

## 「SPSS活用論」

小山明宏\*

### 〔1〕はじめに

本学でSPSS(Statistical Package for the Social Sciences)が導入されたのは、ごく最近であると聞く。SPSSはもともと、わが国では富士通、日立、あるいはIBMのマシン用に導入されていたようであるし、その点、本学のコンピュータ用に改変することは相当御苦労があったのではないと思われるのである。だから、本学でコンピュータを利用できる立場にある方々にはぜひこれを活用していただきたく思うのであるが、残念ながら、聞くところによると、導入以後現在までの経済学部の利用者はせいぜい数人だそうで、甚だ残念なことである。そこで、というわけでもないが、本稿ではその「1人ぼっちの利用」の概況を簡単にまとめ、あわせてSPSSの、一種のPRを行いたいと思う。

筆者が1981年4月に本学に赴任するまで、大学生活をすごした一橋大学では、SPSSは驚異的なまでに使われており、毎日誰かがそのアウトプットを抱えて計算機センター付近をうろろろしていたものである。筆者自身もそのうちの一人で、博士課程の単位修得論文で行った、企業の資本構成に関する実証研究でも、DISCRIMINANTを始めとして、CONDENSATIVE, FREQUENCIES, SCATTERGRAM, PEARSON CORR, NONPAR CORR, REGRESSION, CLUSTER等々、ありとあらゆる手法をゴチャゴチャに、いわば「満艦飾」的に使用した記憶がある。ただし、このような利用のしかたは悪い例なので、ここでは問題をより小さくしぼって、判別分析による簡単な研究例を提示してみよう。

ところで、本学のSPSSに入っているものは、その関係の解説書(東洋経済新報社、SPSS統計パッケージ、I&II)に掲載されているものの全部ではない。だから実際の利用にあたってはセンターの方に問い合わせからのほうが無難である。筆者自身も今後、ひととおり、ありそうなものを使って確かめたく思う。

さて、まず今回のミニ分析例の問題意識をまとめておこう。企業はその事業活動においてさまざまな資源を投入し、あるいは利用している。それは通常、物的資源、人的資源、そして資金という3つに分類されることが多いが、その他にも「情報的経営資源」と称されるものも、たびたび加えられている。その内容としては、たとえば顧客の信用、ブランドの知名度、技術的なノウハウ、流通チャネルの支配力、従業員のモラルの高さなど、いわばモノ、ヒト、カネのどれにも入らないも

\*学習院大学経済学部専任講師

のの総称ともいえるものがあげられている。ただ、これらに共通しているのは、それらが皆、一種の「潜在的成長力」を表わすものだというのである。しかも、それらすべては定性的なものであり、精密な数値で表わすことはできないものと言えるだろう。

このような、「経営資源」とよばれる諸資源をどのように組み合わせて利用すれば企業の成長をより促進することができるか、という問題は、経営戦略論とよばれる幅広い研究分野の中心テーマとされている。そこでは組織面、流通面など数多くの方向からのアプローチが可能であるが、本稿ではごく単純に、財務的側面、とくに研究開発などへの投資と企業成長との関係を、簡単なデータを利用して、みることにする。本来の目的がSPSSの使用例の提示なのであるから、サンプル数も少くてもかまわないであろう。

検討の対象とした産業は医薬品産業である。医薬品産業は、わが国経済が高度成長期にあった頃はもちろんのこと、その後の石油ショックに対してもほとんど何事もなかったような増収増益を示し続けて来ており、とくに注目してよい産業と思われる。また、その扱っている商品が、とりわけ技術的な側面あるいは研究の成果がはっきりと現われるものであるために、研究開発投資への重点のおき方の程度が企業の伸びに強く反映される産業だと言われている。また、当然のことながら、広告宣伝、販売促進等への投資も重要となる。

まず医薬品産業から最近5年間の売上高、総資産、そして経常利益の伸び率を基準として、高成長企業と低成長企業のグループをそれぞれ7社ずつで作った。この手続きはSPSSではなく、FACOM-CLUSTERを利用した。SPSSのCLUSTERは一昨年8月、日本で使われるSPSSが第7版から第8版に改変されるにあたって廃棄されて、現在は存在しないからである。

このようにして選ばれた高成長企業と低成長企業は、いわば機械的に取り扱われたものであると言える。つまりここ数年の実績値をそのままとりあげ、機械的な算法にしたがって似たもの同士（この場合は3種類の成長度指標を用いた3次元でのグループ化）を単純にまとめあげたものにすぎないのである。それは言い換えれば「結果による分類」にあたるものである。これに対して、そのような高成長企業と低成長企業それぞれのグループができることとなった背景には、何らかの理論的要因があったはずである。それを明らかにすることは、さきに述べた経営戦略論の課題の一つであり、ここではその最も単純な、財務的側面を表わす研究開発費、広告宣伝費、販売促進費それぞれの絶対額、対売上高比率、そして最近5年間の伸び率をとりあげて、両グループの差異をどの程度説明しうるかを調べてみる。そしてその他に労働装備率、1人当売上高という指標も加え、11個の変数により、変数選択を伴う判別分析を行ってみようというのが本稿の最終目標である。これは、「結果による分類」の形成要因に、「原因による説明」を加えようとする最も単純な試みである。但し、その際の分析作業が「理論なき計測」になってしまわないた

めに、その背景として経営戦略論の体系が必要なわけである。

〔 2 〕 SPSSによる分析手順

まず、高成長企業グループと低成長企業グループそれぞれ7社ずつから成るデータ・ファイルを作る必要がある。そのプログラムは以下の通りである。

RUN NAME	TEST RUN
PAGE SIZE	NOEJECT
VARIABLE LIST	RD, GRD, RDSL, AD, GAD, ADSL, SP, GSP, SPSL, RLE, SLPE
SUBFILE LIST	HGR(7), MGR(7)
INPUT MEDIUM	CARD
INPUT FORMAT	FREEFIELD
VAR LABELS	RD AMOUNT OF R&D EXPENDITURE /GRD GROWTH RATE OF RD/RDSL PROPORTION OF RD TO SALES/ AD AMOUNT OF ADVERTISEMENT EXPENDITURE/GAD GROWTH RATE OF AD/ADSL PROPORTION OF AD TO SALES/SP AMOUNT OF SALES PROMOTION EXPENDITURE/GSP GROWTH RATE OF SP/SPSL PRO PORTION OF SP TO SALES/RLE RATE OF LABOUR EQUIPMENT/ SLPE SALES PER AN EMPLOYEE
LIST CASES	
READ INPUT DATA	
	データカード
SAVE FILE	IITDRUGS
FINISH	

まずRUN NAME, これは適当な名前をつければよい。ここに記入した文がアウトプットの各頁に見出しとしてプリントされる。次のPAGE SIZEは, NOEJECTとすると, アウトプットの際の無駄な頁替えを避けることができる。いわゆるDENSE PRINTの指

定である。VARIABLE LISTは、使用する変数名を、カードにパンチしてある順に並べてある。その次のSUBFILE LISTは、インプットするカードが複数のグループから成っていることを示すものである。だから、この場合にはデータカード(全部で14ケース)のうち最初の7つのケースがHGRという名の一つのグループであること、そして残る7つのケースがMGRという一群のグループであることを意味している。この2つのサブファイルは、後の判別分析の際に判別対象のグループとなる。さてその次の2枚は入力媒体・書式を示すものである。その後のVAR LABELSのカードは、必須のものではない。これは、後にこのファイルを使って分析をする時、変数名のラベルをつけておけば、変数内容を忘れることもないし、場合によっては計算のたびに変数の横に表形式でラベルが印刷されるからである。だから、もちろん、なくても済むわけだが、できたら今回のように、変数の名付け方そのものはいいかげんであっても何らかの形でつけておけば便利なことが多いのである。

それに続くLIST CASESは、インプットした内容を、確認のためにプリントさせるためのもので、16コラムめからのオプションを省略すると10ケース分プリントされる。

SAVE FILEとそのあとのIITDRUGSは、今回のランでインプットされたデータをIITDRUGSというファイル名で保存せよ、という意味である。ファイル名も、自分でわかりやすくさえあれば、適当につけてしまえばよい。

これでファイルはでき上がりである。次にいよいよ判別分析を始める。プログラムは以下の通りである。

```

RUN NAME          TEST RUN
PAGE SIZE         NOEJECT
GET FILE          IITDRUGS
RUN SUBFILES      (HGR, MGR)
DISCRIMINANT     GROUPS=SUBFILES/VARIABLES=
                  RD TO SLPE/ANALYSIS=RD TO
                  SLPE/METHOD=MAHAL
OPTIONS           5, 6, 7, 10, 11, 12
STATISTICS        1, 2, 3, 4, 5, 6, 9
FINISH

```

このような非常に短いプログラムから、実にたくさんの情報が得られる。

GET FILEは、さき作成したファイルを取り出し、システムに引き渡させるコマンドである。本学のSPSSの便利なところは、さきのSAVE FILEにしても、このGET FILEにしても、その他の特別な補助コントロール・カードがいらないことである。すなわち、

通常、たとえば日立のマシンを利用する場合には次の例のようなカードをカード・デッキの先頭ジョブ・カードの次につけなくてはならない。まずGET FILEの場合には

```
>>SPSSF FT03F001, FN(IITDRUGS.DATA)
```

そしてSAVE FILEの場合には

```
>>SPSSF FT04F001, FN(IITDRUGS.DATA)
```

つまり、ランの内容によってジョブ制御文をつけかえ、入出力用のファイル定義をその都度行わなくてはならないのである。これに対し本学のSPSSでは、システム・ファイルを利用する、しないにかかわらず、全くこの種のカードがいらないのであり、私のようなミスが多いユーザーにとってはもちろんのこと、一般のユーザーにとってもまことに便利なものと言えるであろう。

次のRUN SUBFILESは、すでに作ってあるサブファイル構造をそのまま利用することを意味している。このカードを抜くか、あるいはRUN SUBFILES ALLとすると、14のケース全体を一括して扱うことになり、サブファイル構造は無視される。

その後のカードは、判別分析を行うための手続カードである。VARIABLES=以下の変数はこの判別分析に用いる変数の候補、ANALYSIS=以下の変数は、実際にとりあげる変数のリストである。METHOD=は、変数選択を行うときの選択法を指定するキーワードである。変数選択とは、判別関数の係数算出に先立って判別に有効な変数を候補変数の中から段階的に選ぶということで、指定した選択基準に従って、最適なものから順に変数一つ一つが追加されていくことになる。この場合のMAHALという指定は2グループの間のマハラノビス汎離 $D^2$ を意味している。すなわち、すでに存在しているP個の判別変数に新たな変数一つを追加したと仮定して、新たな $D^2$ の値を他のすべての変数、すなわちグループの全組合せについて算出し、 $D^2$ の最小値を求める。そして候補変数間で $D^2$ の最小値を比較して、その値が最も大きい変数を採択する、というものである。これにより、グループ間のマハラノビス距離が最大になるような変数の組合せを得ることができる。

OPTIONSとSTATISTICSは、それぞれ、出力したい図表や統計量を指定するものである。以下、アウトプットをならべてみよう。表1は最も単純な基本統計量である。これと表2(群内相関行列)は他の統計サブプログラム(三菱のMSL, 富士通のMULVAなど)でも、通常はアウトプットされるものであるが、その後が続いているアウトプットは、SPSSに特有のものと思われるユニークなものである。表3はSTATISTICSの6によるもので、両グループ間で平均が等しいとする仮説を個別の変数について検定するためのF値である。この場合はSPSLとRDSL以外は両グループ間に有意な差が存在していると断言できそうである。その後のアウトプット・リストはそれぞれたどってゆけば、説明なしでもかなり読めると思う。すなわちさきに

MEANS

	GROUP 1 SUBFILE HGR	GROUP 2 SUBFILE MGR	TOTAL
RD	13152.7109	5703.7109	9428.2109
GRD	1.9811	1.3536	1,6674
RDSL	7.8773	6.9287	7,4030
AD	4487.0000	1216.5713	2851.7856
GAD	1.6576	1.1447	1,4011
ADSL	3.1070	1.4461	2,2766
SP	9950.0000	3715.5713	6832,7852
GSP	1.4377	1.0767	1,2572
SPSL	6.5194	5.3480	5,9337
RLE	5046.2852	3850.7141	4448,5000
SLPE	28772.1406	18789.8555	23781,0000

STANDARD DEVIATIONS

	GROUP 1 SUBFILE HGR	GROUP 2 SUBFILE MGR	TOTAL
RD	7391.4531	4107.9102	6924.0742
GRD	.5216	.1705	.4950
RDSL	1.8346	2.3686	2,0940
AD	1986.3457	1078.1575	2288.4807
GAD	.3303	.4241	.4518
ADSL	1.5726	.9398	1,5139
SP	5403.6289	1428.5676	4988,2813
GSP	.4181	.1851	.3628
SPSL	3.4112	2.1123	2,7927
RLE	952.2070	1509,3723	1361,9026
SLPE	4596.5742	2847.4456	6349,8320

表 1

表2

WITHIN GROUPS CORRELATION MATRIX

	RD	GRD	RDSL	AD	GAD	AUSL	SP	GSP	SPSL	RLE
RD	1.0000									
GRD	-.4912	1.0000								
RDSL	.0510	.3870	1.0000							
AD	.7505	-.5868	-.1096	1.0000						
GAD	.3359	-.0718	-.1393	.3534	1.0000					
AUSL	-.4501	.0872	-.0037	.1777	.0180	1.0000				
SP	.7582	-.4842	-.2084	.7402	.4473	-.1123	1.0000			
GSP	.4677	-.7592	-.2267	.5914	.4857	-.0040	.5934	1.0000		
SPSL	-.5199	.1526	.0318	-.0586	.0870	.7511	.0566	.1171	1.0000	
RLE	-.3327	.5127	-.2848	-.3467	-.1057	-.0858	-.2252	-.4116	-.0464	1.0000
SLPE	.5658	-.5217	-.3805	.5893	.4882	-.1760	.5508	.3855	-.2507	-.4353

SLPE

SLPE 1.0000

表3

WILKS' LAMBDA (U-STATISTIC) AND UNIVARIATE F-RATIO WITH 1 AND 12 DEGREES OF FREEDOM

VARIABLE	WILKS' LAMBDA	F	F PROBABILITY
RD	.6884	5.4317	.0380
GRD	.5672	9.1578	.0038
RDSL	.9448	.7017	.5690
AD	.4502	14.6574	.0001
GAD	.6531	6.3736	.0041
AUSL	.6759	5.7531	.0050
SP	.5795	8.7092	.0007
GSP	.7333	4.3637	.0113
SPSL	.9526	.5967	.7777
RLE	.7925	3.1417	.0321
SLPE	.3346	23.8599	.0000

表4

PREDICTION RESULTS -

ACTUAL GROUP	NO. OF CASES	PREDICTED GROUP MEMBERSHIP	
		GP, 1	GP, 2
GROUP 1	7.	7.	0.
SUBFILE MGR		100.0%	.0%
GROUP 2	7.	0.	7.
SUBFILE MGR		.0%	100.0%

PERCENT OF "GROUPED" CASES CORRECTLY CLASSIFIED: 100.00%

説明したマハラノビス距離を基準とした変数選択のステップ1とステップ4（ここではこれが最終ステップとなっている）、そして要約表である。未使用変数の中にF値が1以上のものがなくなり、演算がそこでストップしている。これによってSLPE, GRD, GSP, GAD4つが、両グループの判別に最も有効な変数群として抽出された。そして、以下の分析は、これら4つの変数のみによる関数を用いて行われている。標準化されていない判別係数、標準化後の判別係数、判別スコア、そしてテリトリー・マップなど、SPSSのアウトプットは少くとも私にとっては、他のシステムのものよりも格段に見やすいものに思える。



----- DISCRIMINANT ANALYSIS -----

ANALYSIS NUMBER 1  
 F FOR INCLUSION 1.00000  
 F FOR DELETION 1.00000  
 TOLERANCE LEVEL .00100

MAXIMUM STEPS 22

SOLUTION METHOD - STEPWISE.  
 SELECT THE VARIABLE WHICH WILL  
 MAXIMIZE MINIMUM MAHALANOBIS DISTANCE BETWEEN GROUP PAIRS.

PRIOR PROBABILITIES -

GROUP	1	GROUP	2
SUBFILE		SUBFILE	
MGR		MGR	
	.50000		.50000

VAR ENTERED ON STEP 1 SLPE

			DEGREES OF FREEDOM	SIG.
WILKS LAMBDA=	.3346356	APPROX F=	23.8598	1 12.00 .000
RAD'S V	= 23.85988	V CHANGE=	23.8599	1 .000

F MATRIX - DEGREES OF FREEDOM: 1, 12

GROUP	2	GROUP	1
		23.85989	

----- VARIABLES IN THE ANALYSIS -----

VARIABLE	ENTRY CRITERION	F TO REMOVE
SLPE	6.81712	23.85986

----- VARIABLES NOT IN THE ANALYSIS -----

VARIABLE	TOLERANCE	F TO ENTER	ENTRY CRITERION
RU	.67981	.08480	1.55192
GRD	.72787	13.09548	2.61614
RUSL	.85525	2.60712	.20049
AD	.65274	.42419	4.18783
GAC	.76170	.00815	1.82091
ADSL	.96902	3.36079	1.64370
SP	.69663	.02991	2.48835
GSP	.85140	.01546	1.24664
SPSL	.93718	1.30543	.17049
RLE	.81048	5.75350	.89761

VAR ENTERED ON STEP 4 GAD

DEGREES OF FREEDOM SIG.  
 WILKS LAMBDA = .4654815E-01 APPROX F = 46.0870 4 9.00 .000  
 RAO'S V = 245.7668 V CHANGE = 131.8199 1 .000

F MATRIX - DEGREES OF FREEDOM: 4, 9

GROUP 1  
 GROUP 2 46.08131

----- VARIABLES IN THE ANALYSIS -----

----- VARIABLES NOT IN THE ANALYSIS -----

VARIABLE	ENTRY CRITERION	F TO REMOVE	VARIABLE	TOLERANCE	F TO ENTER	ENTRY CRITERION
GRD	19.01355	55.59660	RD	.60361	.00136	32.55647
GAD	70.21916	9.41986	RDSL	.77919	.03527	32.67256
GSP	32.55829	19.72508	AD	.49094	.07340	33.86679
SLPE	6.81712	23.26019	ADSL	.95346	.42552	34.65639
			SP	.51797	.04767	32.80104
			SPSL	.80653	.07318	32.68689
			RLE	.68301	.50441	34.15466

F LEVEL INSUFFICIENT FOR FURTHER COMPUTATION  
 FILE ITDRUGS (CREATION DATE: 10/09/81)  
 SUBFILE: MGR MGR

----- D I S C R I M I N A N T A N A L Y S I S -----

SUMMARY TABLE

STEP NUMBER	VARIABLE ENTERED	VARIABLE REMOVED	F TO ENTER OR REMOVE	NUMBER INCLUDED	WILKS' LAMBDA	SIG.	RAO'S V	CHANGE IN RAO'S V	SIG. OF CHANGE
1	SLPE		23.85988	1	.33464	.000	23.85988	23.85988	.000
2	GRD		13.09548	2	.15277	.000	66.54733	42.68745	.000
3	GSP		6.03549	3	.09527	.000	113.94691	47.39958	.000
4	GAD		9.41987	4	.04655	.000	245.76678	131.81987	.000

CLASSIFICATION FUNCTION COEFFICIENTS

	GROUP 1 SUBFILE MGR	GROUP 2 SUBFILE MGR
GRD	177.14960	123.24663
GAD	-103.30838	-71.83977
GSP	187.40353	131.67529
SLPE	.01025	.00700
CONSTANT	-372.04517	-178.98315

DISCRIM. FUNCTION	EIGEN VALUE	REL. %	CANN CORR	FUNCTIONS DERIVED	WILKS' LAMBDA	CHI SQUARE	DF	SIG.
1	20.4821100	.00	.9764	0	.04655	30.672	4	.000

REMAINING COMPUTATIONS WILL BE BASED ON 1 DISCRIMINANT FUNCTION(S)

STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

	FUNC 1
GRD	.71503
GAD	-.38104
GSP	.54175
SLPE	.55253

UNSTANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

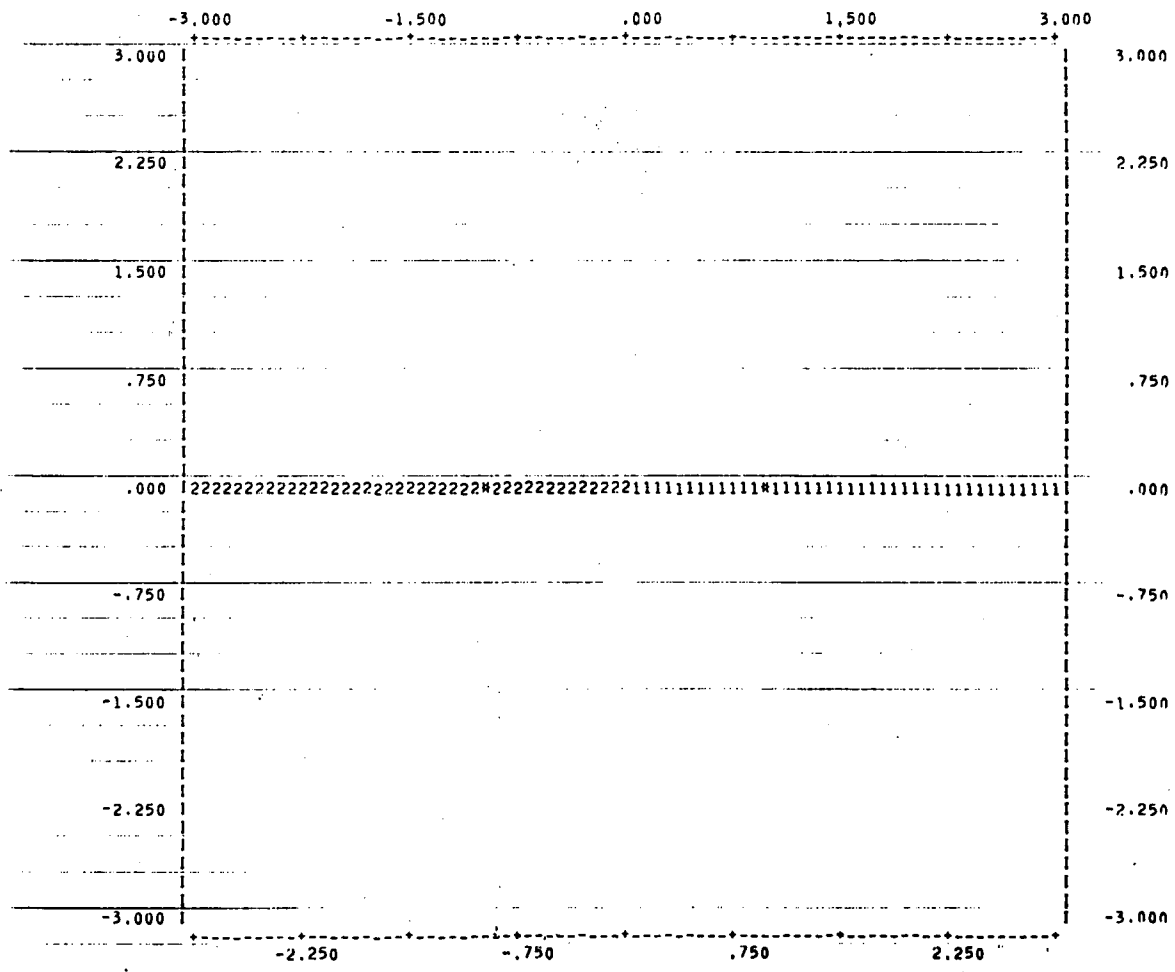
	FUNC 1
GRD	1.44453
GAD	-.84332
GSP	1.49345
SLPE	.00009
CONSTANT	-5.17381

CENTROIDS OF GROUPS IN REDUCED SPACE

GROUP	FUNC 1
GROUP 1	.94090
SUBFILE HGR	
GROUP 2	-.94089
SUBFILE HGR	

CASE SUBFIL	SEQNUM	MISSING VALUES	ACTUAL GROUP	HIGHEST PROBABILITY GROUP	P(X/G)	P(G/X)	2ND HIGHEST GROUP	P(G/X)	DISCRIMINANT SCORES FUNC 1
HGR	1.	0	1	1	1.000	1.000			.926
HGR	2.	0	1	1	.861	1.000			1.197
HGR	3.	0	1	1	.988	1.000			1.069
HGR	4.	0	1	1	.188	1.000			.384
HGR	5.	0	1	1	1.000	1.000			.965
HGR	6.	0	1	1	.789	1.000			1.234
HGR	7.	0	1	1	.987	1.000			.811
HGR	1.	0	2	2	1.000	1.000			-.912
HGR	2.	0	2	2	1.000	1.000			-.949
HGR	3.	0	2	2	.999	1.000			-.881
HGR	4.	0	2	2	.861	1.000			-.684
HGR	5.	0	2	2	.997	1.000			-1.028
HGR	6.	0	2	2	.994	1.000			-1.049
HGR	7.	0	2	2	.983	1.000			-1.081

TERRITORIAL MAP OF DISCRIMINANT SCORE 1 (HORIZONTAL) VS. DISCRIMINANT SCORE 2 (VERTICAL). \* INDICATES A GROUP CENTROID.



### 〔 3 〕 小 結

結果を簡単にまとめよう。最終的な判別関数に残った変数は、それぞれの支出の伸び率と、生産性指標たる1人当売上高であった。この5年間で研究開発、広告宣伝、販売促進のそれぞれにどれくらい重点を移したか、ということが同じ時期の企業全体の成長を規定する、というのは、きわめて常識的ではあるが、GADすなわち広告宣伝費の伸び率の係数符号が負になってしまっているのはうなずけない。しいていえば、今回のサンプル企業には大衆医薬品メーカーが入っておらず、もっぱら医家向中心の会社が多いことから、あまり広告宣伝に力を入れても、それに比例した効果は出にくい、ということであろうか。たしかに標準化後の判別係数をみるとGRD、すなわち研究開発費の伸び率の係数が最も大きく、研究開発の影響度が大きそうだということで、一応の説明はつく。

以上のとおり、SPSSは、独特のファイル構造、分析の多彩さ、そしてアウトプットの見易さなど、とくに社会科学系の研究者にとっては、その名の通り、おおいに推奨されるものである。ただし、もちろん、有用な分析になるかどうかは本人の、解釈のしかたによるものではあるが。