

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年 6月 6日現在

機関番号：62603

研究種目：挑戦的萌芽研究

研究期間：2010～2012

課題番号：22650059

研究課題名（和文） 意図的な欠測を伴う調査設計の指針開発に関する研究

研究課題名（英文） Research on strategies for survey design with planned missing data

研究代表者

前田 忠彦 (MAEDA TADAHIKO)

統計数理研究所・データ科学研究系・准教授

研究者番号：10247257

研究成果の概要（和文）：社会調査において、意図的な欠測部分を作って調査票を分冊化することによって、項目間の相関等をより多くの項目組み合わせについて検討するための調査デザインについて検討した。項目群の組み合わせについて一般の場合と因子分析的な場合を想定した上で、どのような分冊法をとるべきかについて、効用評価の観点と統計的推定の精度の観点から考察して、デザインの指針を提案した。

研究成果の概要（英文）：Strategies for split questionnaire survey design, where some portion of the items are intentionally set to missing in specific forms, thus enabling the estimation of correlations among more combinations of questionnaire items, were investigated in this study. We assumed two settings on group of items, the one is the case of general items and the other is the case of factor-analytic items, and proposed a evaluation of design strategies for these cases, by considering a utility function or precision of statistical estimation.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	600,000	180,000	780,000
2011年度	1,200,000	360,000	1,560,000
2012年度	1,000,000	300,000	1,300,000
年度			
年度			
総計	2,800,000	840,000	3,640,000

研究分野：情報学

科研費の分科・細目：統計科学

キーワード：社会調査，測定デザイン，調査票設計，不完全データ，効用関数，因子分析，分冊型調査票

1. 研究開始当初の背景

意図的な欠測を伴うデザインは、サンプルを何ブロックか分けた上で、共通測定項目群とブロックごとに独自の測定項目群を組み合わせる測定を行うことで、一部の変数についてはサンプルサイズを失う代わりに、相関をとることができる変数の組み合わせを増や

す技法である。

この方法について2005年以降くらいに、欧米を中心に研究例が少しずつ発表されて関心が高まっていると言われており、本研究との表題ともなっている **planned missing data design** という呼ばれ方が一般的であるが、実際には全く新しい話題というわけでは

なく、split questionnaire design など他の名称で技術的には実際の調査にも用いられてきた話題でもある。ところが、そのデザインの仕方あるいはデザインの良さの評価について、統一的な観点から研究対象になってきたと言える状態ではなかった。

研究開始時点では、研究代表者自身は、効用関数のような考え方を取り入れるべきではないかという萌芽的なアイデアを持っており、このアイデアを、単純に調査項目のセットが複数の組に分けられるという状況だけではなく、もう少し多様な調査票設計の文脈において展開することを本研究は意図したものであった。

2. 研究の目的

本研究では、統計的社会調査において、この意図的な欠測を組み入れた調査設計により、相関関係を検討できる調査項目の組み合わせを増やす方法の効用評価法を研究し、調査設計のための支援アルゴリズムを開発し、将来の実用化に備えた知見を蓄積することを目的とする。

こうしたアルゴリズムの開発により、さまざまな統計的な調査において効率の良い測定デザインの普及に資することを目指す。

ここで多様な調査票設計の文脈において展デザインの良さを評価する文脈を、(1)ばらばらの質問項目が複数組に分けられるという文脈だけではなく、(2)複数の組が因子分析的に複数組に分かれている場合を含めることにする。

これら(1)(2)の状況設定の例は4の研究成果の項で述べる。

3. 研究の方法

本研究でとる方法は理論的考察とそれに基づく数値実験、シミュレーション(乱数発生に基づく)解析モデル間の性能比較、実際のデータの取得と解析に基づく実例提供である。実データに基づく解析は付随的な成果であり、4の研究成果では前2者に基づく検討を中心に述べる。

以下では(意図的な欠測を伴って)設計された複数の版の調査票を「分冊型の調査票」(split questionnaire)と手短かに呼ぶことにする。

4. 研究成果

(1) 分冊型調査票のデザインの効率性に関する効用関数アプローチ：一般の質問項目組み合わせの場合

主に平成22年度中の検討成果である。分冊型調査票のデザインの効率性を評価するための効用関数アプローチによる理論の整備と簡単な数値実験を行った。また過去の類似の調査票設計による調査の実例の調査と

それに基づく実用的な調査票フォーム数の検討などの考察を行った。

デザイン名	フォーム	項目数	qx	q0	q0	q0
		Effective n	X	A	B	C
項目選択型1フォーム計画	1	n	■	■	■	□
全項目2フォーム型計画	1	n/2	■	■	■	□
	2	n/2	■	■	□	■
全項目3フォーム型計画	1	n/3	■	■	■	□
	2	n/3	■	■	□	■
	3	n/3	■	□	■	■

図1 分冊型の調査票のデザイン例

ここで検討したのは図1のような状況である。

図1は、例えば項目群相対的な重要度が $X \gg A \geq B \geq C$

と評価できるような項目群が用意されたことを示す。 \gg は左が右に比べて明らかに重要、 \geq は左が右に比べて同程度かそれ以上に重要、ということの意味している。Xは明らかに他の項目群ABCよりも重要度が高く、調査実施に際して決して省略のできない項目群(例えば基本属性項目)、Aは調査の主要な目的変数であって、相対的にはBCよりも重要度が高い、BCもこの順に重要度が高いことを意味している。

考えるべき問題は、対象者の負担を考えて全項目を調査票に収録できないとき、例えば項目数を全体で $(q_x + 2q_0)$ 個に押さえなければいけない状態で、どのように調査票を設計するのが良いかということである。図には3つの計画が示されており、■を入れたセルの測定は行い、□のセルを欠測とする。

図でもXは必ず全ての計画に含まれている。一番上の「項目選択型1フォーム計画」 $B \gg C$ という判断がついた場合に項目群Cを調査には使用せず、XABのみで1種類の調査票を用いるデザインであるから分冊型調査票とは呼ばない。通常はこのように重要度の序列にしたがって、最も重要でない項目群を省き1種類の調査票だけを作成して調査を行う。

二番目の「全項目2フォーム型計画」は、BCの重要度の序列がつきにくい、 $A \gg B = C$ である場合に、対象者を折半してハンス数ずつ(サンプルサイズ $n/2$ ずつ)の対象者にフォーム1: XAB, フォーム2: XACの2分冊で調査を実施するものである。

三番目の「全項目3フォーム型計画」は、ABC間の重要度の序列がつけにくい場合、つまり $X > A = B = C$ の場合に、対象者を三等分して、フォーム1: XAB, フォーム2: XAC, フォーム3: XBCの3分冊を割り当てて調査を実施する方式である。

フォーム2と3ではABCに含まれる項目の周辺分布を知ることができるだけでなく、項目群ABCに含まれる項目相互の同時分布を知る（相関係数等を計算する）ことができる。しかし現実的な場面での問題は項目間の重要度の序列が明快に“>>”という記号が表すように決定しにくいという点であり、こうした状況下では2フォーム計画がよいのか3フォーム計画が良いのかという点を評価しにくい。このことへの対応を効用関数を導入して検討したということが、本研究の第1の成果である。

まず効用評価法として、効用関数を導入した評価方法を理論的に検討した。考察した効用関数は、

- ・ 推定可能な「パラメータ個々の効用」の総和をとる
- ・ 「個々の効用」＝「パラメータの価値」×「検定力」で定まるものとする
- ・ 「パラメータの価値」には、設計者が特定の変数（調査項目）をどの程度重要と思うか、パラメータ（平均、分散、相関係数など）のうちどのパラメータを推定することを重視するか、などのような主観的な要因を導入する。

という特徴を持たせた。要素の積を取る部分には別の考え方もあり得るが、結論的には、調査票設計の支援指針として機能しうるとの見通しを得た。

上記の形の効用関数を仮定した場合の数値実験では、

- ① ある一定のサンプルサイズの箇所を見ると、特定のデザインが一貫して有利になることがある。やはり n をいろいろ変えたときの情報が重要である（図2）。
 - ② 3フォームを用意するデザインの効用が、より少ないフォーム数のデザインの効用を上回るためには、全体のサンプルサイズが大きいことが重要である（図3）
- ことなどが判明した。

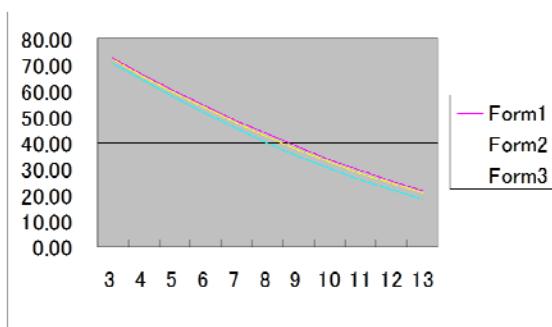


図2 効用関数の例：

（解説）総項目数が40、サンプルサイズが300のように一定で、横軸に項目群ABCに含まれる項目数をとって変化させた場合の効用関数の例：1フォームデザインが（Form1のライン）常に他のデザインを上回っている。

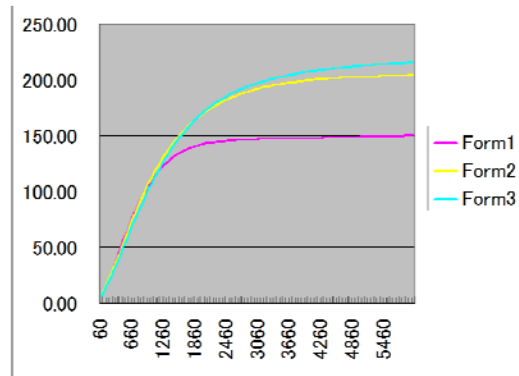


図3 効用関数の別の例：

（解説）：総項目数と項目群ABC内の項目数等を定めたときの、横軸（サンプルサイズ）を変化させた場合の効用関数。サンプルサイズが小さい部分では、1フォーム計画、中程度のところでは2フォーム計画、サンプルサイズ2000程度以上では3フォーム計画の値が大きくなる。

以上のように、効用関数内で与えるパラメータの重要度や検定力の設定に関して分析者が指定する（その値を変化させて、いくつかの条件設定間での比較を行う）ことによって、どのデザインを選択すべきかの指針が得られることになる。これが本研究の一つの成果である。

(2) 分冊型調査票のデザインの効率性に関する統計的評価：因子分析モデルが設定できる場合

項目	案1			案2			案3		
	1_1	1_2	1_3	2_1	2_2	2_3	3_1	3_2	3_3
a1	1	1		1	1		1	1	1
a2	1	1		1	1		1	1	1
a3	1	1		1		1	1	1	1
a4	1	1		1		1	1		
a5	1	1			1	1		1	
a6	1	1			1	1			1
b1	1		1	1	1		1	1	1
b2	1		1	1	1		1	1	1
b3	1		1	1		1	1	1	1
b4	1		1	1		1	1		
b5	1		1		1	1		1	
b6	1		1		1	1			1
c1		1	1	1	1		1	1	1
c2		1	1	1	1		1	1	1
c3		1	1	1		1	1	1	1
c4		1	1	1		1	1		
c5		1	1		1	1		1	
c6		1	1		1	1			1

図4 因子分析的な設定における18項目の3つの分冊に対する配置案

続いて、項目群ABCが因子分析モデルにし

たがケースについて、因子内の項目を分冊間にどのように配置すべきかという問題に関する検討を行い、シミュレーションによって配置方法の良さを評価する方法についての成果を得た。

考えたケースは3因子内の項目（候補）が各6項目、計18項目の項目を複数のフォームに振り分ける分冊方式間の比較である。（図4）。

この図では、案1は、因子aの項目はフォーム1とフォーム2に集め、(1_1, 1_2の列に1を入れてある：以下説明略)、因子bの項目はフォーム1とフォーム3に、因子cの項目はフォーム2とフォーム3に全て含める形になる。各フォーム(1_1, 1_2, 1_3は個別には2因子の項目計12個で構成される。

案2は、各分冊には3因子から4つずつの項目を含ませ、因子内の6項目は3つのフォーム内のうち2フォームでのみ測定されるようにバランスをとったデザインである。案3は各因子内で重要な3項目が設定できるが、(因子a b c内の項目1~3)、残りの3項目は比較的重要度が低いことが分かっている項目という設定である。この場合の項目の重要度は因子負荷量で設定される。

以上のような問題設定の下で、本研究ではシミュレーション研究により、どの案で測定を行うのが良いかを統計的な立場から評価した。シミュレーション研究では、因子間相関係数、分冊方法・欠測処理方法・モデルのパラメタなどを独立変数とした。欠測処理方法（分析方法）として採用したのは、欠測に関して完全情報最尤推定を行う方式（FIML法）と、マルコフ連鎖・モンテカルロ法を用いて欠測を補完しながら推定を行う方法（MCMC法）である。

主な成果は次の通り。

[a]分冊方式のうち、特定因子の項目を全て欠測にして下位尺度を分散させる分冊方式（案1）では、和得点のバイアスに関して全体として有利になる一方で（図5）、因子間相関の推定については不利になる。

尺度の和得点の絶対バイアス(分冊方法・相関)

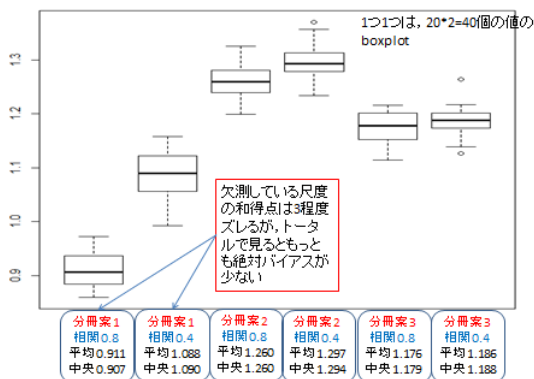


図5 分冊案1が尺度の和得点の推定に関して有利であることを示す結果

因子間相関の絶対バイアス(因子負荷量)

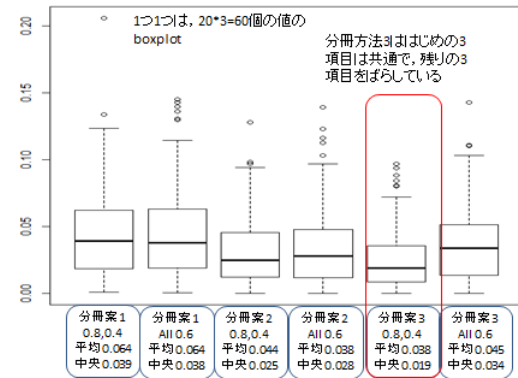


図6 案2に比べて案3が推定精度が有利になることを示したシミュレーション結果

[b]下位尺度内での欠測を全ての分冊に振り分ける方式の間（案2と案3）では、結果に大差はないが、重要な項目を事前に特定できる条件（案3）では、推定精度が良くなる面がある、欠測値処理法間の差もわずかにある。（以上図6）。

以上のシミュレーションからの暫定的な結論としては、次のような指針が得られたと考えられる。

- 因子間相関の絶対バイアスに関して
 - 分冊方法1は避けた方が良い
 - 分冊方法2と3では違いが生じない
 - 個人ごとの尺度の和得点の絶対バイアスの平均に関して
 - 全体的には分冊方法1が最も良いが、完全に欠測している尺度の和得点は大きなバイアスが残る
 - 分冊方法案3の方が2よりも良い（外せない項目＝因子負荷量の大きな項目がある／予め分かっている場合には、案3の方が有利になる）
 - 因子間相関の真値が大きい場合の方が因子間相関、尺度の和得点ともに絶対バイアスが小さい
 - FIMLとMCMCで違いはほぼないが、FIMLの方が若干良い。ただし、尺度の和得点も推定したいときにはMCMCしか使えない。
- などである。

本研究での(2)項の成果を生かし、研究分担者尾崎は新たな研究課題で平成25年度からの科学研究費を受けて研究内容を発展させる予定である。

(3) 実験調査の実施による実データ解析による成果を得た。

平成24年3月に実験的ウェブ調査を実施した。ウェブ調査はA社とB社の2社に対し計3条件(A1, A2, B1条件とする)で実施

し、(2) で言及した因子分析的な項目準備状況を想定しており、効用評価の観点から分冊調査票を設計する際の実例提示のために利用する。この目的の他に、項目の提示順序のランダム化の効果(A1 と A2 の比較)、実施者間の差(A1 と B1 の比較)を行うことができる設計とした。人間関係観や信頼感、不安感に関する実質科学的な観点での分析も行うことができる、例えば同時期の他の全国規模社会調査とも比較可能である。単純集計レベルでも、例えば人のためになるより自分の好きなことをしたい、という項目の支持が高いなど、ウェブ調査の回答者特性を反映したと思える結果が得られた。

また、一般的な信頼と用心深さという心理特性に関する多母集団同時解析を行い、用心深さ一般的信頼感へのパスに年齢層間で差がある、等の実質科学的な成果が得られた。

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 1 件)

- ① Nakamura, T., Maeda, T. et al. A Study of the Japanese national character: The twelfth nationwide survey (2008) —English Edition—, 統計数理研究所調査研究レポート, No.102, 2011 年, pp.1-155. 査読無.
<http://www.ism.ac.jp/editsec/kenripo/index.html>

[学会発表] (計 6 件)

- ① 尾崎幸謙 (2012). 分冊型の調査票に関する統計学的評価, 第 54 回数理社会学会大会. 2012 年 8 月 31 日, 横浜市
- ② 大崎裕子・前田忠彦 (2012). 道徳的信頼と用心深さ—多母集団同時分析を用いた一般的信頼の下位側面の検証—, 第 54 回数理社会学会大会. 2012 年 8 月 31 日, 横浜市.
- ③ Ozaki, Koken (2012) Complexity Index in Cognitive Diagnosis Models. The 77th annual meeting of the Psychometric Society, 2012 年 7 月 9~12 日, Lincoln, U. S. A.
- ④ 前田忠彦 (2012). 評定尺度法の調査項目における反応の構えについて—SSP-I2010 調査の事例から—, 第 53 回数理社会学会. 2012 年 3 月 15 日, 鹿児島市.
- ⑤ 前田忠彦・尾崎幸謙 (2012) 分冊型の調査票設計に関する一考察, 第 53 回数理社会学会. 2012 年 3 月 14 日, 鹿児島市.

6. 研究組織

(1) 研究代表者

前田 忠彦 (MAEDA TADAHIKO)

統計数理研究所・データ科学研究系・准教授

研究者番号：10247257

(2) 研究分担者

尾崎 幸謙 (OZAKI KOKEN)

統計数理研究所・データ科学研究系・助教
研究者番号：50574612

(3) 連携研究者

なし