氏名（本籍）　八木文香（静岡県）
学位の種類　博士（理学）
学位記番号　甲第1150号
学位授与の日付　2018年3月19日
学位授与の要件　学位規則第4条第1項該当
学位論文題目　**Some tests for mean vectors with monotone missing data**
（単調欠測データをもつ平均ベクトルに対するいくつかの検定）

論文審査委員　（主査）教授　瀬尾　隆
教授　佐藤　洋祐　　教授　橋口　博樹
教授　宮岡　悦良　　准教授　柳田　昌宏

論文内容の要旨

統計データ解析の際，データが欠測してしまうことはよく起こることであり，その場合の統計解析手法を与えることは重要である．本論文では，多変量統計解析における欠測値をもつ平均ベクトルの仮説検定問題を取り扱う．特に，欠測のパターンが単調である「単調欠測データ」の場合について，1標本問題から多標本問題までの平均ベクトルの仮説検定問題について考える．ここで，単調欠測データとは，一つの観測ベクトルに対して，一度欠測が生じるとそれ以降の成分がすべて欠測となっているデータのことである．すなわち，k-ステップ単調欠測データは，$\boldsymbol{x}_{ij}$ $(i=1,2,\ldots,k,\ j=1,2,\ldots,n_i)$ を $p_i \times 1$ のランダム標本ベクトル（列ベクトル）とすると

$$
\left(\begin{array}{ccccccc}
\multicolumn{7}{c}{\boldsymbol{x}'_{11}} \\
\multicolumn{7}{c}{\vdots} \\
\multicolumn{7}{c}{\boldsymbol{x}'_{1n_1}} \\
\multicolumn{4}{c}{\boldsymbol{x}'_{21}} & * & \cdots & * \\
\multicolumn{4}{c}{\vdots} & \vdots & & \vdots \\
\multicolumn{4}{c}{\boldsymbol{x}'_{2n_2}} & * & \cdots & * \\
\multicolumn{7}{c}{\ddots} \\
\boldsymbol{x}'_{k1} & * & \cdots & * & * & \cdots & * \\
\vdots & \vdots & & \vdots & \vdots & & \vdots \\
\boldsymbol{x}'_{kn_k} & * & \cdots & * & * & \cdots & *
\end{array}\right)
$$

と表すことができる．ここで $p = p_1 > p_2 > \cdots > p_k > 0$ であり，“$*$” は，欠測値を表し，“$\boldsymbol{x}'_{ij}$” は $\boldsymbol{x}_{ij}$ を転置したベクトル $(1 \times p_i)$ を表している．また，本論文では，観測されたベクトルは多変量正規分布に従うとし，欠測のメカニズムは MCAR (Missing Completely at Random) を仮定して議論する．欠測のメカニズムについては，Little and Rubin (2002) などを参照されたい．

本論文の構成は以下の通りである．

第 1 章では本論文全体の要旨について，先行研究を踏まえて述べる．

第 2 章は，1 標本問題に対するものである．すなわち，単調欠測データをもつ場合の次のような仮説検定問題

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \text{ vs. } H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$$

を考える．ただし，$\boldsymbol{\mu}$ は平均ベクトルで，$\boldsymbol{\mu}_0$ は既知ベクトルとする．欠測値をもたない完全データの場合，この問題に対する検定統計量として，ホテリングの T^2 検定統計量が知られており，その帰無分布は正確に F 分布に従う．それに対して 2-ステップ単調欠測データの場合，Chang and Richards (2009) や Seko, Yamazaki and Seo (2012), Kawasaki and Seo (2016) などによって，ホテリングの T^2 検定統計量に対応する T^2 型検定統計量が提案され，その近似帰無分布などが議論されている．しかしながら，3-ステップ以上の場合，T^2 型検定統計量そのものを導出することが困難なため，Krishnamoorthy and Pannala (1999) では，ホテリングの T^2 検定統計量に類似した検定統計量（T^2_{KP} とおく）を提案し，T^2_{KP} の近似帰無分布を与えている．検定統計量に用いられる平均ベクトルと分散共分散行列の MLE については，Anderson and Olkin (1985) や Jinadasa and Tracy (1992), Kanda and

Fujikoshi (1998) などを参照されたい．

上記のような先行研究に対して，2.1 節では，3-ステップの場合について，Jinadasa and Tracy (1992) による表記法を用いて「簡便な T^2 型検定統計量」($\widetilde{T}^2$ とおく）を導出している．そして，この $\widetilde{T}^2$ は，表記法及び導出法については T^2_{KP} とは異なるものであるが，統計量そのものは本質的に同じものであることを確認し，T^2_{KP} の近似帰無分布の上側パーセント点とは異なる近似上側パーセント点を提案している．これは，完全データの場合の検定統計量の正確な上側パーセント点（F 分布を用いたもの）を基にした線形補間によるものである．この近似上側パーセント点を利用し，$\boldsymbol{\mu}$ の線形結合に対する近似同時信頼区間も提案している．さらに，これに関連して，平均成分の同等性仮説検定問題

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_p \text{ vs. } H_1 : \text{not } H_0$$

についても $\widetilde{T}^2$ を用いて検定できることを示している．ここで，$\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_p)'$ である．また，モンテカルロ・シミュレーションを通して，数値評価を行い，いくつかのパラメータに対して，提案した近似上側パーセント点が良い近似となっていることを示している．最後に，本研究で導出した $\widetilde{T}^2$ を用いた検定の手法を説明するために，Wei and Lachin (1984) のコレステロール値による実データによる数値例を紹介している．2.2 節では，2.1 節を拡張した k-ステップの場合の結果を与えており，モンテカルロ・シミュレーションにより，Krishnamoorthy and Pannala (1999) による近似上側パーセント点と本研究で提案した近似との数値比較も行っている．

第 3 章は，2 標本問題においてデータが単調欠測構造をもつ場合の $\widetilde{T}^2$ 検定統計量に関するものである．まず，仮説検定問題

$$H_0 : \boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)} \text{ vs. } H_1 : \boldsymbol{\mu}^{(1)} \neq \boldsymbol{\mu}^{(2)}$$

について考える．ここで $\boldsymbol{\mu}^{(1)}$ と $\boldsymbol{\mu}^{(2)}$ はそれぞれ，第 1 母集団と第 2 母集団の平均ベクトルであり，各母集団の分散共分散行列は等しいと仮定する．3.1 節では，3-ステップ単調欠測データに基づく $\widetilde{T}^2$ を導出している．また，第 2 章の 1 標本問題のときと同様に，近似上側パーセント点と，$(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})$ の線形結合に対する近似同時信頼区間を提案している．ここで，第 1 母集団の欠測部分（2, 3 段目）の次元は，対応する第 2 母集団のものと等しいと仮定しているが，この結果は，第 1 母集団の 3 段目と，第 2 母集団の 2 段目のサンプル数をそれぞれゼロとすることによって，欠測部分の次元が異なる 2-ステップの場合にも適

用することができる．さらに，Jinadasa and Tracy (1992) によって与えられている 1 標本問題における MLE を，多標本問題へと拡張させることに成功している．この結果を利用し，複数個の母集団における平均ベクトル間の対比較及び対照比較に対する近似同時信頼区間も与える．最後に，いくつかのパラメータに対して，モンテカルロ・シミュレーションを行っている．この結果から，極限分布であるカイ二乗分布の上側パーセント点を使うよりも，今回提案した近似上側パーセント点を利用した場合の方が近似精度が良いことがみてとれる．3.2 節では，これらを k-ステップへ拡張させた場合の結果を与えている．

一方，完全データの場合，ホテリングの T^2 検定は尤度比検定となっているが，単調欠測データの場合，ホテリングの T^2 検定統計量に対応する $\widetilde{T}^2$ 検定統計量は尤度比検定ではない．このことから，第 4 章と第 5 章では，単調欠測データにおける尤度比検定について議論する．第 4 章は，1 標本問題に対するものである．先行研究として，Krishnamoorthy and Pannala (1998) は，単調欠測データに基づく尤度比を，$(\sum_{i=1}^{k} n_i) \times p_k$ のデータを用いた「p_k 次元の平均ベクトルの検定の尤度比」と $(k-1)$ 個の「部分平均ベクトルの検定の尤度比」の積で表している．これらはすべて，完全データの場合の尤度比とみることができ，互いに独立である．このことを利用し，Krishnamoorthy and Pannala (1998) は，特に 2-ステップの場合の修正尤度比検定統計量の分布関数の漸近展開を明示的に与えている．また，完全データの場合の部分平均ベクトルの検定については，ラオの U 統計量が知られており，非正規分布の下での分布関数の漸近展開が，Gupta, Xu and Fujikoshi (2006) によって与えられている．これらの結果を踏まえ，第 4 章では，k-ステップに対する尤度比を先行研究とは別の表記法により与える．さらに先行研究とは異なる導出法である摂動法により，分解した k 個の尤度比検定統計量に対する漸近展開をそれぞれ導出し，カイ二乗近似を個々に改良する修正検定統計量を新たに提案する．また，従来の修正尤度比検定統計量も与える．最後に，モンテカルロ・シミュレーションにより数値評価を行い，2 つの検定統計量のカイ二乗近似精度が良いことを示している．

第 5 章では，第 4 章の多標本問題への拡張として，平均ベクトルの同等性検定問題

$$H_0 : \boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)} = \cdots = \boldsymbol{\mu}^{(m)} \text{ vs. } H_1 : \text{not } H_0,$$

すなわち，one-way MANOVA について考える．ここで，$\boldsymbol{\mu}^{(\ell)}$ は第 ℓ 母集団 $(\ell = 1, 2, \ldots, m)$ の平均ベクトルで，各母集団の分散共分散行列は等しいとする．この問題について，完全データの下では，尤度比検定であるウィルクスの Λ 統計量などの議論がある（例えば，Siotani, Hayakawa and Fujikoshi (1985) などを参照)．一方，単調欠測データの下での議論は，2-ステップにおいて，Seko (2012) や Tsukada (2014) などがあり，Seko (2012) では尤度比検定統計量とその極限分布であるカイ二乗分布の上側パーセント点を用いた検定手順が与えられている．しかしながら，尤度比検定統計量の分布に対する漸近論についての議論はなく，第5章では，多標本問題に対する k-ステップ単調欠測データの下での尤度比検定統計量を与え，その分布の漸近論に基づく結果を与えている．第4章と同じように尤度比を分解し，それぞれの尤度比検定統計量の特性関数や分布関数の漸近展開を導出することによって，修正検定統計量及び修正尤度比検定統計量を提案している．この結果は，第4章で述べた1標本問題の場合の拡張となっている．そして，モンテカルロ・シミュレーションにより近似精度の数値評価を与えており，提案した2つの検定統計量のカイ二乗近似が，修正前に比べて非常に良くなっていることを示している．最後に，提案した手法を説明するために実データを基にした解析例を与えている．

論文審査の結果の要旨

本論文は，1標本問題，2標本問題，そして多標本問題における平均ベクトルの検定問題について述べたものであり，5つの章から構成される．

第1章は序論である．まず，本論文の研究の位置づけと研究の背景について述べている．多変量解析の検定問題の中で，平均ベクトルに関する検定問題は重要かつ基盤となる問題である．このような背景の下，データに欠損が生じた場合の平均ベクトルの検定問題に注目し，本論文の研究テーマと目的を与えている．

第2 章では，データが単調欠測という構造を持つ場合の1 標本問題における平均ベクトルの検定問題を取り扱っている．まず，欠測データの構造を3ステップ単調欠測データに絞り，本論文に使われる記号や定義などを述べ，それを基に平均ベクトルと分散共分散行列の最尤推定量について具体的に与え，平均ベクトルの検定を行う際に鍵となる T^2型検定統計量を与え，その分布の上側パーセント点の線形補間近似を与えている．さらに，3ステップ単調欠測データをより一般の場合のkステップ単調欠測データに拡張することに成功し，平均ベクトル成分間の多重比較法にも触れ，その近似同時信頼区間を与えることにも成功している．また，モンテカルロ・シミュレーションを通して，近似に対する数値評価を与えるとともに，先行研究による近似値との数値比較も行っている．

第3章では，第2章の内容である1標本問題を2標本問題に拡張することを考え，第 2 章と同様に 3 ステップ単調欠測データの場合から議論し，一般の場合のkステ

ップ単調欠測データの場合の結果を与えている．その中で，本論文の結果は母集団間ごとに異なる欠測パターンをもつ単調欠測データの場合にも対応しているものであることを示している．また，多標本問題における単調欠測データの場合の平均ベクトルと分散共分散行列の最尤推定量を明示的に与えており，新たな結果として特筆すべき点である．さらに，第 2 章と同様に，平均ベクトル間の多重比較法として，単調欠測データの場合の近似同時信頼区間を提案し，モンテカルロ・シミュレーションにより，これらの近似に対する数値評価を与えている．

第 4 章では，1 標本問題かつ単調欠測データの下で平均ベクトルの検定に対する尤度比検定統計量を導出し，カイ二乗近似を改善するバートレット補正による修正尤度比検定統計量を導出することに成功している．その導出法は仮説を分解することによって，尤度比をその分解した仮説の尤度比の積で与え，分解した尤度比ごとに特性関数の漸近展開を摂動法によって求め，分布関数の漸近展開を与えるものである．また分解した個々の尤度比に対する漸近展開式から，サンプル数に対してカイ二乗分布への収束が速くなる新たな修正検定統計量を提案するとともに，尤度比ごとの特性関数の漸近展開をまとめることによって，単調欠測データの場合の修正尤度比検定統計量を導出することに成功している．最後に，モンテカルロ・シミュレーションを行い，これらの近似に対する数値的評価を行い，非常に良い近似となっていることを示している．

第 5 章では，第 4 章の議論を多標本問題である多変量一元配置分散分析の場合に拡張した結果を与えている．導出法は第 4 章の 1 標本問題の場合と本質的には同じであるが，尤度比検定統計量の導出，尤度比の分解，特性関数の展開そして期待値の計算において，改めて行う必要があり，非常に複雑である．結果として，第 4 章と同様に，新たな修正検定統計量と修正尤度比検定統計量を与えることに成功している．また，モンテカルロ・シミュレーションを通して，これらの近似も非常によい近似であることを数値的に示すとともに，人工的ではあるが，実データを基に単調欠測データとなるデータを用いた解析例も与えている．

これらの結果は多変量統計解析の理論に対する大きな貢献である．よって，本論文は博士（理学）の学位論文として十分価値があると認める．