

令和 2 年 6 月 5 日現在

機関番号：32634

研究種目：基盤研究(C) (一般)

研究期間：2017～2019

課題番号：17K00056

研究課題名(和文) 多様な状況における多重検定方式の研究

研究課題名(英文) Study on multiple testing procedure for various situations

研究代表者

西山 貴弘 (NISHIYAMA, Takahiro)

専修大学・経営学部・准教授

研究者番号：30516472

交付決定額(研究期間全体)：(直接経費) 2,900,000円

研究成果の概要(和文)：本研究では、多様な状況の下で、平均ベクトルや共分散構造に関する統計的仮説検定の理論および現実問題への応用について研究を行った。特に、1. 高次元データに対する平均ベクトルおよび共分散構造に関する仮説検定、2. ノンパラメトリックな場合における平均間の多重比較法、3. 欠測が生じた場合の分割表における仮説検定、に対していくつかの研究成果が得られた。

研究成果の学術的意義や社会的意義

本研究では、高次元データにおける統計的仮説検定問題に対して研究成果を多く出すことができおり、「ビッグデータ時代」の中で取り扱われることが多いデータに対する有用な統計理論の開発を行った。さらにノンパラメトリックな場合や欠測値を含む場合といった、より現実に起こりえる状況での理論開発を行っており、この分野における理論的發展に大きく貢献することができたと考えられる。加えて、大規模なモンテカルロ・シミュレーションによる数値実験や実データ解析への適用を行っており、実際のデータ解析の場への応用も十分に期待できる。

研究成果の概要(英文)：In this study, we discussed the statistical hypothesis testing for mean vector and for covariance structure in various situations. In particular, we derived some results concerning with following 3 topics. 1. Statistical hypothesis testing for mean vectors and for covariance structure in high-dimensional settings. 2. Multiple comparisons among means for non-parametric situation. 3. Statistical hypothesis testing for contingency table with missing data.

研究分野：統計科学

キーワード：仮説検定 漸近理論 多重比較 高次元データ

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

近年の情報化社会の発展に伴って、様々なデータがインターネットなどを通じて簡単に手に入るようになった。特に最近では日々新しく生まれるデータがますます巨大化し、それらのデータが互いに融合し複雑化する「ビックデータ時代」となっている。このビックデータに関連し、最近では遺伝子データ、画像・音声データなどといった「高次元データ」や、何らかの理由によってデータに欠損が生じてしまっている「欠測値データ」に対する統計解析手法の研究が盛んに行われていた。このようなデータに対する平均ベクトルに関する仮説検定問題は、1標本問題、2標本問題については多くの研究結果があるが、複数個の母集団間の比較に対する検定方式である「多変量多重比較法」に関する研究はほとんど見られなかった。

平均ベクトル間の多変量多重比較法は、例えば医学・薬学分野で臨床試験データ解析に用いられるなど、自然科学や社会科学を問わず様々な分野で用いられる非常に重要な統計手法である。この問題は、通常、同時信頼区間の形を用いて議論されるため、「平均ベクトル間の比較に関する同時信頼区間の構成」が重要な問題となり、総標本数が次元数よりも多いという大標本枠組みの下や、母集団分布に何らかの仮定をおいた下では様々な研究が行われている。しかしながら、それらの結果は次元数が総標本数よりも大きいという「高次元枠組み」の下や、「分布の仮定が成り立たない場合」では使用することができないため、これらの状況に対して使用可能な新たな理論と方法論の開発が望まれていた。

また、多変量解析手法には共分散構造の仮定が必要なものがある。そのため、共分散構造に関する統計的仮説検定は重要な問題であり、近年、高次元データに対する理論研究も多く行われていた。

### 2. 研究の目的

本研究では、これまでの研究成果を踏まえ、多変量統計解析理論の分野である複数母集団における多重比較法、および関連する問題の統計理論に焦点を絞り、平均ベクトルに関する統計的仮説検定の理論および現実問題への応用について以下の研究を行った。

- (1) 高次元データに対する平均ベクトルおよび共分散構造に関する新たな仮説検定方式の提案。
- (2) ノンパラメトリックな場合における平均間の多重検定方式の提案。
- (3) 欠測が生じた場合の分割表における仮説検定方式の提案。

### 3. 研究の方法

研究の目的(1)~(3)に対して、以下の方法で研究を遂行した。

(1) 高次元枠組みの下で、平均ベクトルと共分散行列の同時検定問題に対して、既存の統計量を組み合わせることによって新たな検定統計量を提案する。また、平均ベクトルの同等性検定問題においてよく知られている統計量に対して、新たな近似分布を導出することによって、より精度の高い検定方式を提案する。さらに共分散構造の検定問題に関連して、高次元データに対する独立性検定問題を考え、新たな検定方式を提案する。

(2) Brunner-Munzel 検定を基にした、1変量の場合の Behrens-Fisher 問題に対する新たなノンパラメトリック多重比較法を提案し、帰無仮説の下で検定統計量の近似分布を導出する。

(3) 欠測が生じている  $2 \times 2$  分割表に対する対称性の仮説検定問題を考え、特に欠測メカニズムが MNAR (Missing not at random) の場合を取り扱い、グラフィカルモデルに対する適合度検定を考え、その下で対称性の検定を行う方法論を提案する。

### 4. 研究成果

研究の目的(1)~(3)に対して、期間内に以下の研究成果を得ることができた。

(1) まず、高次元データに対して、一般に  $k$  個の母集団間の平均ベクトルと分散共分散行列の同時検定問題について議論し、非正規性の下でこの仮説検定問題に対する検定統計量を提案し、いくつかの高次元枠組みの下でその近似分布を導出した。さらに検出力について議論を行い、提案統計量の有用性を確認することに成功した。

次に、高次元データに対する独立性検定問題について議論を行った。具体的には、高次元の場合に対応した RV 係数に基づいた検定統計量を提案し、正規性の下でいくつかの高次元枠組みの下でその近似分布を導出した。さらに検出力について議論を行い、提案統計量の有用性を確認することに成功した。また、この検定方式を用いて、どの確率ベクトル間に相関が生じているかを検出するための多重対比較法を提案した。この結果に関連して、非正規性の下でも同様な議論を行い、高次元データにおける無相関性の検定に対する検定方式の提案をした。

さらに、高次元データに対して、多変量 Behrens-Fisher 問題として知られる、2つの母集団の共分散構造が異なる場合の平均ベクトルの同等性検定問題(2標本問題)について議論を行った。具体的には、高次元データにおける多変量 Behrens-Fisher 問題に対してよく知られている統計量について、エッジワース展開とカイ2乗分布に基づく新たな2つのタイプの近似分布を導出し、理論的にその近似の誤差評価を行った。また、すでに知られている正規分布に基づいた近似(漸近正規性)との近似精度の比較を行い、新たに導出し

た近似分布の有用性を確認することに成功した。

(2) 母集団分布の仮定をおかないノンパラメトリックな場合に対して、2 標本問題に対する代表的な検定手法として知られている Brunner-Munzel 検定を基にした、1 変量の場合の Behrens-Fisher 問題に対する新たなノンパラメトリック多重比較法を提案し、帰無仮説の下で検定統計量の近似分布を導出した。さらに検出力や頑健性について数値的に評価を行い、提案方式の有用性を確認することに成功した。

(3) 欠測データ解析に関連して、欠測が生じている  $2 \times 2$  分割表に対する対称性の仮説検定問題を議論した。ここでは、特に欠測メカニズムが MNAR (Missing not at random) の場合を取り扱い、グラフィカルモデルに対する適合度検定を考え、その下で対称性の検定を行う方法論を提案した。

## 5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 4件/うち国際共著 3件/うちオープンアクセス 1件）

1. 著者名 Hyodo Masashi, Nishiyama Takahiro, Pavlenko Tatjana	4. 巻 157
2. 論文標題 On error bounds for high-dimensional asymptotic distribution of L2-type test statistic for equality of means	5. 発行年 2020年
3. 雑誌名 Statistics & Probability Letters	6. 最初と最後の頁 108637 ~ 108637
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.spl.2019.108637	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する
1. 著者名 Hyodo Masashi, Nishiyama Takahiro, Pavlenko Tatjana	4. 巻 178
2. 論文標題 Testing for independence of high-dimensional variables: V-coefficient based approach	5. 発行年 2020年
3. 雑誌名 Journal of Multivariate Analysis	6. 最初と最後の頁 104627 ~ 104627
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.jmva.2020.104627	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する
1. 著者名 Masashi Hyodo, Takahiro Nishiyama, Tatjana Pavlenko	4. 巻 No.18-08
2. 論文標題 Testing independence in high-dimensional data: V-coefficient based approach	5. 発行年 2018年
3. 雑誌名 Technical Report, Statistical Research Group, Hiroshima University	6. 最初と最後の頁 1-23
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 無
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 該当する
1. 著者名 Hyodo Masashi, Nishiyama Takahiro	4. 巻 27
2. 論文標題 A simultaneous testing of the mean vector and the covariance matrix among two populations for high-dimensional data	5. 発行年 2017年
3. 雑誌名 TEST	6. 最初と最後の頁 680 ~ 699
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s11749-017-0567-x	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Hyodo Masashi, Nishiyama Takahiro	4. 巻 印刷中
2. 論文標題 Simultaneous testing of the mean vector and covariance matrix among k populations for high-dimensional data	5. 発行年 2019年
3. 雑誌名 Communications in Statistics - Theory and Methods	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) doi.org/10.1080/03610926.2019.1639751	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

[学会発表] 計10件 (うち招待講演 2件 / うち国際学会 5件)

1. 発表者名 Takahiro Nishiyama, Masashi Hyodo and Tatjana Pavlenko
2. 発表標題 Asymptotic distribution of test statistic for two sample test under high-dimensional setting
3. 学会等名 10th International Workshop on Simulation and Statistics (国際学会)
4. 発表年 2019年

1. 発表者名 Takahiro Nishiyama, Masashi Hyodo and Tatjana Pavlenko
2. 発表標題 On error bounds for high-dimensional asymptotic distribution of L2-type test statistic
3. 学会等名 International Symposium on Theories and Methodologies for Large Complex Data (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 西山 貴弘、兵頭 昌、Tatjana Pavlenko
2. 発表標題 高次元データに対するRV係数に基づく独立性検定
3. 学会等名 日本計算機統計学会
4. 発表年 2018年

1. 発表者名 Takahiro Nishiyama and Masashi Hyodo
2. 発表標題 Simultaneous test for mean vectors and covariance matrices among k populations for high-dimensional data
3. 学会等名 International Conference on Computational Statistics (国際学会)
4. 発表年 2018年

1. 発表者名 小川 颯、兵頭 昌、西山 貴弘
2. 発表標題 複数の高次元確率ベクトル間の無相関性の検定
3. 学会等名 統計関連学会連合大会
4. 発表年 2018年

1. 発表者名 伊井 裕亮、田畑 耕治、西山 貴弘
2. 発表標題 無視できない欠測における2×2分割表に対する対称性の検定
3. 学会等名 統計関連学会連合大会
4. 発表年 2018年

1. 発表者名 Takahiro Nishiyama, Masashi Hyodo and Tatjana Pavlenko
2. 発表標題 Testing independence in high-dimensional data: V-coefficient based approach
3. 学会等名 International Conference of the ERCIM WG on Computational and Methodological Statistics (招待講演) (国際学会)
4. 発表年 2018年

1. 発表者名 西山 貴弘、村上 秀俊
2. 発表標題 Nonparametric multiple comparison procedure in the one-way layout
3. 学会等名 統計関連学会連合大会
4. 発表年 2017年

1. 発表者名 兵頭 昌、小川 颯、西山 貴弘
2. 発表標題 多標本問題に対するユークリッド距離を利用した平均ベクトルと分散共分散行列の同時検定
3. 学会等名 日本数学会秋季総合分科会
4. 発表年 2017年

1. 発表者名 Takahiro Nishiyama, Masashi Hyodo
2. 発表標題 Simultaneous test for mean vectors and covariance matrices in high-dimensional settings
3. 学会等名 New Zealand Statistical Association and the International Association of Statistical Computing (Asian Regional Section) Joint Conference (国際学会)
4. 発表年 2017年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

<p>専修大学研究者情報データベース  <a href="https://kjs.acc.senshu-u.ac.jp/sshhp/KgApp?resId=S001699">https://kjs.acc.senshu-u.ac.jp/sshhp/KgApp?resId=S001699</a></p>
---

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
--	---------------------------	-----------------------	----