

令和 4 年 5 月 18 日現在

機関番号：32660

研究種目：若手研究

研究期間：2018～2021

課題番号：18K18014

研究課題名（和文）等分散性が崩れたデータや欠測値を含むデータにおける推測理論の構築及びその応用

研究課題名（英文）Statistical inference theory and its application to data with unequal covariance matrices and missing data

研究代表者

川崎 玉恵（Kawasaki, Tamae）

東京理科大学・理学部第一部応用数学科・講師

研究者番号：30778212

交付決定額（研究期間全体）：（直接経費） 1,500,000円

研究成果の概要（和文）：本研究では、多変量解析において仮説検定問題の基礎である平均ベクトルに関する仮説検定問題を議論した。特に、データに欠測値を含んでいる場合と、2つの母集団分布における母分散共分散行列が母集団間で等しくない場合を背景とした研究を行った。欠測値を含んだデータに関する研究では、特に欠測構造のひとつである2-step単調欠測データを仮定し、等分散性が崩れた場合に関する研究では、Bennett型の検定統計量に着目して議論を行い、それぞれ検定統計量の提案とその近似分布について導出した。

研究成果の学術的意義や社会的意義

平均ベクトルに関する仮説検定問題は、すべてのデータが揃い、複数の母集団を考えた場合にはその母集団分布の母分散共分散行列が等しいことを仮定出来るのであれば、有効な検定手法が存在する。しかしデータサイエンスが進み、多様なデータが存在する現代においては、欠測値の存在や等分散性の崩れた場合のデータに関わる問題など、既存の手法通りでは対処できない問題が身近で重要な問題となっている。本研究成果はその解決法の1つとして、統計学のさらなる発展を目指すものとなっている。

研究成果の概要（英文）：In this study, we discussed the hypothesis testing problem regarding the mean vector, which is the basis of hypothesis testing problems in multivariate analysis. We focused the case when the data contain missing values and when the population covariance matrices of the two population distributions are not equal across the populations. In the study on the data containing missing values, we discussed the assumption of two-step monotone missing data, which is one of the missing structures. In the study on the case of the unequal covariance matrices, we focused on the Bennett type test statistic. We derived the test statistics and their approximate upper percentiles for each assumption.

研究分野：多変量解析

キーワード：数理統計学 統計的仮説検定

1. 研究開始当初の背景

統計学は、品質管理や臨床科学などの分野で古くから広く活用されており、近年のデータサイエンスや人工知能等の技術にも必要不可欠な、現代社会において非常に注目度の高い研究分野のひとつである。データが容易に手に入り、誰もが簡単にデータ分析することのできるようになった一方で、複数の母集団間で等分散性が仮定できないような場合や、データが何かしらの原因で欠測値を含んだものとして構成される場合は多く存在する。これらの問題については、これまで多数の研究がなされており、有効な手段も多くある一方で、結果が複雑すぎるものや、仮定の厳しいものも多く、精度は高く、かつ使いやすい容易な形で表された統計的推測理論の構築等、挑戦すべき研究課題はまだ多く残されている。申請者はこれまで、上記の研究課題に関連した成果をいくつかあげてきている。

本研究課題ではこのような研究背景から、これまでの研究成果を基に基礎理論の構築と応用問題における統計的手法の開発を目的として研究を行った。具体的には、多変量解析において、平均ベクトルや部分平均ベクトルに対する仮説検定問題を基盤とし、

- (1). 「等分散性が崩れた場合」
- (2). 「データに欠測値を含んでいる場合」

において、上記の問題を解決する方法のひとつとして研究を行った。

2. 研究の目的

上記の通り、現代社会における統計学の重要性は増している。

- (1). 申請者はこれまでに、2つの母集団間における等分散性が崩れた問題、多変量解析では特に、母分散共分散行列が母集団間で等しくない場合の問題について、いくつかの研究成果をあげてきた。本研究課題ではその発展として、統計学の世界で未解決となっているこの問題に対し、部分平均ベクトルにおける仮説検定問題に関する新たな検定統計量を提案し、その分布について理論的に議論し、いくつかの性質を与え、解決策のひとつとなる手法の開発を目指した。既存の手法には、検定統計量の分布の近似精度を追求するあまりに結果が複雑な形で与えられており、実際のデータ分析の場では取り扱いの難しいものや、データや母分散共分散行列の構造に課されている仮定や条件が厳しいものも多くある。そこで本研究課題の目的は、仮定をなるべく設けずに、近似精度を保ちつつも、簡便な形で扱いやすい近似分布の導出を目指した。
- (2). 申請者はこれまで、データに欠測構造を仮定した場合の漸近理論に基づく研究成果をいくつかあげている。欠測構造を仮定するメリットのひとつは、漸近理論をはじめとした理論的な議論が行いやすい点であるといえる。本研究課題ではこれまでの研究結果の発展として、新たな検定統計量の提案と、これまであまり漸近分布が議論されてこなかった検定統計量の帰無分布の導出とその漸近的性質、簡便な近似分布の提案を、平均ベクトル、部分平均ベクトルの検定問題で取り扱う。さらに上記(1)の研究テーマと融合させることにより、欠測値を含むデータに対する母分散共分散行列が異なる場合の問題を考え、さらなる研究の発展を目指した。

3. 研究の方法

以下の方法で研究を行った。

(1). 母分散共分散行列が異なる場合

- ① 2標本問題における部分平均ベクトルの検定問題について議論を行った。部分平均ベクトルについては、RaoによるU統計量が知られており、本研究課題の問題に適用するため、Bennettによる2標本問題を1標本問題として扱う手法を適用し、部分平均ベクトルに関する2つの母分散共分散行列が異なる場合の検定統計量を与えた。
- ② 2つの母集団のサンプルサイズが等しい場合と異なる場合で議論を分け、欠測値を含まない完全にデータが揃っている場合は正確な帰無分布が得られることを示した。2-step単調欠測データをデータ構造に仮定した場合は、検定統計量として尤度比検定統計量を用いることとし、その近似帰無分布について申請者の先行研究を適用する形で提案した。
- ③ モンテカルロ・シミュレーションを用いて、2-step単調欠測データの検定統計量とその近似パーセント点の近似精度を評価し、本数値実験から分かる性質をいくつか示した。

(2). データに欠測値を含んでいる場合

- ① 条件付き平均ベクトルの仮説検定問題について、2標本問題を議論した。この問題に関する同時信頼区間の導出をしようと試みたところ、完全データの下での議論がなされていないことが分かったため、まずは完全データの下で同時信頼区間の議論を行った。
- ② 続いて、2標本問題に関する条件付き平均ベクトルの仮説検定について、欠測構造に2-step単調欠測データを仮定して、この場合の検定統計量の導出を行った。この場合の検定統計量は、尤度比検定統計量を用いて提案したため、その近似精度を高めるため、漸近理論による修正尤度比検定統計量と、修正尤度比検定統計量の修正係数を線形補間によって与えた検定統計量の2つの提案を行った。
- ③ また、1標本問題における平均ベクトルの検定統計量の帰無分布として、F分布の自由度を調整した近似帰無分布の導出について議論した。
- ④ 2標本問題における部分平均ベクトルの検定問題について、RaoのU統計量を基にした検定統計量を議論し、その近似帰無分布を議論した。

4. 研究成果

- (1). 母分散共分散行列が異なる場合の議論については、部分平均ベクトルに関する2標本問題について議論した。部分平均ベクトルの検定統計量としてRaoのU統計量を用いて、2標本問題の確率ベクトルに対して重み付きの差をとることで1標本問題に帰着させるBennettによる手法を用いて議論を行った。完全データの場合は1標本問題となり、正確な分布を用いることが可能となり、2-step単調欠測データの下では、申請者による1標本問題における先行研究の結果を適用することで解決する方法を示した。

- (2). 欠測値を含むデータについては、2-step単調欠測データを仮定した下記の2つのテーマについて研究成果をあげた。

- ① 2標本問題の条件付き平均ベクトルの仮説検定に対する検定統計量としては、RaoによるU統計量が既にあるが、この検定統計量は尤度比検定統計量となっているため同時信頼区間の導出を行うことが出来ない。そこで本研究課題では、ホテリングの T^2 検定統計量のような二次形式で与えられる新たな検定統計量を提案し、その帰無分布について2通りの近似解を与えた。1つ目は、提案した検定統計量の構造に着目した近似である。ホテリングの T^2 検定統計量は、ベクトル部分が多変量正規分布に、行列部分がウィシャート分布に従う統計量であるが、今回提案した検定統計量は、行列部分はウィシャート分布に従っているが、ベクトル部分が多変量正規分布には従わないことが分かった。しかし、検定統計量の極限分布はカイ二乗分布であることが確認できたため、ベクトル部分を多変量正規分布とみなすことにより、近似的にF分布に従うような第二自由度とその係数を与えた。2つ目もF分布を用いた近似解であるが、提案した検定統計量を近似的にカイ二乗分布の比となるような形で表現し直し、F分布の第一自由度と第二自由度、その係数の推定結果を導出した。以上から、自由度を調整したF分布による近似帰無分布を与えた。さらに、1つ目、2つ目の近似帰無分布を用いることによって、先行研究では与えることが出来なかった同時信頼区間の導出を行うことが出来た。本結果は、Kawasaki, Naito and Seo (2020)に掲載されている。
- ② 2標本問題における条件付き平均ベクトルの仮説検定問題を考えた。この問題に対する最尤推定量を与え、その結果を用いて尤度比検定統計量を構成した。ここで、尤度比検定統計量の極限分布はカイ二乗分布であるが、サンプルサイズが少ない場合には近似精度がよくないことが知られており、本研究課題では、2つの結果を与えた。1つ目は、尤度比検定統計量を大標本漸近枠組みのもとで展開することにより修正尤度比検定統計量を提案した。この修正尤度比検定統計量は、極限分布であるカイ二乗分布への収束が尤度比検定統計量よりも早くなるような修正となっている。そのため、カイ二乗分布のパーセント点を用いて検定を行う際、サンプルサイズが少ない場合でも元の尤度比検定統計量よりも精度の良い検定を行うことが期待できる。2つ目は、修正尤度比検定統計量の修正係数を、欠測値の含まれない完全データのもとの尤度比検定統計量に対する修正係数を用いて、式が簡便で使いやすいものとして提案した。この完全データのもとの修正係数は、1つ目で与えた修正係数に対し、欠測値を含むデータのサンプルサイズを0とすることで得ることができ、その式はもとの修正係数の式よりも簡便なものとなっている。そこで、2-step単調欠測データのデータ構造を生かし、2-step単調欠測データの分布が間にくるような2つの完全データを用意し、それら2つの完全データに関する修正係数と欠測の比率を用いて線形補完を行い、新たな修正係数を提案した。この修正係数を用いた修正尤度比検定統計量を使うことで、1つ目の提案と同様、サンプルサイズが少ない場合でも近似精度が良くなることが分かった。本結果は、Kawasaki and Seo (2021)に掲載されている。

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 3件）

1. 著者名 Tamae Kawasaki and Takashi Seo	4. 巻 50
2. 論文標題 Modified likelihood ratio test for sub-mean vectors with two-step monotone missing data in two-sample problem	5. 発行年 2021年
3. 雑誌名 Austrian Journal of Statistics	6. 最初と最後の頁 88-104
掲載論文のDOI（デジタルオブジェクト識別子） 10.17713/ajs.v50i1.928	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Tamae Kawasaki and Takashi Seo	4. 巻 7
2. 論文標題 Tests for mean vector using approximate degrees of freedom with two-step monotone missing data	5. 発行年 2020年
3. 雑誌名 Proceedings of the 62nd ISI World Statistics Congress	6. 最初と最後の頁 259-264
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Kawasaki Tamae, Naito Toshiki, Seo Takashi	4. 巻 9
2. 論文標題 T^2 Type Test Statistic and Simultaneous Confidence Intervals for Two Sub-mean Vectors	5. 発行年 2020年
3. 雑誌名 International Journal of Statistics and Probability	6. 最初と最後の頁 1~8
掲載論文のDOI（デジタルオブジェクト識別子） 10.5539/ijsp.v9n1p1	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計4件（うち招待講演 0件/うち国際学会 3件）

1. 発表者名 Tamae Kawasaki and Takashi Seo
2. 発表標題 Tests for mean vector using approximate degrees of freedom with two-step monotone missing data
3. 学会等名 62st World Statistics Congress（国際学会）
4. 発表年 2019年

1. 発表者名 Tamae Kawasaki and Takashi Seo
2. 発表標題 Likelihood ratio tests for sub-mean vectors with two-step monotone missing data in two-sample problem
3. 学会等名 Joint International Society for Clinical Biostatistics and Australian Statistical Conference (国際学会)
4. 発表年 2018年

1. 発表者名 川崎 玉恵, 瀬尾 隆
2. 発表標題 2-step単調欠測データのもとでの部分平均ベクトルの検定に対する修正尤度比検定統計量
3. 学会等名 科研費シンポジウム「融合する統計科学」
4. 発表年 2018年

1. 発表者名 Tamae Kawasaki and Takashi Seo
2. 発表標題 T2 type test statistic and simultaneous confidence intervals for sub-mean vectors in two-sample problem
3. 学会等名 14th International Conference of the ERCIM WG on Computational and Methodological Statistics (国際学会)
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------