

---

---

**研究ノート**

---

---

## 日赤・健康管理センターに蓄積した検診データを 活性化する情報システムの研究と開発

～判別分析結果の関連分析～

[中間報告その5]

健康増進研究会 : 野村 武○ 小山 和作 川島 英敏  
山崎 明治 井上 靖士 波多江 修

〈目次〉

1. いきさつ
2. 判別分析結果のまとめ
3. 不良データと想定されるデータを除いて再分析した結果
4. 優良データに絞って再分析した結果
5. データ項目数を削減する試行
6. データと計算の仕組みを「一つの判別システム」と捉えて
7. 考察

**Research on activation of healthcare examination results  
accumulated by JRC Kumamoto Healthcare Center  
: Pretest by Discriminant Analytical Method**

Study team ;

Takeshi NOMURA (Prefectural University of Kumamoto)

Wasaku KOYAMA (Japanese Red Cross Healthcare Center)

Hidetoshi KAWASHIMA (JRC Healthcare Center)

Akitaka YAMASAKI (Fukuoka Institute of Technology, Junior College)

Yasuto INOUE (Fukuoka Institute of Technology, Junior College)

Osamu HATAE (Nishinippon Newspaper Co.)

Abstract ;

Japanese Red Cross Kumamoto Healthcare Center has been accumulating huge data of health examinations for the 25 years. Researches designed to develop information systems concerning these data has been done by the team of JRC Kumamoto Healthcare Center, Prefectural University of Kumamoto and Fukuoka Institute of Technology, Junior college.

This collaboration team consist of members of different walks of life such as medical staff, statisticians and system engineers.

For the past three years, JRCKHC has been trying to add the cause of patients' deaths to the examination results accumulated before.

To investigate the relationship between the health examination results and the causes of death, our team tried to judge the relationship between the healthcare examination data before and the cause of death, applying the discriminant analysis.

This is the report of pre-experiments with small data.

## 1. これまでの経過

健康増進研究会は、予防医学、情報システム学、統計数学の専門家で構成している。長年にわたり、日赤熊本健康管理センターが蓄積してきた検診データを使って、統計学的な手法を中心にした学際的な研究を続けてきた。

詳細な内容は、それぞれの中間報告にゆずるとして、簡潔に研究の経過をまとめると次のとおりである。

初めに、汎用プログラムの「三次元集計表」を、21項目の数値データの中から三項目を抜き出して適用してみた。プログラムの扱いに慣れる意味もあった。が、分かりやすい表を出力できたというに止どまった。

次いで、熊本県内の地域細分化を人口動態のデータから算出した。熊本市への傾斜が強い中で、県内を8地域に細分する試案をまとめた。センターの健康管理データは県内各地に一様に分散していると聞いており、いつかは役立てたいと準備したわけである。

さらに、因子分析法も適用した。野村や波多江は、学生の進路分析、市町村の住みよさ調査などで同法を活用した経験を持っており、この分析では健康診断の21項目のデータに適用した。その簡便法といわれるデータマイニング法も使用したが、これらの結果は「際立った因子群」を抽出するに至らなかった。

最後に判別分析法を適用した。センター側で検診データに死因を結び付けたらとの発想があり、苦労して数百のデータが準備された。数学的にはデータに死因という「外的基準」が付けられたことになり、分析法も判別分析が適用できることになった。

データ数に限りがあったため、分析精度の面で疑問が出てきたが、極めて興味深い結果が生まれてきたために、予備調査であることを強調しながら、いくつかの角度から傾向をつかむ努力をした。

この後に、データを補強した本格分析を予定している。

## 2. 判別分析結果のまとめ

前回の報告書に判別分析結果のまとめを記述したが、以下にも総括した。

### (1) 表－1の読み方

一連の判別分析の中核をなす結果は、前回報告書の〔別表－1〕から〔別表－4〕に表示している。今回の報告書には、さきの4表をまとめて〔表－1〕として載せている。死因をもとに心臓血管系と脳部血管系とがん腫瘍系の三つのグループと、いまひとつ心臓・脳部血管系とがん脳部血管系の二つに大別した場合の分析結果である。つまり、比較する相手を変えて判別分析を行ったわけである。

各データには、分析で出てきた係数群を乗じた総和＝判別得点が表示されており、この得点の符号と絶対値とで判別結果が分かる。

〔表－1〕判別分析の4表の結果をまとめた表

分類 コード	死 因 表 現	判 別		得 点	
		別表－1	別表－2	別表－3	別表－4
1-1	急性心不全(拡張心筋)	☆	2.591	5.497	4.447
1-2	急性心筋梗塞	☆	0.857	1.869	1.076
1-3	冠動脈血栓・腹動脈瘤	☆	0.749	3.298	2.852
1-4	大動脈弁閉鎖不全		-1.452A	1.930	2.067
1-5	心筋梗塞		1.157	0.348	-0.067B
1-6	心筋梗塞		0.056	0.804	1.011
1-7	急性心筋梗塞		1.447	0.756	-0.030B
1-8	心不全、心筋梗塞		-0.655B	1.313	1.038
1-9	急性心不全、心筋梗塞	☆	2.385	3.117	2.330
1-10	心筋梗塞		0.220	0.691	0.522
1-11	心筋梗塞脳梗塞肺炎		0.735	0.734	0.264
1-12	虚血性心疾患心筋梗塞		0.159	3.223	2.587
1-13	急性心筋梗塞	×	-0.188B	0.163	-0.140B
1-14	心筋梗塞	☆	1.914	2.543	1.816
1-15	転倒でのショック死		-0.981A	1.451	1.628
1-16	心筋梗塞		0.010	2.688	1.385
2-17	心不全		0.802	2.932	2.061
2-18	不整脈による心不全	☆	2.317	1.810	1.187
2-19	急性心不全		0.270	3.898	3.666
2-20	急性心臓死		0.713	0.746	-0.103A
2-21	心不全	☆	2.514	2.223	1.561
2-22	急性心不全		0.918	-0.107B	-0.648B
2-23	心不全(海水浴中)		-0.023A	1.010	1.181
2-24	急性心不全	×	-0.739B	-0.729B	-0.775B
2-25	心不全(心房細動)		3.677	1.432	0.104
2-26	急性心不全		1.394	0.169	-0.365
2-27	急性心不全	×	-2.588B	-2.365B	-1.291B
2-28	急性心不全腎不全		0.256	6.166	4.604
2-29	急性心不全		0.841	0.136	-0.238B

2-30	心不全		3.588		1.374	0.186
2-31	心不全	☆	2.261		0.834	2.261
2-32	心臓発作(急死)		-0.651A		0.910	0.925
2-33	不整脈による突然死	☆	0.614		2.635	1.303
2-34	急性心不全	☆	3.114		1.339	0.498
2-35	?	×	0.126B		0.156B	1.161B
<hr/>						
3-36	脳血管障害(脳卒中)		-1.701	4.025		2.047
3-37	くも膜下出血	×	-0.312	-1.262B		-1.199B
3-38	小脳出血		-3.955	2.349		0.872
3-39	くも膜下出血		-0.696	0.949		-0.297B
3-40	脳出血		0.209B	3.793		4.607
3-41	脳出血	☆	-2.735	2.991		1.825
3-42	脳内出血	☆	-2.125	2.182		1.747
3-43	くも膜下出血		-1.840	1.337		-0.184B
3-44	くも膜下出血	×	-1.381	-2.290B		-3.277B
3-45	くも膜下出血		-1.657	0.622		0.032
3-46	脳出血	☆	-2.943	1.419		0.598
3-47	くも膜下出血		-2.481	-0.168B		-0.695B
3-48	小脳出血	☆	-0.825	1.396		0.953
<hr/>						
4-49	脳梗塞、風邪、老衰		1.027A	2.132		1.854
4-50	脳梗塞	☆	-1.475	2.466		1.393
4-51	脳梗塞、肺炎		0.275A	2.429		1.406
4-52	脳梗塞、肺炎		0.566A	0.712		-0.553B
4-53	心不全、脳梗塞	☆	-2.554	0.931		0.539
4-54	脳梗塞	☆	-1.758	5.076		4.316
4-55	脳血栓	☆	-1.620	2.135		1.795
4-56	脳梗塞	☆	-2.239	2.428		1.398
4-57	間質肺炎胆石脳梗塞		1.396A	1.145		0.449
4-58	脳梗塞		-0.390	2.016		2.347
4-59	脳梗塞、急性心不全		0.053A	1.571		1.030
<hr/>						
5-60	肝臓がん肺がん			0.598B	-2.345	-1.347
5-61	膵臓がん肝不全	☆		-0.975	-1.046	-1.971
5-62	肝がんB型肝炎腎不全心不全			1.105B	-0.076	0.028B
5-63	肝がん肝硬変(慢性肝炎)	☆		-2.531	-2.074	-2.752
5-64	肝臓がん(多臓器転移)	☆		-1.329	-2.020	-1.829
5-65	原発性肝がん・腹膜炎イレウス	☆		-2.871	-0.928	-1.661
5-66	肝臓がん			0.053B	-1.630	-1.559
5-67	肝腫瘍慢性肝炎(B型)	☆		-1.153	-3.404	-2.365
5-68	肝臓がん	×		2.392B	0.025B	0.165B
5-69	肝腫瘍慢性肝炎			1.762B	-0.330	0.024B
5-70	肝腫瘍慢性肝炎	☆		-4.565	-6.162	-6.471
5-71	原発性肝臓がん	☆		-2.858	-3.023	-3.695
<hr/>						
6-72	膵臓がん			-0.470	0.153B	-0.150
6-73	膵臓がん			-0.295	-0.122	-0.333
6-74	膵臓がん肝転移	×		0.138B	0.455B	0.257B
6-75	膵臓がん	☆		-2.345	-0.785	-1.788
6-76	膵臓がん大量出血でショック死	☆		-2.116	-2.997	-2.849

6-77	膵臓がん	☆	-2.838	-3.560	-4.572
6-78	膵臓がん	☆	-3.089	-2.396	-2.384
6-79	膵臓がんにて		-0.783	0.338B	-0.161
6-80	膵臓がん	×	0.719B	0.306B	0.286B
6-81	膵臓がん肝転移		0.401B	-0.639	-0.307
6-82	膵臓がん	☆	-2.627	-5.109	-4.024
6-83	膵臓がんにて	☆	-2.459	-0.787	-1.881
<hr/>					
7-84	肺がん(腺癌)大腿骨転移	×	0.660B	0.295B	0.546B
7-85	肺がん呼吸不全		-0.316	-1.852	-1.351
7-86	肺がん呼吸不全		-0.457	0.656B	-0.088
7-87	肺がん	☆	-1.120	-2.185	-1.987
7-88	肺がん(腺癌)	×	0.977B	1.808B	1.557B
7-89	肺がん	☆	-3.863	-5.423	-5.025
7-90	肺がん	☆	-1.241	-2.726	-2.213
7-91	肺がん脳症	×	1.113A	0.227B	0.419B
7-92	肺がん膀胱がん術後(4年)		-1.424	0.318B	-0.390
7-93	肺がん(左肺上葉)	☆	-0.612	-2.842	-2.338
7-94	肺がん	☆	-2.885	-3.310	-3.020
7-95	肺がん	☆	-3.290	-2.461	-3.424
7-96	肺がん	☆	-2.414	-1.265	-2.616
7-97	肺がん	×	1.457B	-0.118	0.321B
7-98	脳腫瘍がん性胸膜炎呼吸不全	×	-0.326	0.832B	0.216B

判定コード A：異常で当然か

B：なぜ異常となったか

×：三つの判別得点がいずれも異常で、除外候補

☆：三つの判別得点が優良なデータ(符号を含め0.5以上)

[表-1]の中の「別表-1の部分」で簡単に解説する。分類コードの1-、2-は心臓血管系である。その中でデータ1-1は心不全・心筋梗塞が死因で、+2.59の得点は妥当な結果である。1-4は弁膜症が死因で、-1.45の得点はこのグループではないと指摘しており、これも妥当な判別であろう。また、1-8は心不全・心筋梗塞が死因だが、得点は-0.66と出ていてグループではないという判定ととらえている。

下部の3-、4-は脳血管系である。別表-1の分析では、結果が負に出れば妥当となる。従って3-40、4-49などの+得点は、このグループに入れることは妥当でないと判定されたことになる。

## (2) 妥当率

判別分析の結果を見て、妥当な結果が出たと判断される率=妥当率を集計す

ると、[表-2] から [表-4] となった。全体的に60%以上で、平均して70～80%になったことは大きな収穫である。

[表-2] 心血管系：脳血管系

グループ	データ	異常当然	なぜ異常	妥当率	除当然妥当率	妥当率	除当然妥当率
1	16	2	2	12/16=0.75	12/14=0.86		
2	19	2	3	14/19=0.74	14/17=0.82	23/35=0.74	26/31=0.84
3	13	0	1	12/13=0.92	12/13=0.92		
4	11	5	0	6/11=0.54	6/6=1.00	18/24=0.75	18/19=0.95

[表-3] 脳血管系：悪性腫瘍系

グループ	データ	異常当然	なぜ異常	妥当率	除当然妥当率	妥当率	除当然妥当率
3	13	0	3	10/13=0.77	10/13=0.77		
4	11	0	0	11/11=1.00	11/11=1.00	21/24=0.88	21/24=0.88
5	12	0	5	7/12=0.58	7/12=0.58		
6	12	0	3	9/12=0.75	9/12=0.75	27/39=0.69	27/38=0.71
7	15	1	3	11/15=0.73	11/15=0.79		

[表-4] 心血管系：悪性腫瘍系

グループ	データ	異常当然	なぜ異常	妥当率	除当然妥当率	妥当率	除当然妥当率
1	16	0	0	16/16=1.00	=1.00		
2	19	0	4	15/19=0.79	=0.79	31/35=0.89	31/35=0.89
5	12	0	1	11/12=0.92	=0.92		
6	12	0	4	8/12=0.67	=0.67	28/39=0.71	28/39=0.71
7	15	0	6	9/15=0.60	=0.60		

[表-5] 心・脳血管系：悪性腫瘍系

グループ	データ	異常当然	なぜ異常	妥当率	除当然妥当率	妥当率	除当然妥当率
1	16	0	3	13/16=0.81	=0.81		
2	19	1	6	12/19=0.63	12/18=0.67	43/59=0.73	43/58=0.74
3	13	0	5	8/13=0.62	=0.62		
4	11	0	1	10/11=0.91	=0.91		
5	12	0	3	9/12=0.75	=0.75		
6	12	0	2	10/12=0.83	=0.83	29/39=0.74	29/39=0.74
7	15	0	5	10/15=0.67	=0.67		

ただし、ここでの妥当との判断は、あくまでもデータに付加された死因を正しいものとしてのわれわれの判断であり、この後に述べる疑わしいものや紛らわしいものについては、可能な範囲での医学的なフォローアップが望まれる。

報告の中で、死因表現という言葉を使っているのはこの辺りを考えたからである。中身を見ると、死因表現どおりに明瞭に判別されたデータもあれば、反対に疑問に思われるデータもあり、医学データの難しさ、複雑さが思いやられる。

今回は予防医学の面からの本格的な吟味を前に、データ処理の観点から常識的に考えられるいくつかの関連分析を行ったので、順に報告する。

### 3. 不良と想定されるデータを除いて再分析した結果

[表-1]の中で、判別得点に特徴があるものを抽出した。

(1) まず、疑問だと思われるデータに目を付け、表の中に

A：死因表現から見て、異常と出て当然と思われるもの

B：死因表現と違った判別がなぜか出たと思われるもの

というサインを付けた。

そして、判別分析の妥当率＝妥当な判別ができた割合＝を計算し、加えて「除当然妥当率」＝Aを除いた妥当率＝除いて当然の処理後の妥当率を計算して、前出の[表-2]から[表-5]までに付け加えた。

(2) [表-1]で次に吟味したことは、異常データの共通性である。それぞれの分析を行う度に異常と表示されるデータは変化しているが、中にはどの分析をやってもおかしい結果を示したデータがある。このいつも異常を示すデータを野村の判断でピックアップしたところ、14件となった。表の中に×印で表示している。

この14件を除いて、再度判別分析を行った。その結果各データの判別得点は変化した。その状況の一例として、心臓血管系－脳血管系の場合の変化を示すと[表-6]のとおりである。



[表-6] 判別得点の変化：心系-脳系  
 ～異常データ14件を除いて再分析～

No.	当初	改善	No.	当初	改善
1-1	2.591	3.307	2-31	2.261	3.724
1-2	0.875	1.271	2-32	-0.651	0.666
1-3	0.749	1.391	2-33	0.614	-0.439
1-4	-1.452	-0.775	2-34	3.114	5.248
1-5	1.157	2.743	2-35	0.126	—
1-6	0.056	-0.725	3-36	-1.701	-5.635
1-7	1.447	0.899	3-37	-0.32	—
1-8	-0.655	-0.655	3-38	-3.955	-6.182
1-9	2.385	2.647	3-39	-0.696	-1.993
1-10	0.220	-0.748	3-40	0.209	0.576
1-11	2.489	2.429	3-41	-2.735	-4.622
1-12	0.159	1.119	3-42	-2.125	-2.615
1-13	-1.188	—	3-43	-1.840	-4.621
1-14	1.914	2.010	3-44	-1.381	—
1-15	-0.981	-1.493	3-45	-1.115	-1.418
1-16	0.010	-0.414	3-46	-2.943	-4.710
2-17	0.802	1.060	3-47	-2.481	-3.868
2-18	2.317	2.172	3-48	-0.825	-1.322
2-19	0.270	1.346	4-49	1.027	-0.166
2-20	0.713	1.115	4-50	-1.475	-3.583
2-21	2.514	3.245	4-51	0.275	0.140
2-22	0.918	1.570	4-52	0.566	-1.753
2-23	-0.023	0.483	4-53	-2.554	-2.607
2-24	-1.739	—	4-54	-1.758	-1.800
2-25	3.677	4.487	4-55	-1.620	-1.860
2-26	1.394	2.972	4-56	-2.239	-4.399
2-27	-2.588	—	4-57	1.396	0.582
2-28	0.256	1.195	4-58	-0.390	0.346
2-29	0.841	2.622	4-59	0.053	0.997
2-30	3.588	6.344			

さらに、全データについて、判別得点の符号と絶対値の変化を分類して、変化の傾向を分かりやすくしたのが [表-7] である。

[表-7] に見られるとおり、全体の3/4のデータは、判別得点が大幅に

良い方向に変わっており、余り変らないデータを含めると88%にも達している。

[表-7] 不良データを除く前と後の状況の変化

	12/34		34/567		12/567		1234/567		合計	
逆転しよくなった	4	8%	4	8%	1	2%	6	7%	15	6%
大幅によくなった	29	55	40	76	53	86	64	76	186	74
余り変わらず	12	23	2	4	5	8	2	2	21	8
少し悪くなった	1	2	4	8	3	5	3	4	11	4
逆転して悪化した	2	4	2	4	0	0	4	5	8	3
大幅に悪くなった	5	9	1	2	0	0	5	6	11	4
	53		53		62		84		252	

#### 4. 優良データに絞って再分析をした結果

こんどは2. と反対の関連分析である。

[表-1] を眺めて、「妥当な判別が行われていて、加えて判別得点が高いデータ」をピックアップしてみた。いわば、優良群の抽出である。表の中では☆印を付けている。

符号の合致と得点（絶対値）の0.50以上を基準として、野村の判断による抽出を行ったところ、優良データは全体で39件（39.8%）となった。

データ数が不足であること、項目数がデータ数に比べて多すぎることは承知の上で、予備調査の延長線上の試行を行った。

結果は、心系-脳系だけが計算不能となり、残りは一応まともな結果が出ている。判別得点も30~60点ときわめて高い。

全体として、項目数に比してデータ数が不足していると考えられるので、現象のチェックはこんご行うことにして、ここでは実験を打ち切った。

#### 5. データ項目数を削減する試行

なんども記述したように、今回の予備調査では、データ数に比べてデータ数が21項目と多すぎる。21項目のデータにしては、データ数が少なすぎるという方が良さそう。この項目を医学的な判断ではなく、統計学的な判断で削減することを試みた。

[試行-1]

ここでは、判別分析の処理の中で出力される「データ項目ごとの相関比（標準化）」を検討した。つまり、相関比が大きければ、その項目が結果に対して大きな貢献をしていることになり、小さければ、その項目の貢献度は小さいことになる。

四つの分析で、相関比はいろいろと変動するので、それぞれの数値が最下位から下方6位までに位置する回数、下方5位までに位置する回数を調べた。

[表-8]はその結果である。

[表-8] 項目数削減の吟味  
～相関比の低い項目を探す～

	6位迄の回数	5位迄の回数	結論
肥満度	6	5	
一日高	2	2	
一日低	3	1	
血糖空	9 ③	4	⑥
努力肺	4	2	
総コレ	1	1	
HDL	6	6 ④	⑨
中脂肪	2	2	
尿酸値	5	5	
白血球	5	4	
赤血球	15 ①	15 ①	①*
ヘモグ	8 ④	8 ③	③*
血清ア	8 ④	6 ④	⑤*
尿素窒	7 ⑦	6 ④	⑥
クレア	4	3	
TB	4	4	
GOT	7 ⑦	6 ④	⑥
GPT	7 ⑦	3	
γGT	3	3	
LDH	8 ④	8 ③	③*
TP	12 ②	11 ②	②*

\*印の5項目を候補とした

下位に位置することが多い順に項目を並べると、赤血球、TP、ヘモグ、LDH、血清アの5項目となった。

この5項目のデータを外して、判別分析を行った。結果はこれまでとあまり変わらないように見受けられる。

[試行-2]

いまひとつ別の削除方法を試行した。判別分析で出力される「標準化された項目係数」の比較である。判別得点の計算に使われるプール分散共分散の項目係数（標準化）の大小をチェックした。

係数の絶対値0.3以下がいくつあるか、0.5以上がいくつあるかなどの吟味を行い、[表-9]を得た。

[表-9] 項目削減の検討表

～プール分散共分散の係数（標準化）をチェックして～

	12-34	34-567	12-567	1234-567	<0.3	>0.5	判定
肥満	-0.002234	+0.273124	+0.318646	+0.341934	20	00	*
圧高	-0.601357	+0.222851	+0.082192	+0.199310	21	10	△
圧低	-0.207081	+0.987547	+1.161948	+0.924285	10	21	
血糖	+0.488377	+0.211298	+0.585412	+0.271279	10	20	△
肺活	+1.547933	-0.896053	-0.040476	-0.367792	10	20	△
総コ	+0.989480	+0.032398	+0.979690	+0.472486	10	21	
HD	-0.551683	+0.585868	+0.183180	+0.479395	10	21	
中脂	+0.138761	-0.165015	+0.196894	+0.024257	31	00	*
尿酸	-0.0617	+0.808257	+0.899186	+0.818191	10	21	
白血	+0.299841	-0.225383	+0.374483	+0.137736	21	00	*
赤血	+1.600072	+0.275615	+0.892701	+0.802198	10	21	
へモ	-2.197881	-1.143843	-2.186443	-1.681884	00	31	
血ア	-0.162003	+0.052869	+0.306516	+0.188155	21	00	*
尿窒	-0.726662	-0.325498	-0.608367	-0.555143	00	21	
クレ	+0.506222	+0.580587	+0.575099	+0.534256	00	31	
T B	-0.100597	-0.165457	-0.578664	-0.560118	20	11	△
G O	+0.973807	-0.401593	+0.114230	-0.042378	11	10	△
G P	-1.396000	+0.909125	+0.127353	+0.204979	11	20	
γ G	+1.053913	-0.544711	+0.006355	-0.134011	11	20	
L D	+0.769710	-0.686466	-0.658518	-0.660079	00	31	
T P	+0.137547	-0.873545	-0.966722	-0.940403	10	21	

注 \* : 削減する候補 肥満度、中脂肪、白血球、血清ア  
△ : 準候補 この中から血圧高、血糖空を選択した

順位の状態をにらんで、肥満度、中性脂肪、白血球、血清ア、血圧高、血糖空の6項目をピックアップした。

この6項目を外して分析を行った結果は、これまでとあまり変わらないように見受けられる。

[試行-1]、[試行-2]については、結果の詳細な吟味はこれからである。

## 6. データと計算仕組みを「一つの判定システム」と捉えて

ここに判別分析で得られた「四つの項目係数群（+定数）」がある。この係数群に、新規の死因未知データを乗じて、総和を算出すると、それは未知のデータの判別得点となる。

$$\text{判別得点} = \Sigma (\text{項目データ} \times \text{項目係数}) + \text{定数}$$

この一式を「一つの死因判定システム」と捉えて、つぎのような試行を行った。新しいシステム考案へ向けての最初の模索である。

現在、不完全ながら出現しているこの「判定システム」に、未知のデータを投入して、判別得点を求めた。結果は[表-10]のとおり。

[表-10] 判定システムによる計算

未知のデータ 肥満：95 圧高：130 圧低：90 血糖：219 肺活：3000  
 総コ：201 HDL：48 中脂：148 尿酸：90 白血：6000 赤血：450  
 ヘモ：130 血ア：74 尿窒：15 クレ：9 TB：2 GOT：24  
 GPT：19 γG：143 LDH：455 TP：69

	心：脳	脳：腫	心：腫	心脳：腫	判定
得点計算	+6.403	+7.518	+4.132	+5.427	
論理演算	心>脳	脳>腫	心>腫	心脳>腫	⇒心臓系

つまり、予備実験で作られた判定システムに、一つの死因未知のデータを投入して、判別得点の符号および絶対値を論理演算したところ、答えは心臓血管系と出た。

判定システムの一つの案は上述の方法であるが、今一つ現在のデータ群に未知のデータを挿入して、四つの判別分析を行う方法も考えた。試行したが、統

計学的な疑問点が残っており、報告は今後に譲ることにする。

## 7. 考察

前回の報告から、多変量解析法を判別分析法に変えた。検診データに死因を結び付けたデータ約200件が用意できたからである。統計数学的にいえば、数値データに外的基準（死因）が付加されたことで、判別分析が適用できることになった。初めての試行であるため、データは量的な数値に限ることにした。

### [試行のねらい]

最近になって、一連の試行のねらいが明確化してきたような気がする。もちろん、まだ期待のレベルのものでしかなく、こんごどのように進んでいくかは流動できてある。

直接的な目標は次のように考える。

一般に健康な人は、年の経過とともに何らかの病気がもとで亡くなっていく。この健康人の検診データと、その後の当人の死因との間にはどのような関係があるのだろうか。この関係を疫学的な分析によって捉えてみたい。

長期的な目標は、予防医学にいろいろと貢献し得るような研究成果の蓄積である。個性的な状況が強い「人の健康問題」について、数学的なアプローチがどこまで貢献できるのか、まったく未知の研究であるが、可能な限り進んでみたいと考える。

### [強調点]

具体的に強調しておきたいことは、判別分析法は21項目という多項目のデータを同時に分析して、各項目ごとの関わり具合を数字で表してくれる手法ということである。

つまり、人間の能力では不可能なことをコンピューターは迅速に処理してくれるわけで、人間はその結果を読み取り、さらに経験と考察力を加えて、総合的な判断をしていくということになる。

以下には、予防医学の視点を加えた総合的な考察はまだできていないので、取りあえず統計学的、情報処理学の立場からの技術的な考察に限って展開する。

### (1) 大筋の結論

データ数の不足のために、分析作業は正確性を欠いている。一応予備調査と位置付けているが、判別分析の手法は十分に適用できそうである。

これからは、医学全般の視座からの検討を加えていくことが必要で、関係者は強い関心を寄せている。

こんごどのように展開していくか、いろいろな側面が考えられ、健康増進研究会も準備を始めている。

### (2) データに関する課題

データに関しては、①検診データの項目脱落防止の対策②プライバシー保護のチェック③死因情報収集の課題がある。

①と③はセンター内部の課題である。②のプライバシー保護はデータの取扱状況を説明することによって、理解を深める努力をしたい。個人名を削除したデータを取り扱っているが、フォローアップ調査はセンター側の課題となる。

ほかに、分析プログラムの取り決めに従い、脱落のあるデータは棄却することにしたが、会としてはいまひとつの平均値を代入する方法に納得がいかなかったからで、今後は棄却と代入の両面作戦も再検討する。

棄却量に勝る多くのデータが準備されることを望みたいものである。

### (3) 判別分析のプログラムに関する課題

SPSS から購入した多変量解析法のプログラムは、マニュアルの不備からかなり使いづらいところがあった。会の山崎はプログラムを自作し、関連書の具体例をことごとくトレースして、種々のチェックをした。さらには、データ処理上の必要から、プログラムの制作・改変もこまかく対応した。これによって会としては、安心して実験を進めることができた。

前回にも提示した三グループ以上のデータの同時分析（判別分析）も、マスターしておきたいところである。

### (4) データ数の不足に対する対策

一般的に、分析にあたっては、データ数は変数の数より多いのが原則であろう。今回は、当初は7種の死因分類からスタートしたが、変数21に対して死因ごとのデータ数が10～20と少なかった。データ収集の事情から仕方のないこと

だった。

ここで工夫をして、心臓血管系、脳部血管系、がん腫瘍系と大きくまとめる発想をして、実験を前進させた。

本格実験では、初めの考えに戻して、もっと細かく分類して分析を進めたい。

また、本格分析では、数学的な精度をいろいろな角度から吟味していかねばならないと考えている。

#### (5) 判別の結果の70～80%は妥当な結果が出た

今回の予備調査の分析で、死因と判別結果とに妥当な関係が見られる割合＝妥当判別率＝は、最低でも60%以上で、平均すると70～80%になったことは、大きな収穫といえる。

判別分析法に慣れていなかったこともあるが、厄介と想定していた検診データと死因の関係が、これほどに分かりやすく出てくるとは予想していなかった。

#### (6) 不良データの抽出について

死因の中に不良データがあるという発想は、医学的にはなじまないことかもしれない。しかし、死因を追及するのではなく、死因を判別しようという発想からは、このような数学的な考え方が生まれてくる。

各種の死因領域が明瞭に存在するとすれば、その間の境界領域にはあいまいな死因領域も存在するという発想である。

不良データを削除した再分析の結果は、あまり変わらないというデータを含めると、88%にも達している。もちろん、若干の逆転・悪化が11%も出ているが、不良データを除去した処理は、抜群の効果をもたらしたといえる。判別得点が二倍近く増加している例が多いことも上げられる。後には、10倍以上の得点が出てきており、判別分析に関する認識が大きく変ってきた。

#### (7) 優良データの抽出に関して

優良データに絞ってみるとは、あまり好ましい試行とはいえない。

結果は、心：脳では結果が出力されない事態を招いたりしたが、その後数学的なチェックを進めている。

しかし、ほかの3分析では判別得点が30～60点と、初めの分析に比べて10～20倍の大きさになり、関係者には一つの傾向がつかめた感じがする。



### (8) データ項目を削減する試行

二通りの削減案を試行した。根拠を定めた削減案を作り、実行したが、大ざっぱに言えば大きな効果は見えてこない。代わりばえがしなかった。これから詳細分析を行って結論を出すことにしたい。

削減する項目について医学サイドからの一つの試案もあったが、これまでの医学の経験を先行させることは好ましくないかも知れないと考え、野村試案を実行した。

これからじっくり議論を重ねたいところである。

### (9) 「判定システム」の模索

情報システム学の立場からいうと、今回の実験の一つの目標は「判定システムの開発」にあるといえる。とくに健康維持・増進が重要になってきた今日、新しいアイデアで健康増進に役立つ情報システムが開発されるならば、きわめて望ましいことである。

まだまだ具体化までの道は遠い。が、この学際的な研究活動が意欲的に進められることは、意義深いと考えている。

以上

### 参考文献

「健康診断データの判別関数による分析」山崎明治ほか、福岡工業大学研究論集、36-1

付：健康増進研究会会員

野村 武 熊本県立大学・名誉教授・研究会代表

小山 和作 日赤熊本健康管理センター・名誉所長・研究会副代表

川島 英敏 日赤熊本健康管理センター・課長

山崎 明治 福岡工業大学・短期大学部・教授

井上 靖士 元福岡工業大学・短期大学部・教授

波多江 修 元西日本新聞社・局次長・研究会幹事