

科学研究費助成事業 研究成果報告書

平成 26 年 5 月 13 日現在

機関番号：62603

研究種目：若手研究(B)

研究期間：2011～2013

課題番号：23700347

研究課題名(和文)南極データベースの構築とそのデータ解析

研究課題名(英文)Formulation of antarctic-related database and its data analysis

研究代表者

小森 理 (Komori, Osamu)

統計数理研究所・大学共同利用機関等の部局等・その他

研究者番号：60586379

交付決定額(研究期間全体)：(直接経費) 1,700,000円、(間接経費) 510,000円

研究成果の概要(和文)：南極観測隊の健康管理と病気の予防を，統計的データ解析によって強化促進すること目的としてこのプロジェクトは開始された。まず滞在期間中または滞在前に実施された健康診断のデータと，その滞在期間中の観測された気象データ(温度，湿度，気圧など)の整備を進めた(データベースの構築)。ただ観測隊員の健康診断のデータの整備統合にはかなりの労力と関連部局との協力が必要となり(個人情報の取り扱い等)，今回のプロジェクト期間では終えることは出来ず今後とも引き続き継続して取り組むこととなった。またデータ整備とは別に健康診断のデータ解析のための新たな統計的手法の開発も行った。

研究成果の概要(英文)：This project has started for the purpose of promoting the health of personnel of Japanese Antarctic Research Expedition (JARE) and disease prevention by statistical data analysis. We tried to collect and arrange the data sets of health checkup examination measured during and before the expedition, and meteorological data sets such as temperature, moisture, atmosphere pressure measured in Antarctic. However, it took a long time to collect the health checkup data because of personal information protection, for which an ethics committee was set up to discuss the utility of the data. Hence it turned out that this project has not yet been done completely, so it also continues hereafter. On the other hand, some new statistical method has been developed for the analysis of health checkup data.

研究分野：統計科学

科研費の分科・細目：情報学・統計科学

キーワード：データ処理 医療統計

1. 研究開始当初の背景

南極での観測活動は僻地で行われるため、その観測隊員の健康状態は通常の日本国内での健康状態とは異なる様相を呈すると考えられていた。そのため観測期間中またはその前後で入念な健康診断が実施されており、そのデータも紙ベースではあるがかなりの量が蓄積されてきていた。ただそのデータは手つかずのままであり、観測隊員の健康の管理を今後さらに促進するためにはそのデータの整備と統計的な解析が必須の項目として認識されていた。

2. 研究の目的

(1) 上記の健康診断のデータの整備を進めるとともに、南極で観測されている気象データ、隊員の基本情報の整備加工をすすめることにより、環境要因、性別年齢等の個人的な要因から病気や怪我の因果関係を明らかにすることを目的としていた。またこの病気の原因を探る手段として新たな統計的手法の開発を目指していた。

(2) 南極の観測隊員の越冬中の疾病動向の把握は隊員の健康管理に欠かせないと同時に、医療学研究を推進するうえでも重要なことである。それは南極での医療設備の向上、隊員選抜の基本資料、派遣医師らの事前準備等につながるためである。日本基地での死亡事故は遭難の1件のみであり、いまだ致命的な疾患が発生しておらず、他国と比べて非常に充実した安全管理体制と医療体制を確保しているといえる。ただ、急性心筋梗塞を含む急性心疾患を南極の医療設備で治療することは現時点でも極めて困難である。そこで申請者は越冬前の健診のデータから急性心筋梗塞等の重篤な病気の事前予測と、それによる発症の予防を検討し、南極観測隊員の健康管理に貢献することを考えた。

3. 研究の方法

(1) データ整備について

健康診断のデータについては個人情報保護とも関わりがあるため、国立極地研究所の先生との密接な議論を通してそのデータの整備を目指した。気象データについてはすでに気象庁が一般に公開しているものもあるため、それをうまく利用し労力と時間をできるだけ削減した。隊員の基本情報は国立極地研究所の図書館所蔵の「南極資料」を参照しまとめを行った。

(2) 統計的手法の開発

病気の予知予防を目指す統計手法として今まで重回帰分析や数量化、主成分分析や因子分析等が一般に使われてきた。今回予測精度向上のため統計の分野で近年注目されてきた機械学習の手法に注目し新たな統計的手法の開発に取り組んだ。

(3) 個人情報の取り扱い

個人情報を含む医療データを解析するにあたり、その個人情報の機密性の確保は、研究を推進するうえできわめて重要な問題である。2003年には個人情報保護法が施行され、学術分野でもその重要性が認識されはじめている。そこで今回の南極データの使用に関しても、連結可能匿名化の暗号手法を用い観測隊員の氏名の匿名化をはかることを考えた。具体的には岡本(日本公衆衛生雑誌, 2003)が提案する JIS コードと乱数を使った手法を用いる。これはアルファベットの大文字小文字と0から9の数字の組み合わせとなり、全部で 5.74×10^{16} 通りの組み合わせで暗号化するため解読するのはほぼ不可能である。ただ注意すべきこととしては、観測隊員には複数年度南極観測に参加している者もいるため、観測隊次ごとに暗号化をする必要がある。また女性の隊員が少数であることと、年齢別で個人が判定できてしまう可能性があり、厳密な匿名化をどうするか検討も不可欠である。

4. 研究成果

(1) データ整備について

本プロジェクト開始前から続けてきた国立極地研究所の先生方とのデータ整備に向けての議論を本格化させた。

2011, 2012 年度においてはまず過去50年分の観測隊員の健康診断の入手に着手するにあたり、国立極地研究所の先生方の御協力を得て、研究所内にデータ取り扱いのための倫理委員会の設立を実現化した。今まで健診データ等の医療データに関する解析は前例がなく、倫理委員会も一からの立ち上げとなり、関連部局の先生方には多大な時間と労力をおかけした。特に委員会の基本方針の作成に尽力して下さった医師の先生にはここで心からの感謝の意を表したい。

2013 年からは健康判定委員会のもとで一隊員分の予備的なデータ解析が行われその有効性を検証して頂いた。有効性が確認され次第全ての隊員に関して同様な解析をすることになった。

2014 年度の4月には健康判定委員会での健康診断データ活用についてのガイドラインが提出して今後の本格的なデータ解析の下地ができた。

匿名化の作業も本格化した。申請者も「2013 年南極医学医療ワークショップ」において、個人情報保護に関する現在の状況を取りまとめ発表を行った([学会発表])。

気象データのデータベース化は比較的スムーズに事が運んだ。基本データはすでに気象庁で公開されているデータを少し加工することでデータ解析用のデータは用意することができた。ただ記号や数値の統合修正のような細かい作業には多くの時間がかかるためデータ整備のためにスタッフを雇用した。また過去50年度分の隊員の基本情報も

まとめた。年齢に関する一例を図1に示す。これまでの最高齢は58歳，最年少者は20歳となり，中央値は32歳であることが確認される。

越冬隊員の場合は選抜の要件として健康診断の結果が大きく関与しているため，おおむね一般よりも健康状態の良い集団のデータとなるが，閉じた人間関係，代わりがききにくく過重になりがちな労働，長らく太陽の登らない期間があること（極夜）など健康に悪影響を与える社会的，物理的環境を多く抱えている。そのため，越冬隊メンバーの選抜や越冬中の健康管理を行うにあたっては，医学・生理学的な性質だけでなく疫学的な特徴にも注目していくことに一定の意義がある。このような見地からデータベースの構築の意義を研究報告書してまとめた（[図書]）。

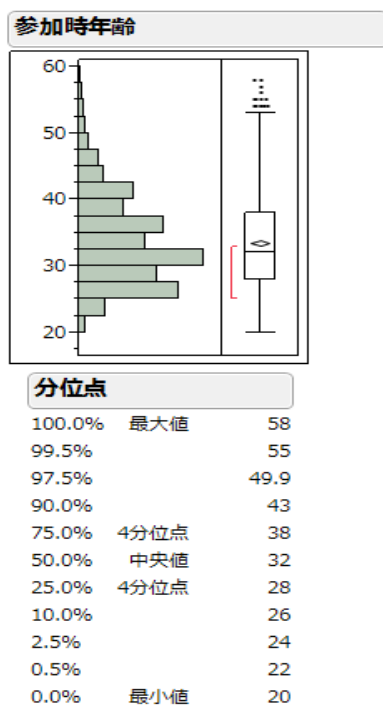


図1．歴代の南極観測隊員の年齢の分布

(2) 統計的手法の開発

データ整備に時間がかかる中，健診データの解析のための新たな統計的手法の開発も行った。

病気または怪我をしたグループと健常者のグループの2群を想定した今回の健診データにおいて，通常検定の枠組みで使われているt統計量の一般化を考察し，その理論的な性質を明らかにした。また実データまたはシミュレーションデータでの検証も行っており，今後健診データの整備が完了した段階で今回のデータにも適用を考えている（[学会発表]）。

病気や怪我をする人の検査値が持っているとする異質性を考慮した新たな回帰モデルを提案した。Kernel法の考えを取り入れ二つの回帰モデルを同時に当てはめる

switching regressionの手法を拡張し，推定精度の向上を試みた（図2）。（[学会発表]）。

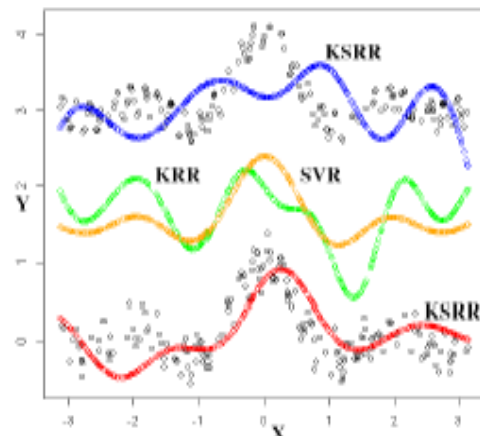


図2．提案手法(KSRR)と既存手法(KRR, SVR)との比較。黒丸はシミュレーションから発生させたデータ。KSRRの2つの回帰曲線(青と赤)の当てはめが良いことが確認できる。

(3) 今後の研究の方向性

プロジェクトの期間中には本格的なデータ整備，その後のデータ解析の作業に踏み込むことができなかった。今回はその前の労力と時間のかかる準備作業を地道に推し進めた形となった。しかしこの作業なくしてその後の十分なデータ整備と解析には進めないことは明らかであり，今後推し進められる健診データ解析の研究の土台作りの意味においては大きな貢献があったと思う。今後とも継続してこの研究課題に取り組む予定である。

5．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕(計 0件)

〔学会発表〕(計 3件)

Osamu Komori, Maximization of a Generalized t-Statistic for Two-Class Discrimination Problem, XXVI the International Biometric Conference, 2012年8月28日, Kobe International Conference Center

小森理, 小集団の匿名化の実際と課題 データの暗号化の科学, 2013年南極医学医療ワークショップ, 2013年7月20日, 国立極地研究所

小森理, Kernel Switching Ridge Regression, 日本統計学会春季集会, 2014年3月8日, 同志社大学今出川キャンパス

〔図書〕(計 1件)

小森理他, 統計数理研究所, 南極観測隊データベースの進捗報告と今後の計画(共同研究レポート), 2012年, p1-p21.

〔産業財産権〕

出願状況（計 0 件）

取得状況（計 0 件）

〔その他〕

なし

6．研究組織

(1)研究代表者

小森理（KOMORI, Osamu）

統計数理研究所 特任助教

研究者番号：60586379

(2)研究分担者

なし

(3)連携研究者

なし