

健診データにおける検査項目間の 関連性評価に関する研究

森田 裕也[†] 山崎 恭[†]
[†] 北九州市立大学 国際環境工学部

1. はじめに

近年、超高齢社会における平均寿命と健康寿命の差(寿命ギャップ)が、憂慮すべき深刻な問題となっている[1]。健康寿命を延ばすには、健康を損なう前に対策を講じる必要があるが、現在、個人の健康状態に関する最も基本的なデータである健康診断データ(健診データ)を活用し、将来の健康状態を予測して病気の発症前に対処しようとする動きが活発化している。そこで、本研究では長期かつ大規模な健診データを解析して個人の将来の健康状態を予測することを目的とし、その実現に向けた第一段階として、健診データにおける各検査項目間の関連性を解析した結果について報告する。

2. 健診データの概要

本研究では、検査機関で取得された健診データに対して完全匿名加工を施したデータを使用した。なお、匿名加工に際しては、元の健診データの統計的な特徴を失わないような処理を施している。データは2009年～2018年の10年間に約23万人から取得されたものであり、検査と問診から成る88の項目を含んでいる。

3. 検査項目間の関連性

各検査項目間にどの程度の相関があるかを調べた。このうち、一部の検査項目について、検査項目間の相関を相関係数で表したヒートマップを図1に、また、相関係数を絶対値の大きな順に並べたときの検査項目の組み合わせを表1に示す。これらの結果から、ヘモグロビンとヘマトクリット、最高血圧と最低血圧のように、類似した検査項目間には強い相関のあることが確認される。

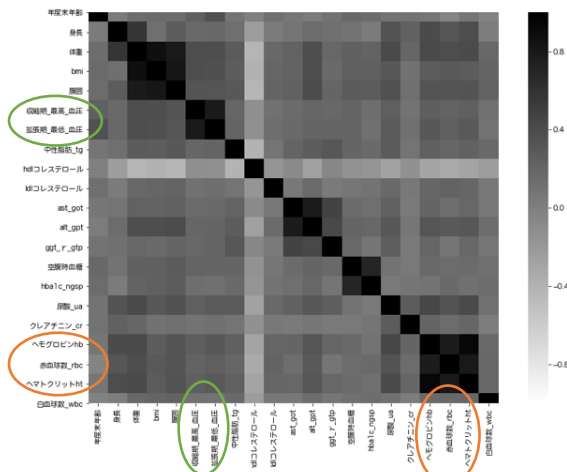


図1. 検査項目間の相関

表1. 相関係数の大きな検査項目の組み合わせ

	検査項目	検査項目	相関係数
1	ヘモグロビン	ヘマトクリット	0.945
2	循環器通院状況	投薬__高血圧症	0.918
3	BMI	腹囲	0.904
4	体重	腹囲	0.889
5	体重	BMI	0.876
6	投薬__高血圧症	血圧判定	-0.860
7	赤血球数	ヘマトクリット	0.801
8	最高血圧	最低血圧	0.795

次に、図1と同じ検査項目を対象に主成分分析を行った。累積寄与率を図2に示す。図2より累積寄与率が0.8を超えるのは第9主成分であり、元の情報の80%以上が4割程度の主成分に凝縮されていることが確認される。以上の結果から、これらの検査項目間には一定程度の相関のあることが確認される。

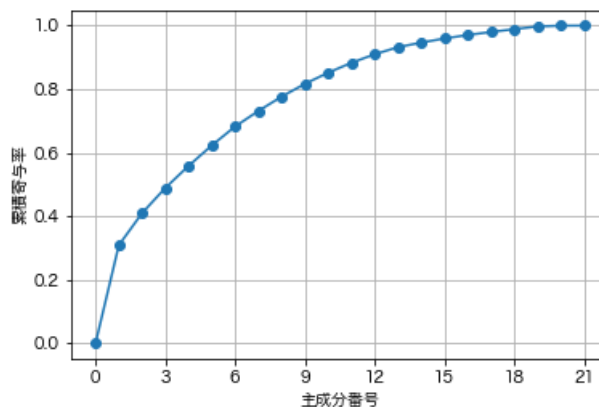


図2. 累積寄与率

5. まとめ

健診データにおける各検査項目間の関連性について解析を行った。今後、さらに解析を進め、個人の将来の健康状態を予測する手法を確立することが課題である。

謝辞: 本研究の一部は、平成31年度北九州市学術・研究振興事業調査研究助成金の助成を受けたものです。

参考文献

[1]内閣府、「平成30年度版高齢社会白書(概要版)」,
https://www8.cao.go.jp/kourei/whitepaper/w2018/html/gaiyou/s1_1.html (2019年12月24日参照)。