
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

～判別分析による死因判別の研究と考察～

〈中間報告その6・総括〉

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. 総論
2. 判別分析結果の総括
3. 検査項目の課題
4. 結び

[参考文献]

[メンバーの経歴]

Activation Research on healthcare examination results accumulated by JRC Kumamoto Healthcare Center : test by Discriminant Analytical Method

Study team

Takeshi NOMURA: Prefectural University of Kumamoto, Honorary Proff.

Wasaku KOYAMA: JRC Healthcare Center, Honorary Director

Akitaka YAMASAKI: Fukuoka Institute of Technology, Jr. Col., Proff.

Yasuto INOUE: Kyushu Sangyo Univ., Lecturer

Oasamu HATAE: Kumamoto Technical College, Lecturer

Abstract

Japanese Red Cross Kumamoto Healthcare Center has been accumulating huge data of healthcare examinations for the last 25 years.

Researches designed to develop information systems concerning these data has been done by the team of JRC Kumamoto Healthcare Center, Fukuoka Institute of Technology, Jr. Col. and Prefectural Univ. of Kumamoto.

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

For the past three years. JRC KHC has been trying to add the causes of patients' death to the examination results accumulated before.

To investigate the relationship between the healthcare examination results and the causes of death, our team tried to judge the relationship between the healthcare examination data before and the causes of death, applying the Discriminant Analysis Method.

We got good results about analysis. But to make them more sure, we had to face to the Protection of Personal Privacy.

So we decided to stop this researches until last fall.

1. 総論

表記の研究は、医学と統計数学と情報システム学の三分野の専門家群による学際的な研究活動であった。

最後の二～三年の研究を一言で言えば、健管センターに蓄積された個人データ（21の検査項目）の中から、死亡した人のデータを抽出して、その死因をデータに付け加え、死因別にグループ分けをして、データの判別分析を進めた。

判別分析とは、二つのグループ間についてどれくらい明瞭に違いがあるかを数値で表示するものである。実際の分析活動は、データ数の不足に悩みながらも、統計数学上の技法を織り込んで、相当の判別が可能になるところまで進むことができた。

さらに研究を進めるためには、個人のプライバシーの問題が重要になってきた。統計分析作業では完全にプライバシーを守れるのだが、死因について個人情報への遡及的な追跡調査などが必要になってくると、現在の常識では踏み込めないことになった。

また、データ数を増やしたい要望も強いが、これもプライバシー保護が重視されて、死因の追加・確認などが容易にはできない状況になってきた。

これらの事情で、研究活動は当面区切りを付けざるを得ないことになった。

2. 判別分析結果の総括

①最近の三年間は、分析法には判別分析を採用した。種々の試行を繰り返した結果、「データの分析方法」としてほぼ固めることができた。つまり全体的には、妥当性のある分析を行うことができるようになった。

②数学的には、検討項目の判別寄与度を考察した。例えば、赤血球数とヘモグロビンのように高い相関を示す項目の組み合わせについては、多重共線性（マルチコリニアリティ）現象が発生するため、二つの対策を立てた。

ア. 組み合わせのグループで項目間のそれぞれの相関係数を計算し、0.8以上となったら相関比が小さい方の項目を外す。次いでボックスのM検定を行い、X値が計算不能なら相関比の最小の項目を除く。残りの項目について判別分析を行う。

イ. 判別関数の係数と各変数の平均値のグループ間差の符号が、一致しているかを調べ、不一致なら除外する。

③分析結果として、8つの主要死因の組み合わせ別の判別得点一覧表を出力した。平均的にはおよそ70~80%のデータが死因と一致、のこりが不一致という結果となった。

この先は、死因の細かな遡及調査が必要となる。とくに判別得点の不一致群のチェックは大きな意味を持つ。

また、データ数を現在数より大幅に増やしたいとの要望も持っている。

しかし、いまの医学界の常識では、プライバシー保護のためにこれが実現できない状況にある。

④8つの主要な死因のうち、心臓疾患グループは死亡の原因というよりもいずれかの病気の結果という要素を大きく持っている。分析結果も明らかに異常であるが、それはリーズナブルであると考えている。他の死因グループと同等には扱えない。どんな工夫をすべきかは固まっていない。

付け加えれば、脳死グループにも異常の傾向が感じられる。

この他、8つの死因以外のマイナー・グループの分析を進めることも重要になってくる。このためのデータ量の増加は大きな課題であった。

3. 検討事項のまとめ

①2-③に述べたように、死因の組み合わせによって判別に関与する検査項目は、その中身と数が異なってくる。この関連を詳細に分析すると、こんご新しい議論がいろいろと出てきそうである。

②個々の死因は、実際には担当した医師の判断で決められたものであって、ほとんどは検死に基づくというような確たるものではない。分析の精度を上げていくためには、この判断に疑問を投げ掛けざるを得ない。遡及調査の必要性・可能性の議論が必要である。

③判別分析手法

手法については、相当に徹底した考察と工夫を重ねた。ただ今回は元データに空白項目(受診漏れ)があると、データを棄ててきた。これについては、

空白項目の救済策も考えており、こんごの検討課題とする。

④検査項目について

21の検査項目の詳細は過去の報告書に記載しているが、数字データのみを採用した。こんごアナログデータというか、生活習慣病データというか、多元的なデータを検討することが不可欠となる。数値データと非数値データが混在したデータの分析方法も考えなければならない。難しくなるうが、避けて通れない課題である。

⑤死因との結合ではなく、病気との結合へ前進したいものである。

健康管理の建て前から議論すると、死亡する前の病気の段階での分析が重要になってくる。果たしてデータが揃うか、分別できるか、などと課題は多そう。

⑥データ採取から発病・死亡までの期間

いまはデータが少ないために、データの採取から死亡までの期間については無視して、数年間のデータを無条件で取り扱ってきた。

また、女性のデータが男性より少なく、必要量だけ集らないという悩みもある。女性についての分析もぜひ実現したいところである。

⑦地域性のチェック

熊本県は地域の多元性に富んでいる。つまり、山あり、海あり、火山あり、温泉ありと変化に富んだ環境がある。この研究で人口動態から見た地域性の分析を試みようとしたが、まだ納得できる結論を出すに至っていない。

沖縄とか長野県とかの県単位の地域性の比較も重要であり、こんご県内、県外の比較分析という視点も忘れてはならない。

4. 結び

野村自身が大学での活動を辞めたために、研究活動は大幅に制約を受けるようになった。辞職後も約一年半にわたって研究を継続してきて、十年を超す長期間にわたった学際チームの研究となったが、大変ありがたいことであったと心から感謝している。

研究会全体としては、「健康増進・維持に役立つ情報システムの開発」を遠い

目標にして、こつこつと努力を重ねてきた。こんごも機会があれば、研究を継続したいテーマであるが、諸般の事情で一応の区切りをつける。

〈参考文献〉

[野村による報告] (熊本県立大学総合管理学部:アドミネストレーション)

- ①健管センターの蓄積データを活性化する研究・1…多種分析方法の整理
- ②健管センターの蓄積データを活性化する研究・2…多重集計、県内細分化
- ③健管センターの蓄積データを活性化する研究・3…因子分析法
- ④健管センターの蓄積データを活性化する研究・4…判別分析1
- ⑤健管センターの蓄積データを活性化する研究・5…判別分析2
- ⑥健管センターの蓄積データを活性化する研究・6…判別分析3 (未発表)
- ⑦健管センターの蓄積データを活性化する研究・7…総括

[山崎による報告] (福岡工業大学:研究論集)

- ①健診データの判別関数による分析・1
- ②健診データの判別関数による分析・2
- ③健診データの判別関数による分析・3 (分析方法と判定システムの調査)

〈メンバーの経歴〉

- 野村 武:熊本県立大学・名誉教授
小山 和作:日赤熊本健康管理センター・名誉所長
川島 英敏:日赤熊本健康管理センター・課長
山崎 明治:福岡工業大学短大部・教授
井上 靖士:九州産業大学・非常勤講師
波多江 修:熊本県技術短期大学・非常勤講師

以上