Feature Selection Methods for Microarray Data," *Research Gate Free paper (15)*, 1-16.

41. *_____ (2016a) "Matroska Feature Selection Method for Microarray Data." *Biotechno 2016*, 1-6.

42. *_____ (2016b) "Discriminant Analysis of the Linear Separable Data -Japanese automobiles-." *Journal of Statistical Science and Application*, vol4, No.07-08, 165-178. DOI: 10.17265/2328-224X/2016.0708.001.

43. *_____ (2016c) "The Best Model of the Swiss Banknote Data-Validation by the 95% CI of error rates and discriminant coefficients -." *Optimization, and Information Computing*, Vol.3, 322-

335, 2015. DOI: 10.19139/soic.20151202.

44. *_____ (2016d) "The K-fold Cross Validation for Small Sample Method." *Data Analytic 2016*, 1-6.
45. *_____ (2016f) *The New Theory of Discriminant Analysis after R Fisher*, Springer. DOI: 10.1007/978-981-10-2164-0
46. *_____ (2017a) Validation of Matroska Feature Selection Method for Microarray Data by LINGO Program 1 using Japanese Automobile and Swiss Banknote Data. Biotechno 2017, pp.1-24.
47. Simon N, Friedman J, Hastie T, Tibshirani R (2013) "A sparse-group lasso." *J. Comput. Graph. Statist*, 22:231-245
48. Vapnik, V. (1995) *The Nature of Statistical Learning Theory*. Springer-Verlag.
49. 新村秀一 (1984)「医療データ解析, モデル主義そしてOR」. 『オペレーションズ・リサーチ, 29/7』, 415-421.
50. _____訳著 (1986)『SASによる回帰分析の実践』. 朝倉書店.
51. _____ (1996)「重回帰分析と判別分析のモデル決定 (2) : 19変数をもつC.P.D.データのモデル決定」。『成蹊大学経済学部論集』, 27/1, 180-203.
52. _____ (1998)「数理計画法を用いた最適線形判別関数」. 『計算機統計学, 11/2』, 89-101.
53. 新村秀一, 垂水共之 (2000)「乱数データを用いた最適線形判別関数の評価」. 『計算機統計学, 12/2』, 107-123.
54. 新村秀一 (2004)『JMP活用　統計学とっておき勉強法』. 講談社.
55. _____ (2007a)「改定IP-OLDFによるIP-OLDFの問題点の解消」.『計算機統計学, 19 / 1』, 1-16.
56. _____ (2007b)「数理計画法による判別分析の10年」. 『計算機統計学, 20 / 1 & 2』, 59-94.
57. _____ (2010a)『最適線形判別関数』. 日科技連出版.
58. _____ (2010b)「線形計画法による改定IP-OLDFの計算時間の改善」. 『計算機統計学, 22/1』, 37-57.
59. _____ (2011a)「合否判定データによる判別分析の問題点」. 『応用統計学, 40/3』, 157-172.
60. _____ (2011b)『数理計画法による問題解決法』. 日科技連出版.
61. _____ (2012)「コラム「SAS/JMPとの歩み」, SAS Technical News, 春, 夏, 秋, 冬号」.
62. _____ (2015a)「いかに研究成果を世界に発信するか―判別分析の4つの問題と新事実

—」. 『SASユーザー会』, 484-493.

63. **_____（2016a）「判別分析の新理論と遺伝子解析」,『第9回コンピューテーショナル・インテリジェンス研究会』, 77-84.
64. **_____（2016b）「判別分析の新理論と遺伝子解析のための新手法2」,『成蹊大学経済学部論集』, 第47巻第1号, 43-77.
65. **_____（2016c）「R.A. Fisher以後の判別分析の新理論（1）—遺伝子解析の新手法2（LINGO Program3）の検証—」,『成蹊大学経済学部論集』, 第47巻第2号, 1-42.
66. 竹内啓（20011）書評:小西定則「多変量解析入門—線形から非線形へ—」, 新村秀一「最適線形判別関数」. 統計, 71-74.
67. 田邉國士（2011）「応用数理の遊歩道（67）帰納という原罪」. 『応用数理』, 304-309.
68. 三宅章彦, 新村秀一（1980）「最適線形判別関数のアルゴリズムとその応用,『医用電子と生体工学』, 18/6, 452-454.

Researchers can download author's papers with * or ** before author's name from the Research Gate and Research Map