氏名（本籍）	小野澤　瑞　季（長野県）
学位の種類	博士（理学）
学位記番号	乙第 1241 号
学位授与の日付	2021 年 3 月 18 日
学位授与の要件	学位規則第 4 条第 2 項該当
学位論文題目	**Profile analysis and tests for mean vectors with two-step monotone missing data (2-ステップ単調欠測データのもとでのプロフィール分析と平均ベクトルに対する検定)**
論文審査委員	（主査）教授　瀬尾　　隆
	教授　石渡恵美子　　教授　宮岡　悦良
	准教授　黒沢　　健　　教授　眞田　克典

論文内容の要旨

本論文では, 2 ステップ単調欠測データをもつプロフィール分析の検定, 及び平均ベクトルの同等性検定について議論する. 実際の統計データ解析において, データが何らかの原因により全て得られず欠測を含んでいる場合が多い. 本論文で扱うデータは, 欠測のパターンが単調である 2 ステップ単調欠測データを仮定している.

第 1 章では, 平均ベクトルの検定問題やプロフィール分析に関する研究の背景を, 先行研究を踏まえ述べている.

第 2 章では, 2 ステップ単調欠測データのもとで 2 標本問題, 及び多標本問題におけるプロフィール分析に対する検定について議論する. プロフィール分析は, 複数個の母集団の平均に関する比較を行う仮説検定手法のひとつとして知られており, 平行性, 一致性, 平坦性の 3 つの仮説について検定が行われる. 母集団平均ベクトルを $\boldsymbol{\mu}^{(\ell)}(\ell = 1, \ldots, k)$ とするとき, 3 つの仮説は次のように表される.

$$H_P : \boldsymbol{C}\boldsymbol{\mu}^{(1)} = \cdots = \boldsymbol{C}\boldsymbol{\mu}^{(k)} \ \text{ vs. } \ A_P : \text{not } H_P,$$

$$H_L|H_P : \boldsymbol{1}'\boldsymbol{\mu}^{(1)} = \cdots = \boldsymbol{1}'\boldsymbol{\mu}^{(k)} \ \text{ vs. } \ A_L : \text{not } H_L|H_P,$$

$$H_F|H_P : \boldsymbol{C}(\boldsymbol{\mu}^{(1)} + \cdots + \boldsymbol{\mu}^{(k)}) \ \text{ vs. } \ A_F : \text{not } H_F|H_P.$$

ただし $\boldsymbol{1}_p$ は $(1,\cdots,1)'$ の p 次元ベクトル, $\boldsymbol{C}$ は $\boldsymbol{C}\boldsymbol{1}_p = \boldsymbol{0}$, rank $p-1$ の $(p-1)\times p$ 行列である．完全データの場合, これらの仮説に対し, Hotelling の T^2 型統計量による検定や尤度比検定等が与えられている．$N_1^{(\ell)}$ 個の p 次元観測ベクトル $\boldsymbol{x}_j^{(\ell)} \equiv (\boldsymbol{x}_{1j}^{(\ell)\prime}, \boldsymbol{x}_{2j}^{(\ell)\prime})'$ $(j = 1,\ldots,N_1^{(\ell)},\ \ell = 1,\ldots,k)$ が $N_p(\boldsymbol{\mu}^{(\ell)}, \boldsymbol{\Sigma})$ に従い, $N_2^{(\ell)}(= N^{(\ell)} - N_1^{(\ell)})$ 個の $p_1(= p - p_2)$ 次元観測ベクトル $\boldsymbol{x}_{1j}^{(\ell)}$ $(j = N_1^{(\ell)}+1,\ldots,N^{(\ell)},\ \ell = 1,\ldots,k)$ が $N_{p_1}(\boldsymbol{\mu}_1^{(\ell)}, \boldsymbol{\Sigma}_{11})$ に従うとする．ただし

$$\boldsymbol{\mu}^{(\ell)} = (\boldsymbol{\mu}_1^{(\ell)\prime}, \boldsymbol{\mu}_2^{(\ell)\prime})', \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

とし, $\boldsymbol{\mu}_j^{(\ell)} (j = 1, 2)$ は $\boldsymbol{\mu}^{(\ell)}$ の p_j 次元分割ベクトル, $\boldsymbol{\Sigma}_{jm}$ $(j,\ m = 1, 2)$ は $\boldsymbol{\Sigma}$ の $p_j \times p_m$ 分割行列とする．まず, 2 標本問題 $(k = 2)$ における 2 ステップ単調欠測データのもとでのプロフィールに対する検定統計量をそれぞれ

$$T_{Pm}^2 = (\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)})'\boldsymbol{C}' \left\{ \boldsymbol{C} \begin{pmatrix} \dfrac{N}{N^{(1)}N^{(2)}}\widehat{\boldsymbol{\Sigma}}_{11} & \dfrac{N}{N^{(1)}N^{(2)}}\widehat{\boldsymbol{\Sigma}}_{12} \\ \dfrac{N}{N^{(1)}N^{(2)}}\widehat{\boldsymbol{\Sigma}}_{21} & \widehat{\mathrm{Cov}}[\widehat{\boldsymbol{\mu}}_2^{(1)}] + \widehat{\mathrm{Cov}}[\widehat{\boldsymbol{\mu}}_2^{(2)}] \end{pmatrix} \boldsymbol{C}' \right\}^{-1} \boldsymbol{C}(\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)}),$$

$$T_{Lm}^2 = (\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)})'\boldsymbol{1}_p\{\boldsymbol{1}_p'\widehat{\Xi}\boldsymbol{1}_p\}^{-1}\boldsymbol{1}_p'(\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)}),$$

$$T_{Fm}^2 = (\boldsymbol{C}\widehat{\boldsymbol{\mu}})'\{\boldsymbol{C}\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}})\boldsymbol{C}'\}^{-1}(\boldsymbol{C}\widehat{\boldsymbol{\mu}})$$

として与えた．ただし $\widehat{\boldsymbol{\mu}}^{(\ell)}$ $(\ell = 1, 2)$, $\widehat{\boldsymbol{\Sigma}}$ はそれぞれ, $\boldsymbol{\mu}^{(\ell)}$, $\boldsymbol{\Sigma}$ の最尤推定量であり, Anderson and Olkin (1985) で与えられている．また, $\widehat{\boldsymbol{\mu}}$ は $\boldsymbol{\mu} = \boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)}$ の最尤推定量である．完全データの場合, 検定統計量はホテリングの T^2 型統計量で与えられ, その分布は正確に F 分布に従うことが知られている．しかし, データが欠測している場合, 検定統計量の正確な分布を得ることが難しい．そこで, Chang and Richards (2009), Seko et al. (2012) 等は, 平均ベクトルの同等性検定に対する検定統計量の近似上側パーセント点を与えている．本研究では, Seko et al. (2012) で提案されている完全データの F 分布をもとにした線形補間による T^2 型統計量の近似上側パーセント点を, 2 標本問題におけるプロフィール分析に適応している．

さらに, $k \geq 3$ のときの k 標本問題における平行性仮説検定について考えている．検定統計量は, 尤度比検定統計量として与えられ, その極限分布は χ^2 分布に従う．しかし, 標本の大きさが小さいときにはよい近似ではないので, Bartlett 補正による修正尤度比をもとにし, 線形補完を行った検定統計量を提案している．

また, 1,000,000 回のモンテカルロ・シミュレーションにより提案した近似上側パーセント点とその極限分布の上側パーセント点を使った type I error を数値的に比較し, 提案した近似上側パーセント点がよい近似になっていることを示した．

第 3 章では, 2 ステップ単調欠測データのもとでの 1 標本問題, 及び 2 標本問題の平均ベクトルの同等性検定について考えている．この問題について，1 標本問題では, Krishnamoorthy and Pannala (1999), 2 標本問題では, Yu et al. (2006) により, 検定統計量が与えられ, T^2 型検定統計量に対する F 分布を用いた近似上側パーセント点が提案されている．また, Yagi et al. (2019, 2021) では，これらの検定統計量の分布関数の漸近展開が近似的に与えられ, それを利用した近似上側パーセント点や Bartlett 補正に基づく変換統計量が提案されている．本研究では, 先行研究とは異なる新たな検定統計量を提案する．第 ℓ 群の観測データ行列 $\boldsymbol{X}^{(\ell)}$ を

$$\boldsymbol{X}^{(\ell)} = \begin{pmatrix} \boldsymbol{X}_{11}^{(\ell)} & \boldsymbol{X}_{12}^{(\ell)} \\ \boldsymbol{X}_{21}^{(\ell)} & * \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_{1(12)}^{(\ell)} & \\ \boldsymbol{X}_{21}^{(\ell)} & * \end{pmatrix}$$

と表す ($\ell = 1, 2$)．ただし "$*$" は欠測部分を表す．また $\boldsymbol{X}_{11}^{(\ell)}$ は $n_1^{(\ell)} \times p_1$ ブロック行列, $\boldsymbol{X}_{12}^{(\ell)}$ は $n_1^{(\ell)} \times p_2$ ブロック行列, $\boldsymbol{X}_{21}^{(\ell)}$ は $n_2^{(\ell)} \times p_1$ ブロック行列とし

$$\begin{aligned} \mathrm{vec}(\boldsymbol{X}_{1(12)}^{(\ell)\prime}) &\sim N_{n_1^{(\ell)}p}(\mathrm{vec}(\boldsymbol{\mu}^{(\ell)}\boldsymbol{1}_{n_1^{(\ell)}}'), \boldsymbol{I}_{n_1^{(\ell)}} \otimes \boldsymbol{\Sigma}), \\ \mathrm{vec}(\boldsymbol{X}_{21}^{(\ell)\prime}) &\sim N_{n_2^{(\ell)}p_1}(\mathrm{vec}(\boldsymbol{\mu}_1^{(\ell)}\boldsymbol{1}_{n_2^{(\ell)}}'), \boldsymbol{I}_{n_2^{(\ell)}} \otimes \boldsymbol{\Sigma}_{11}) \end{aligned}$$

と仮定する．さらに $\overline{\boldsymbol{x}}_{(12)1}^{(\ell)}$, $\boldsymbol{S}_{(12)1}^{(\ell)}$ をそれぞれ $\boldsymbol{X}_{(12)1}^{(\ell)}$ の標本平均ベクトル, 不偏標本分散共分散行列とし, $\overline{\boldsymbol{x}}_{1(12)}^{(\ell)}(= (\overline{\boldsymbol{x}}_{11}^{(\ell)\prime}, \overline{\boldsymbol{x}}_{12}^{(\ell)\prime})')$, $\boldsymbol{S}_{1(12)}^{(\ell)}$ を $\boldsymbol{X}_{1(12)}^{(\ell)}$ の 標本平均ベクトル, 不偏標本分散共分散行列とする．ただし

$$\begin{pmatrix} \boldsymbol{X}_{11}^{(\ell)} \\ \boldsymbol{X}_{21}^{(\ell)} \end{pmatrix} = \boldsymbol{X}_{(12)1}^{(\ell)}, \quad \boldsymbol{S}_{1(12)}^{(\ell)} = \begin{pmatrix} \boldsymbol{S}_{1(12),11}^{(\ell)} & \boldsymbol{S}_{1(12),12}^{(\ell)} \\ \boldsymbol{S}_{1(12),12}^{(\ell)\prime} & \boldsymbol{S}_{1(12),22}^{(\ell)} \end{pmatrix}.$$

まず初めに, "(ℓ)" を取り除いた 1 標本問題について考える．仮説

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

に対する 簡便な T^2 型検定統計量は

$$Q = Q_1 + Q_2$$

と表される．ただし

$$Q_i = N_{3-i}\widehat{\boldsymbol{\eta}}_i'\widehat{\Delta}_{ii}^{-1}\widehat{\boldsymbol{\eta}}_i, \quad N_i = \sum_{j=1}^{i} n_j, \quad i = 1, 2,$$
$$\widehat{\boldsymbol{\eta}}_1 = \overline{\boldsymbol{x}}_{(12)1}, \quad \widehat{\boldsymbol{\eta}}_2 = \overline{\boldsymbol{x}}_{12} - \boldsymbol{S}_{1(12),21}\boldsymbol{S}_{1(12),11}^{-1}\overline{\boldsymbol{x}}_{11},$$
$$\widehat{\boldsymbol{\Delta}}_{11} = \frac{N_2 - 1}{N_2}\boldsymbol{S}_{(12)1}, \quad \widehat{\boldsymbol{\Delta}}_{22} = \frac{N_1 - 1}{N_1}\left[\boldsymbol{S}_{1(12),22} - \boldsymbol{S}_{1(12),21}\boldsymbol{S}_{1(12),11}^{-1}\boldsymbol{S}_{1(12),12}\right]$$

である (Krishnamoorthy and Pannala (1999) 参照)．Yagi et al. (2019) では，Q_1 と Q_2 の漸近展開を与えることによって，Q の帰無分布の近似的な漸近展開が与えられている．Q_1 と Q_2 は独立ではないが，漸近的に独立であることを利用して Q の近似的な帰無分布を導出している．そこで本研究では, Q_2 の変わりに Q_1 と独立な R_2 を使った新たな検定統計量

$$Q_M = Q_1 + R_2$$

を提案している．ただし

$$R_2 = \frac{Q_2}{1 + Q_{2d}}, \quad Q_{2d} = \frac{N_1}{N_1 - 1}\overline{\boldsymbol{x}}_{11}'\boldsymbol{S}_{1(12),11}^{-1}\overline{\boldsymbol{x}}_{11}.$$

Q_1 と R_2 が独立なため, 検定統計量の正確な漸近展開を導出することができる．この Q_1 と R_2 に対してそれぞれ漸近展開を与えることによって，Q_M の帰無分布の漸近展開を与えている．さらにその結果から, χ^2 近似を改良する Bartlett 補正に基づく変換統計量をいくつか提案している．また, 提案した近似上側パーセント点や変換統計量の近似精度の評価を, 1,000,000 回のモンテカルロ・シミュレーションにより数値的に行い, 比較している．

また, 2 標本問題について, 仮説

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

を考えている．1 標本問題のときと同様, 検定統計量を 2 つの独立な統計量の和で与えている．2 標本問題についても, それぞれの統計量の漸近展開を導出し, 検定統計量の帰無分布の漸近展開を与えている．さらに, いくつかの変換統計量を提案し, 提案した近似上側パー

セント点や変換統計量の近似精度の評価を, 1,000,000 回のモンテカルロ・シミュレーションにより数値的に行い, 比較している.

論文審査の結果の要旨

本論文は，多変量解析におけるプロフィール分析，および，１標本問題，２標本問題に対する平均ベクトルの検定問題について述べたものであり, 4 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．具体的には，多変量解析の検定問題の一つである平均成分の比較を行うプロフィール分析について先行研究の紹介を行うとともに, 本論文で取り扱う単調欠測データの下での検定問題等の先行研究を与え，本論文の研究テーマと目的を与えている.

第２章では, プロフィール分析について, データが欠損していない完全データの下での従来の検定法について紹介している．具体的には，３つの検定問題（平行性、一致性、平坦性）について，それぞれの検定統計量とその帰無分布を系統的に与えている．その上で，母集団が２つの場合のプロフィール分析について，データが単調欠測という構造を持つ場合の上記の３つの検定問題を取り扱っている．特に，欠測データの構造が 2 ステップ単調欠測データの場合に絞り, 平均ベクトルと分散共分散行列の最尤推定量を具体的に与え，プロフィール分析に用いられる検定統計量である T^2型の検定統計量を与えている．さらに完全データの際に用いられる F 分布を利用して, 線形補間による近似上側パーセント点を与えている.

また，この問題を３つ以上の母集団である多標本問題に拡張することを考え，プロフィール分析の一つである平行性仮説検定に注目し，尤度検定統計量を導出している．さらに完全データの下でのこの問題は，尤度比検定統計量のカイ二乗近似を改良する修正尤度比検定統計量が提案されているが，２ステップ単調欠測データの場合, このような修正尤度比検定統計量を導出することは容易ではなく，本論文では，尤度比検定統計量の近似上側パーセント点として, バートレット補正項に線形補間近似を適用した新たな近似を提案している. 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い，近似精度の数値的評価を与えている.

第３章では，第２章で取り扱った 2 ステップ単調欠測データの下で，平均ベクトルの T^2型検定統計量に注目し，帰無分布の漸近展開近似を改良する新たな検定統計量を提案している．具体的には，まず１標本問題における平均ベクトルの検定について，従来の T^2型検定統計量は２つの統計量に分解されることから，分解された後半

部分の統計量を前半部分の統計量とは正確に独立になるものに置き換えることによって，新たな検定統計量を提案し，この帰無分布の漸近展開をより精密な形で導出することに成功している．さらにこの漸近展開近似の構造を用いて，この検定統計量のカイ二乗近似をオーダーの意味で改善するバートレット型変換統計量をいくつか提案している. これは導出した検定統計量そのものを変換するものや分解した統計量を個々に変換して構成するものとなっている. また同様の議論を２標本問題へ拡張することにも成功している．２標本問題については，先行研究にある１標本問題に提案されている自由度調整法によるＦ近似法を２標本問題にも応用し，２標本問題に対する従来の T^2型検定統計量のＦ近似法も与えている．最後にいくつかのパラメータに対するモンテカルロ・シミュレーションを通して，本論文で与えた近似に対する数値評価を与え，先行研究による近似値との数値比較を行っている．第４章は本論文の全体のまとめを述べている．

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．以上のことから，本論文は博士（理学）の学位論文として十分価値があるものと認める．