

質問紙調査における無回答の発生過程およびその個人差について

島田 大祐 (名古屋大学 大学院情報学研究科, shimada.daisuke@d.mbox.nagoya-u.ac.jp)

片平 健太郎 (名古屋大学 大学院情報学研究科, katahira.kentaro@b.mbox.nagoya-u.ac.jp)

The process of item nonresponse in a questionnaire survey and its individual differences

Daisuke Shimada (Graduate School of Informatics, Nagoya University, Japan)

Kentaro Katahira (Graduate School of Informatics, Nagoya University, Japan)

Abstract

The relationship between the occurrence of item nonresponse and individual characteristics in questionnaire surveys can cause bias in parameter estimation. The purpose of this study was to determine whether individual differences truly exist in the occurrence of item nonresponse and, if so, the extent to which individual differences can predict the occurrence of item nonresponse. Based on the assumption that the number of item nonresponses for each respondent obeys a Poisson distribution, two models were examined for the number of item nonresponses for each respondent in the data for 300 items measuring personality ($N = 307,313$). The negative binomial distribution, which is derived by assuming that the expected value of the number of item nonresponses follows a gamma distribution, fits the distribution of the number of item nonresponses better than the model that assumes no difference in the expected value of the number of item nonresponses (Poisson distribution). Simulations based on the estimated negative binomial distribution showed that the occurrence of item nonresponse can be predicted by the individual difference variable with an accuracy of up to approximately 76%. Finally, a logistic regression model was used to search for individual difference variables that predicted the occurrence of item nonresponse.

Key words

item nonresponse, individual difference, poisson distribution, negative binomial distribution, personality

1. 問題と目的

質問紙調査において、データの欠測は一般的に生じる現象である。質問紙調査における回答データの欠測は、各回答者における無回答 (nonresponse) によって生じる。無回答は、全項目無回答 (unit nonresponse) と一部項目無回答 (item nonresponse) の2種類に大別される (高橋・阿部・野呂, 2015)。全項目無回答は、ある回答者がすべての質問項目において回答しないという行動を指す。一方、一部項目無回答は、ある回答者が質問項目単位で回答をしないという行動を指す。本研究では、一部項目無回答を検討対象とするが、簡略化のため、以後の本文中における「無回答」は一部項目無回答を指すものとする。

質問紙調査において欠測データは頻繁に生じることが報告されている。Eekhout, de Boer, De Vet, & Heymans (2012) によると、2010年に有力雑誌で出版された疫学研究のうち約76%で欠測データが報告されており、その対処法として最も支配的なものはリストワイズ除去 (listwise deletion) であった。リストワイズ除去とは、1項目以上の欠測データがある回答者のデータを分析対象外とする欠測データ処理法である。心理学における質問紙調査でも同様の傾向がみられる。Van Ginkel, Sijtsma, Van der Ark, & Vermunt (2010) によると、パーソナリティ心理学における有力雑誌において1995年から2007年までに出版された論文の半数以上でデータの欠測が報告されており、最もよく用いられた欠測補完の方法はリストワイズ除去であった。

しかし、リストワイズ除去には大きく2つの問題点がある (Little & Rubin, 2019)。まず、サンプルサイズの低下による検出力の低下が挙げられる。欠測値が含まれる回答者データを分析対象外とすると、本来よりもサンプルサイズが小さくなる。そのためリストワイズ除去を実施すると、真の効果を検出できない確率が高くなる。

第二の問題点は、推定結果にバイアスが生じることである。データの欠測が何らかの変数と関連を持つ場合、リストワイズ除去によって分析対象外となる回答データにばらつきが生じる。具体的に言えば、無回答を起こしやすい個人特性を持つ回答者のデータは、そうでない回答者データよりも分析対象外になりやすい。この結果、検討対象のパラメータ推定にバイアスが生じる。このように、データの欠測が何らかの変数に依存して生じる場合、リストワイズ除去を行うと、推定のバイアスが問題になる。

第二の問題点に関して、質問紙調査における無回答の発生と関連を持つ要因が探索されてきた。それは三つのカテゴリに大別される。第一に、データ収集に関する要因が挙げられる。これには、調査実施者 (West & Blom, 2016; West & Olson, 2010)、対面調査かオンライン調査かといったデータ収集の方法 (Denscombe, 2009; de Rada & Dominquez-Alvarez, 2014; Smith, Roster, Golden & Albaum, 2016) などの要因が含まれる。第二の要因は、質問紙の構成に関するものである。これに関しては、質問内容 (Groves et al, 2006) や、“わからない”などの中立的な選択肢の有無 (Krosnick, Holbrook, Berent, Carson, Michaelm Kopp, Mitchell, Presser, Ruud, Smith, Moody, Green, & Conaway, 2002) といった変数と無回答の発生との関連が検討されてきた。第三の要因は、回答者の特性・属性

である。回答者が持つ何らかの個人差変数が無回答の発生と関連を持つ場合、解析結果にバイアスが生じる可能性がある。本研究では、第三の要因である、回答者の属性・特性に焦点を当てる。

個人差変数に関する先行研究においては、高齢者 (Tsiampalis & Panagiotakos, 2020; Kutschar, Weichbold & Ostebrink, 2019; Caan, Hiatt & Owen, 1991) や性別 (Tsiampalis & Panagiotakos, 2020; Kutschar et al., 2019; Guadagnoli & Cleary, 1992)、教育水準 (Tsiampalis & Panagiotakos, 2020; Guadagnoli & Cleary, 1992) といった回答者の属性と無回答の発生の関連が検討されてきた。また、回答者の健康状態 (Tsiampalis & Panagiotakos, 2020)、認知機能の低下 (Kutschar et al., 2019; Fisher, Burgio, Thorn & Hardin, 2006) は、無回答の発生を予測する変数であることも示されてきた。これらの先行研究は、無回答の発生という現象が、回答者の個人差変数に依存することを示唆する。

以上の先行研究を踏まえて、本研究では二つの目的を設定する。本研究の第一の目的は、無回答の発生過程に個人差が真にあるかどうかを明らかにすることである。以上に挙げた個人差変数に関する先行研究では、各個人差変数の効果を個別に検討した。しかし、観測された無回答の発生に関する分散が、本当に個人差変数によって説明できるのか、それとも無回答という現象が持つランダム性 (e.g. 個人内変動) によって生じた結果であるのかについて、明らかでない。そのため、本研究では個人差変数の種類を限定せずに、無回答の発生に関する個人差について基本的な検討を行う。これによって、無回答の発生過程が個人差変数に依存するかどうかを明らかにすることを試みる。これを明らかにできれば、無回答の発生というアウトカムが、個人差変数にどの程度依存するのか、個人差変数からどの程度の精度で予測できるのか、といった一般的な議論が可能になると期待される。

第一の目的を達成するために、本研究では統計モデリングの手法を用いる。類似した現象に対して統計モデリングのアプローチを採用した研究としては、Okada, Vandekerckhove & Lee (2018) が挙げられる。Okada et al. (2018) は、Web 上で実施された質問紙調査において回答を途中でやめてしまう行動、すなわち回答からの離脱行動 (quitting behavior) を検討するために、統計モデリングのアプローチを取った。そこでは、打ち切り幾何モデル (censored geometric model) によって回答からの離脱行動がモデリングされ、パラメータに関して異なる仮定を置いた複数のモデルが比較・検討された。その結果、全質問項目における回答率が全回答者で共通であるという仮定を置いたモデルよりも、全項目中の回答率が回答者毎に異なるという拡張を施したモデルの方が高い予測性能を示した。このように、Okada et al. (2018) は、回答からの離脱という現象が個人差変数に依存して生じるものか、それとも何らかのノイズによって一見して分散が生じているのかについて、実証的に検討したと言える。また、回答からの離脱は、無回答の発生という行動の一形態であり、本研究で着目する現象の特別な場合である

と言える。そのため Okada et al. (2018) で用いられたような統計モデリングのアプローチは、本研究で着目する無回答の発生という、より一般的な現象にも適用可能であると考えられる。したがって、Okada et al. (2018) のように、異なる仮定を置いた複数のモデルを比較・検討することによって、無回答の発生の背後にあるメカニズムを明らかにする一助となることが期待される。よって本研究では、無回答の発生が個人差変数に依存する現象かどうかについて、確率分布を用いた基本的な検討を行う。

また、第一の目的に関して、本研究では無回答の発生が個人差変数から最大でどの程度の精度で予測可能であるかを検討する。そのために、本研究で実施する統計モデリングにおいて無回答の発生を記述するために妥当であると判断された確率分布の推定結果を利用する。これによって、この上限値を数値的に求めることが出来る。このようにして、無回答の発生を個人差変数から予測・説明する研究に対して、一つの参照点を提供することを試みる。

本研究の第二の目的は、回答者のパーソナリティ特性と無回答の発生の関連を明らかにすることである。先行研究では、回答者の属性、健康状態などの個人差変数と無回答の関連が検討されてきた。しかし、回答者のパーソナリティ特性と無回答の発生の関連については直接的に検討されていない。パーソナリティは、思考や感情、行動における特定のパターンの個人差を生じさせる個人の内的要因である (谷, 2015)。加えて、パーソナリティは、心理学をはじめとした行動科学における代表的かつ重要な個人差変数である。そのため、回答者のパーソナリティによって無回答の発生率がどのように異なるのかを明らかにすることは、今後の研究にとって意義があると考えられる。

第二の目的を達成するために、本研究では以下の二点に着目した検討を行う。第一に、本研究では、構築したモデルの予測性能を、既存の研究よりも妥当な方法で評価する。無回答の発生に関する先行研究 (Tsiampalis & Panagiotakos, 2020; Kutschar et al., 2019) では、全回答データをモデルの構築に用いており、そのモデルを解釈している。しかし、この方法では、構築したモデルがデータに対して過剰適合している場合、結果の妥当性が毀損される。よって本研究では、全データをモデルの構築に用いる訓練データと、モデルの予測性能を確かめるテストデータに分割する。そして、テストデータにおけるモデルの予測精度を算出し、これをもとにモデルの性能を比較・検討する。これによって、予測モデルの性能について、客観的に評価する。

加えて、モデルの予測性能に関しては、第一の目的における統計モデリングの結果に基づいた検討を行う。具体的には、推定した確率分布をもとにした予測精度の上限値を参考にして、回答者のパーソナリティから無回答の発生を予測するモデルの性能を定量的に評価する。

第二の着目点として、本研究では機械学習を用いて無回答の発生を予測する。無回答の発生と個人差変数の関連を検討した先行研究 (Tsiampalis & Panagiotakos, 2020; Kutschar et al., 2019) では、ロジスティック回帰モデルやポアソン回帰モデルといった一般化線形モデルが用いられ

ていた。本研究では、これ以外の機械学習で用いられるモデルも同時に構築する。例えば、Lasso 回帰・Ridge 回帰といったパラメータに制約を設ける回帰モデルを用いれば、モデルのデータに対する過剰適合を防ぐことが期待される。また、ランダムフォレスト・勾配ブースティングといった決定木を基にしたアンサンブル法は、説明変数の交互作用や非線形的な効果を検出できる。関連分野における機械学習の適用例としては、オンライン調査における不適切回答を予測するために複数の機械学習の予測モデルを比較・検討した尾崎・鈴木 (2019)、オンライン調査における努力を最小限化するような回答の予測に Lasso 回帰を用いた三浦・小林 (2016) が挙げられる。これらの研究で用いられた手法は、同じくデータの質にかかわる現象である無回答の発生の予測においても有効であると考えられる。よって、以上に挙げた種々のモデルを検討することで、予測性能の高いモデルを構築し、一般化線形モデルでは見られない効果を検出できる可能性がある。本研究では、一般化線形モデルに加えて、機械学習で用いられるその他の予測モデルを構築し、それらの性能を比較・検討する。その後、性能が良いと考えられるモデルを解釈する。

5 因子モデル (McCrae & Costa, 1987) におけるパーソナリティの各特性次元の概念的定義より、各次元は無回答の発生と以下のような関連を持つと考えられる。神経症傾向 (Neuroticism) は、無回答の発生と正の関連を持つと考えられる。神経症傾向が高い個人は回答時におけるストレスに対する脆弱性が高いため、不注意による無回答が誘発されやすいと推測される。一方、質問紙回答に対する動機づけの観点から、外向性 (Extraversion)・開放性 (Openness)・協調性 (Agreeableness) は、それぞれ無回答の発生と負の関連を持つと考えられる。外向性が高い個人は質問紙への回答時における快感情を経験しやすいと推測される。開放性が高い個人は、学術研究における調査内容に対して興味を持ちやすいと考えられる。協調性が高い個人は、調査者に対して協力的なふるまいを取りやすいと推察される。このように、これら三つの特性次元は、調査回答に関する高い動機づけを持つため、無回答と負の関連を持つと予測される。最後に、誠実性 (Conscientiousness) は無回答の発生と負の関連を持つと考えられる。これは、誠実性が高い個人は回答時に注意を働かせやすく、質問項目の見落とし等のミスを生じさせにくいと推測できるためである。

2. 分析対象データ

2.1 データ収集の概要

本研究では、大規模データセット (Johnson, 2014) を二次的に利用した。これは、Johnson (2014) がインターネット上で実施した、パーソナリティ心理学に関する調査で得られたデータである。⁽¹⁾ 回答集計期間は 2001 年 1 月から 2011 年 5 月であった。

この調査は Web サイト上で誰でも参加可能であった。この Web サイトの主要な目的は、回答者に対してパーソナリティの 5 因子モデルに関する教育を行うことであった。そ

のため、回答者は回答終了後、自分自身のパーソナリティについて 5 因子モデルに基づくフィードバックを受けた。

2.2 調査回答者

このデータセットに記録された回答者数は $N = 307,313$ (男性 122,164 人、女性 185,149 人) であった。回答者の年齢の平均値は $M = 25.19$ 歳 ($SD = 10$ 歳) であり、年齢の範囲は 10 歳から 99 歳であった。

2.3 除外基準

Web サイトで収集された回答データは、Johnson (2005) の除外基準に従ってスクリーニングされた。Johnson (2005) の基準は、(a) 同一の回答者による複数の回答データ、(b) 明らかに回答に対して注意を払っていない回答者のデータ、(c) 内的一貫性が不十分な回答者のデータ、(d) あまりに多くの無回答項目数を示した回答者のデータを除外するものである。これらの基準に該当した回答者のデータを除外したものが、Web サイト上に公開されている。

これらの四つの除外基準のうち、本研究と大きく関連するものは (d) である。Johnson (2014) は、(d) の無回答項目数に関して具体的な数値を記述していない。しかし、公開された回答データを確認すると、Johnson (2014) の回答データには、11 項目以上の無回答を示した回答者のデータは記録されていなかった。そのため、本研究における分析対象者は、後述する尺度 (IPIP-NEO-300) の 300 項目中、10 項目以下で無回答を示した回答者である。

2.4 分析対象項目

本研究ではこのデータセットのうち、回答者の年齢・性別および回答者のパーソナリティ特性を測定する 300 項目を分析対象とした。

回答者のパーソナリティ特性は、IPIP-NEO-300 (Goldberg, 1999) によって測定された。これは 5 因子モデル (Costa & McCrae, 1987) に基づいて作成された尺度である。神経症傾向 (Neuroticism)、外向性 (Extroversion)、開放性 (Openness)、協調性 (Agreeableness)、誠実性 (Conscientiousness) の 5 つの次元がそれぞれ 60 項目 (5 件法) によって測定された。Johnson (2014) の回答データには欠測データが含まれているため、ここにおける無回答の発生を検討対象とした。

3. 無回答の発生過程の個人差に関する検討

以後の分析は、オープンソースの統計ソフトウェアである R ver. 3.6.3 (R Core Team, 2020) を用いて実施された。本章では、無回答の発生における個人差に関する検討を行う。

3.1 2 種類の確率分布の当てはめ

Johnson (2014) のデータにおける無回答項目数の分布を図 1 に示す。無回答項目数の平均値は $M = 1.21$ 、標準偏差は $SD = 1.73$ であった。1 項目以上で無回答を示した回答者は、全体の約 52% を占めていた。

本研究では、各回答者の無回答は項目間で独立に生じ、

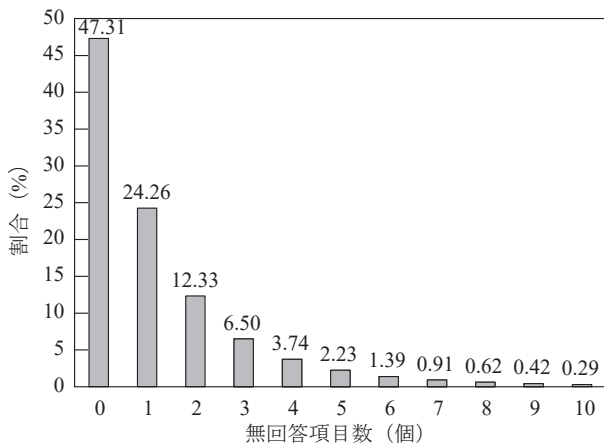


図1：無回答項目数の分布

注：グラフの上部の数値は各無回答項目数を示した回答者の割合を表す。

その発生確率はどの項目でも一定であるという仮定を置く。この仮定の下では、300項目における各回答者の無回答項目数は二項分布に従う。

さらに、この二項分布はポアソン分布によって近似されると仮定する。ポアソン分布は、多数回の試行における稀な事象に関する確率分布である。Johnson (2014) のデータには、300項目中で最大10項目の無回答を発生させた回答者のデータが記録されているため、各回答者における無回答項目数を記述するモデルとしてポアソン分布は適切であると考えられる。そこで、本研究では、Johnson (2014) のデータの回答者集団における無回答項目数の分布に関してポアソン分布および負の二項分布の二種類の確率分布を当てはめる。ポアソン分布においては、全回答者が単一の期待値 (λ) に従って無回答を生じさせると仮定される。そのため、各回答者の無回答項目数がポアソン分布に従うという仮定の下で、回答者集団の無回答項目数もポアソン分布それ自体に従う場合、無回答の発生率に個人差が存在しないと解釈される。Johnson (2014) のデータから、ポアソン分布のパラメータ (λ) を最尤推定したところ、パラメータ推定値は $\hat{\lambda} = 1.21$ であった。

一方、負の二項分布は、各回答者が従うポアソン分布のパラメータである無回答項目数の期待値 (λ_i) がガンマ分布に従うと仮定することで導かれる確率分布である。すなわち、ここでは無回答の発生確率が回答者毎に異なる。したがって、各回答者の無回答項目数がポアソン分布に従うという仮定の下で、回答者集団の無回答項目数が負の二項分布に従う場合、無回答の発生率に関して個人差が存在すると示唆される。なお、ガンマ分布は、非負の連続量に関する確率分布として一般的に用いられる。そして、 λ_i は計数データの期待値であるため、必ず0以上の値を取る。以上の点から、本研究では、 λ_i がガンマ分布に従うという仮定することで導かれる確率分布である負の二項分布を用いた。最尤推定の結果、サイズパラメータ (r) の推定値は $\hat{r} = 0.817$ 、成功確率パラメータ (q) の推定値は $\hat{q} = 0.403$ であった。

これら二つの各確率分布の実データに対する当てはまり

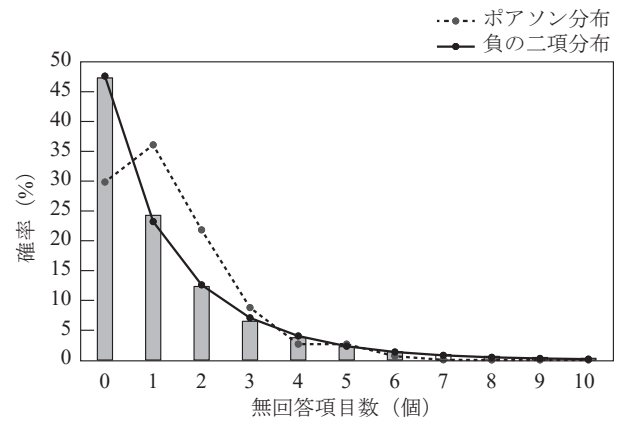


図2：無回答項目数の分布および2つの確率分布の予測値

を確認する。各確率分布の予測値と、実際の無回答項目数の分布を重ね合わせたグラフを図2に示した。このグラフから、ポアソン分布は無回答項目数が0個の回答者数を過少に予測していることが読み取れる。一方、負の二項分布はポアソン分布よりも当てはまりが良い。また、ポアソン分布のAICは1069408.5、負の二項分布のAICは933643.9と、負の二項分布の方が低い値であった。これらの点から、負の二項分布は、ポアソン分布に比べて、実データにおける無回答項目数の分布により適合していると考えられる。以上の検討より、無回答の発生に関しては、回答者毎に無回答を生じさせる確率が異なる可能性が高いと言える。

3.2 ポアソン分布の仮定に関する検討

前章においては、各回答者における無回答項目数がポアソン分布に従うという仮定の下で、無回答の発生過程の個人差について検討した。しかし、この仮定の妥当性について、検証が必要である。ポアソン分布は、ある事象の生起確率が p であるベルヌーイ試行を n 回繰り返したときの二項分布 $Binomial(n, p)$ について、 p が小さく n が大きい極限を取ることによって導出される。したがって、各項目における無回答の発生確率 (p_i) がほとんど全ての回答者で十分に小さく、項目数が十分に大きいと言えない限り、各回答者における無回答項目数の分布をポアソン分布によって近似することは出来ない。その場合、各回答者における無回答項目数がポアソン分布に従うという仮定は妥当でない。

ここでは、各回答者における無回答項目数がポアソン分布に従うという仮定の妥当性を検討する。そのために、各回答者の各項目における無回答の発生確率 (p_i)、および本データの質問項目数 (n) が、ポアソン分布によって二項分布を近似するための条件を満たしているかどうかを確認する。

まず、各回答者における無回答の発生確率 (p_i) が取りうる値の範囲を推測する。負の二項分布から、各回答者の無回答項目数が従うポアソン分布のパラメータ (λ_i) の分布 ($Gamma(k, \theta)$) が導出される (式 (1))。

$$\lambda_i \approx Gamma(k, \theta) \quad (1)$$

負の二項分布のパラメータ（サイズパラメータ r , 成功確率パラメータ q ）から、 λ_i が従うガンマ分布のパラメータ（形状パラメータ k , 尺度パラメータ θ ）を導出することが出来る（式 (2)・式 (3)）。

$$k = r \quad (2)$$

$$\theta = \frac{q}{1-q} \quad (3)$$

負の二項分布のパラメータの最尤推定値と式 (2)、式 (3) より、各回答者の λ_i が従うガンマ分布のパラメータ推定値が求められる。結果として、形状パラメータは $\hat{k} = 0.82$, 尺度パラメータは $\hat{\theta} = 1.48$ と推定された。そして、これらのパラメータを持つガンマ分布の分位点関数によると、 λ_i の 99% 分位点は $\lambda_i = 6.17$ であった。このことから、ほとんどの回答者における λ_i は $0 \leq \lambda_i \leq 6.17$ の範囲に存在すると推測される。したがって、各回答者の各項目における無回答の発生確率 (p_i) は $0 \leq p_i \leq 6.17/300$ の範囲に分布すると考えられる。そのため、この範囲の p_i と $n = 300$ をパラメータとする二項分布 (*Binomial* ($p_i, n = 300$)) についてポアソン分布による近似が妥当かどうかを確認すればよい。

ここでは、項目数を $n = 300$ に固定し、各項目における無回答の発生確率 (p_i) を操作した二項分布 (*Binomial* ($X | p_i, n = 300$)) を真のモデルとした。また、 p_i について $p_i = [1/600, 1/150, 7/300]$ の三条件を設けた。ポアソン分布のパラメータ λ は $\lambda = np$ と導かれるため、 $p_i = 1/600$ の条件では、対応するポアソン分布は *Poisson* ($X | \lambda_i = 0.5$) であり、 $p_i = 1/150$ の条件におけるそれは *Poisson* ($X | \lambda_i = 2$) である。最後に、 $p_i = 7/300$ の条件におけるポアソン分布は *Poisson* ($X | \lambda_i = 7$) であるが、この値は λ_i が従うと推定されるガンマ分布の 99% 分位点 ($\lambda_i = 6.17$) よりも大きい値である。そのため、このデータにおいて $\lambda_i > 7$ の回答者はほとんど存在せず、ほぼ全ての回答者の λ_i は $0 \leq \lambda_i \leq 7$ の範囲に存在すると推測される。

各条件における二項分布およびポアソン分布を図 3 から図 5 に示した。結果として、いずれの条件においても

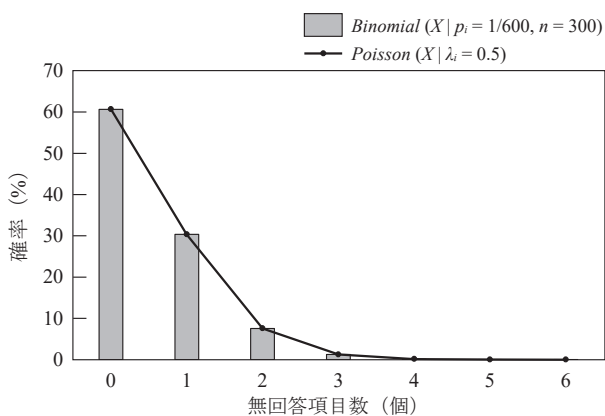


図 3 : $p_i = 1/600$ 条件における二項分布、および対応するポアソン分布

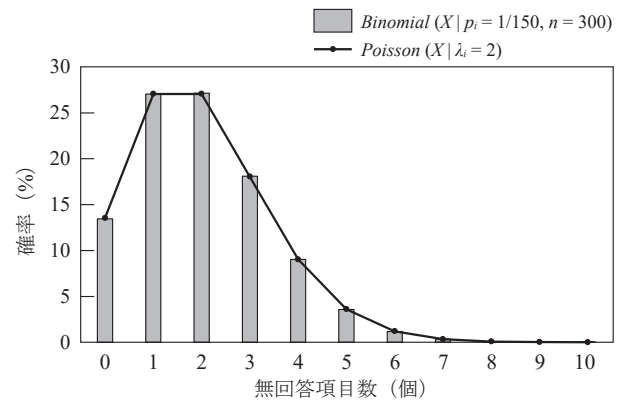


図 4 : $p_i = 1/150$ 条件における二項分布、および対応するポアソン分布

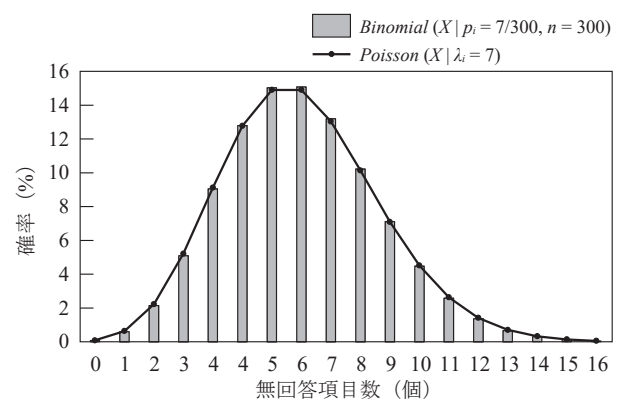


図 5 : $p_i = 7/300$ 条件における二項分布、および対応するポアソン分布

ポアソン分布は二項分布の良い近似になると示された。特に、 $p_i = 7/300$ の条件においてもポアソン分布は二項分布からほとんど乖離していないことから、ほとんどの回答者の無回答項目数に関して、ポアソン分布が十分な近似性能を持つと示唆される。そのため、本データのほとんどの回答者における無回答の発生確率 (p_i) は十分に小さく、項目数 (n) は十分に大きいと判断できる。したがって、本データにおける各回答者の無回答項目数がポアソン分布に従うと仮定することは妥当であると考えられる。

4. 無回答の予測可能性に関する理論的検討

4.1 シミュレーションの概要

前章において、無回答の発生が負の二項分布によって記述できることが示された。ここでは、前章においてデータから推定された負の二項分布をもとに、無回答の発生が個人差変数から最大でどの程度の精度で予測可能かについて、明らかにすることを試みる。無回答の発生を個人差変数から説明した研究 (Tsiampalis & Panagiotakos, 2020; Kutschar et al., 2019) では、各回答者が無回答項目を 1 つでも発生させるか、1 つも発生させないか、という 2 値のアウトカムを予測するモデル (e.g. ロジスティック回帰モデル) が用いられた。このように各回答者における無回答の有無を予測するモデルが持つ予測精度の限界

値を、本研究の検討対象とする。なお、本研究において、モデルの予測精度は、各回答者における無回答の有無を正しく予測出来る確率を指す。

まず、無回答の予測の理論的な限界を求めるために、無回答を引き起こす素因が特定されている理想的な状況を考える。また、このとき、各回答者がどの程度その素因を持っているかを、100%の精度で特定できる理想的なモデルがあるとする。これは、各回答者の無回答の発生確率を何らかの個人差変数によって完全に条件付けることが出来る理想的な状況といえる。

しかし、このような理想的なモデルであっても予測を誤る可能性がある。ある回答者が無回答を起こす高い素因を持っていたとしても、300項目内では無回答が偶然に生じない可能性があるためである。また、それとは逆に無回答を引き起こす素因が少ない回答者であっても、偶然に無回答が生じる可能性も存在する。これらの場合、このモデルは無回答の発生の予測を誤ることになる。以上の点から、理想的なモデルであっても、各回答者における無回答の有無の予測精度は100%に満たないと予測される。そして、この理想的なモデルの予測精度は、個人差変数から無回答の発生を予測する際の予測精度の理論的な上限値と解釈できる。

4.2 シミュレーションの手続き

4.2.1 疑似回答者の生成

本章では、上記の点について検討するために、以下の手続きによるシミュレーションを実施した。はじめに、疑似的な回答者を乱数から生成した。具体的には、300項目内における無回答項目数が負の二項分布に従うと仮定した。なお、このシミュレーションで仮定する負の二項分布のパラメータには、実データにおける最尤推定値を用いた。負の二項分布のパラメータの最尤推定値から、各回答者の無回答項目数の期待値 (λ_i) が従うガンマ分布のパラメータ推定値を求めることが出来る(式(2)、式(3))。第3章における最尤推定の結果、形状パラメータは $\hat{k}=0.82$ 、尺度パラメータは $\hat{\theta}=1.48$ であったため、これらのパラメータを持つガンマ分布から100万個の λ_i を独立にサンプリングした。そして、 λ_i は各回答者が持つ無回答の発生の素因として解釈される。したがって、この手続きによって、無回答生起の素因が大小さまざまである100万人の疑似回答者を生成したことになる。

4.2.2 理想的な予測モデル

次に、想定する予測モデルについて述べる。ここでは理想的なモデルを想定するため、この予測モデルが各回答者における無回答項目数の期待値 (λ_i) を100%の精度で推定できる理想的な状況を設定した。よって、この理想的なモデルは各回答者が持つ無回答発生の素因を特定できることになる。

続いて、このモデルによって無回答の有無を予測する。この予測は、各回答者における無回答項目数の期待値 (λ_i) に関して、以下の2つの場合に分けて行えばよい。第一に、

回答者 i の無回答項目数の期待値が $\lambda_i < 1$ の場合、その人は無回答を1つも生じさせない、と予測する。これは、 $\lambda_i < 1$ のポアソン分布において $X=0$ となる確率は、それ以外の各値を取る確率よりも必ず大きくなるためである。このとき、このモデルが回答者 i における無回答の有無を正しく予測できる確率 $Accuracy_i$ は、ポアソン分布 ($\lambda_i < 1$) において実現値が0を取る確率(式(4) 上部)によって表される。

第二に、回答者 i の無回答項目数の期待値が $\lambda_i \geq 1$ の場合、その回答者 i は無回答を1項目以上で生じさせると予測する。このとき、回答者 i が無回答を発生させると正しく予測できる確率 $Accuracy_i$ は式(4) 下部で表される。これは、確率変数 X が1以上、すなわち0以外の実現値を取る確率である。そのため、ポアソン分布 ($\lambda_i \geq 1$) において実現値が0となる確率を1から減ずることで $Accuracy_i$ を求められる。

$$Accuracy_i = \begin{cases} \text{Poisson}(X=0 | \lambda_i) & \text{if } \lambda_i < 1 \\ 1 - \text{Poisson}(X=0 | \lambda_i) & \text{if } \lambda_i \geq 1 \end{cases} \quad (4)$$

4.2.3 予測精度の算出

次に、このモデルの予測精度を求める。予測精度は、全ての疑似回答者の中で、このモデルが無回答の有無を正しく予測できる割合である。そのため、100万人の疑似回答データについて、無回答の発生を正しく予測できる確率 $Accuracy_i$ を求めた。その後、 $Accuracy_i$ の全回答者に関する平均値を理想的なモデルの予測精度とした。

4.3 シミュレーションの結果

以上の手続きによるシミュレーションの結果、この理想的なモデルは76.09%の予測精度を示した。⁽²⁾ よって、個人差変数から無回答の発生を予測する際、どんなに性能が高い予測モデルであっても、その予測精度は約76%であると考えられる。したがって、各回答者における無回答の発生に関して、個人差変数から最大で約76%の精度で予測可能であると示唆される。

5. パーソナリティ特性と無回答の発生の関連

第3章の結果から、無回答の発生が個人差変数に依存すると示唆される。本章では、本研究の第二の目的を達成するため、個人差変数の中でもパーソナリティ特性に着目する。具体的には、回答者のパーソナリティ特性と、回答者の性別・年齢から無回答の発生を予測するモデルを構築し、それを解釈する。

5.1 欠測値補完

無回答の発生をパーソナリティ特性から予測するためには、無回答を示した回答者のパーソナリティ特性に関して完全なデータを得ることが望ましい。しかし、IPIP-NEO-300に対する回答データには欠測値が含まれているため、これらの欠測値を何らかの方法で補完したものを予測モデルの構築に用いる必要がある。

本研究では欠測値補完の方法として、単一代入法の一つ

である確率的回帰代入法 (Stochastic regression imputation) を用いた。確率的回帰代入法は、欠測値を以下の方法で補完する方法である。まず、補完対象となる変数における欠測値について、それ以外の欠測がない変数を説明変数とする回帰モデルを構築する。次に、そのモデルの予測値に正規分布 ($\mu = 0, \sigma^2 =$ 構築した回帰モデルにおける残差の分散) から生成した乱数を加えた値を欠測値に代入する。このようにして、IPIP-NEO-300 における欠測値を補完した。

なお、この補完データをそのまま用いると、無回答を発生させた回答者においてのみ、小数値のデータや、1 未満もしくは 5 より大きいデータが存在することになる。しかし、小数値の有無や、極端に大きい・小さい値の有無は、無回答を発生させた回答者と完全回答を示す回答者を判別する手がかりとなる可能性がある。そのため、これらの値をそのままにしてモデルを構築すると、予測モデルがこの手がかりを利用してしまふ危険性がある。モデル構築におけるアーティファクトを防ぎ、パーソナリティ特性の効果を妥当な方法で検討するため、代入値の小数点第 1 位を四捨五入した。その後、1 未満になったデータは 1 に、5 より大きい値の回答データは 5 に変換した。

5.2 予測モデル

IPIP-NEO-300 における無回答の発生を予測するモデルを構築した。データセットを訓練データとテストデータに 1:1 の比率でランダムに分割し、各モデルの予測精度を確認した。

本研究におけるアウトカムは、各回答者における無回答の有無である。すなわち、ある回答者において、パーソナリティを測定する 300 項目のうち 1 つでも無回答を示すか、あるいは 1 つも無回答を示さないかを予測することを目標とした。説明変数は、回答者の年齢、性別、パーソナリティの各因子得点 (因子内項目の平均値) である。なお、説明変数はすべて標準化された。

5.2.1 ロジスティック回帰

ロジスティック回帰モデルは、2 値を取る従属変数の値を説明変数によって予測する線形モデルである。本研究では、訓練データにおける 10 分割交差検証によって最適なモデルを探索した。なお、実装にはパッケージ `caret` のメソッド `glm` を用いた。

5.2.2 Lasso 回帰・Ridge 回帰

Lasso 回帰および Ridge 回帰は、訓練データに対する過適合を防ぐことを目的とする、パラメータに制約をかける線形モデルである。これら 2 つのモデルには、ロジスティック回帰の各係数に関するペナルティ項が設けられている。Lasso 回帰では、各係数の絶対値の和に対してペナルティを与える (L1 正則化)。一方、Ridge 回帰では各係数の二乗和に対してペナルティを与える (L2 正則化)。訓練データにおける 10 分割交差検証によって正則化パラメータ (ペナルティ項) を最適化し、最良のモデルを探索した。実装には、パッケージ `caret` のメソッド `glmnet` を用いた。

5.2.3 ランダムフォレスト

ランダムフォレストは、データ全体から抽出されたブートストラップ標本に対して、独立に複数の決定木を構築するモデルである。予測に際しては、それぞれの決定木による予測結果を総合する。なお、ランダムフォレストでは、参加者だけでなく説明変数についてもブートストラップ標本抽出を行う。各決定木に用いる説明変数の個数を最適化し、最良のモデルを探索した。なお、ここではモデルの構築に際して訓練データ全体をそのまま用いた。実装には、パッケージ `caret` のメソッド `rf` を用いた。

5.2.4 勾配ブースティング

勾配ブースティングは、複数の決定木を構築するという点でランダムフォレストと共通している。一方、勾配ブースティングでは、一つ前に構築した決定木における判別の正誤を受けて、次の決定木を構築する。したがって、このモデルには各決定木の間に完全な独立性はなく、逐次的に決定木が作られる。訓練データにおける 10 分割交差検証によって、7 つのハイパーパラメータ (1. 構築する決定木の個数、2. 決定木が持つ深さの上限、3. 学習率、4. 決定木の葉の追加による損失減少の下限、5. 各決定木に用いる説明変数の個数、6. 各決定木に用いる回答者の数、7. 決定木の葉の重みの下限) を最適化し、最適なモデルを探索した。実装には、パッケージ `caret` のメソッド `xgbTree` を用いた。

5.3 モデルの予測精度

各モデルの訓練データおよびテストデータにおける予測精度を表 1 に示す。テストデータにおいて予測精度が最も高かったものは、ロジスティック回帰モデル、Ridge 回帰、勾配ブースティングであった (約 53.9%)。先の分析において予測精度の上限値が約 76% と示されたが、本研究におけるモデルの予測精度 (テストデータにおける予測精度) の最大値は約 54% であった。このことから、無回答の発生を予測する個人差変数として、本研究で用いた説明変数のいずれかは、十分ではないものの一定の情報を持つと示唆される。

表 1: 訓練・テストデータにおける各予測モデルの予測精度

判別モデル	訓練	テスト
ロジスティック回帰	54.05 %	53.87 %
Lasso 回帰	54.08 %	51.54 %
Ridge 回帰	54.07 %	53.91 %
ランダムフォレスト	52.32 %	52.08 %
勾配ブースティング	55.33 %	53.91 %

5.4 説明変数の予測方向

テストデータにおける予測精度が比較的高かった予測モデルのうち、ロジスティック回帰モデルを詳細に解釈する。これは、テストデータにおける予測精度が同等であった勾配ブースティングと Ridge 回帰よりも解釈可能

性が高いためである。

ロジスティック回帰モデルにおける各説明変数の偏回帰係数(β)および調整オッズ比(AOR)を表2に示す。なお、性別には男性=1、女性=2と数量化されたものの標準化得点を用いた。

表2：ロジスティック回帰モデルの偏回帰係数および調整オッズ比

説明変数	β	AOR
性別	-0.064***	1.07
年齢	-0.109***	1.11
神経症傾向	-0.070***	1.07
外向性	-0.127***	1.14
開放性	-0.070***	0.93
協調性	-0.010	0.99
誠実性	-0.018**	0.98

注：*** $p < .001$, ** $p < .01$ 。

解釈対象のロジスティック回帰モデルにおけるサンプルサイズは138,291⁽³⁾と非常に大きいため、偏回帰係数の解釈における有意水準を0.1%に設定した。協調性および誠実性の偏回帰係数はほとんど0に近く、その他の説明変数に比べて絶対値が小さかった。また、協調性および誠実性の主効果は有意ではなかった($ps > .001$)。これらの点から、協調性および誠実性は、無回答の発生の予測において重要度が相対的に低いと示唆される。

性別・年齢(女性)・外向性・神経症傾向の主効果が有意であった($ps < .001$)。また、偏回帰係数の値に着目すると、これらの説明変数は無回答の発生と正の関連を持つことが示唆される。一方、開放性の主効果も有意であったが($p < .001$)、偏回帰係数の値は負であった。そのため、開放性は無回答の発生と負の関連を持つことが示唆される。

6. 総合考察

本研究の第一の目的は、無回答の発生過程に個人差が真に存在するかどうかを明らかにすることであった。そのために、回答者集団における無回答項目数の分布に関して二つの確率分布を当てはめた。結果として、無回答の発生確率が個人ごとに異なるという仮定を置いたモデル(負の二項分布)は、無回答の期待値が個人間で共通であるという仮定を置いたモデル(ポアソン分布)よりも適合度が高かった。この結果から、無回答の発生は回答者の個人差変数に依存する可能性が高いと言える。

また、各回答者における無回答項目数がポアソン分布に従うという仮定の妥当性を検討した。結果として、本データにおけるほとんどの回答者における無回答項目数について、ポアソン分布は二項分布を十分に良く近似していた。この結果から、各回答者における無回答項目数のモデルとして、ポアソン分布を用いることは妥当であると言える。以上の検討から、各回答者の無回答項目数はポアソン分布に従う一方で、無回答項目数の期待値に

は個人差が存在すると示唆される。

個人差変数から無回答の発生を予測するモデルを構築した先行研究(Tsiampalis & Panagiotakos, 2020; Kutschar et al., 2019)では、回答者の属性や状態によって無回答の発生率が異なると示された。本研究における確率分布による検討から、先行研究で見られた個人差変数の効果が個人内変動などのノイズによって生じた疑似的な効果ではなく、回答者間に存在する特性や状態(個人差変数)の分散によって系統的に生じた効果であることが示唆される。先行研究では、個別の個人差変数と無回答の発生が検討されてきたが、本研究では無回答の発生メカニズムに関して基礎的かつ一般的な知見を提供したと言える。

また、無回答の発生の背景にある確率分布に関する検討の結果(負の二項分布)から、無回答の発生が個人差変数から最大でどの程度の精度で予測できるかを、シミュレーションによって求めた。その結果、理想的な予測モデルであっても、各回答者における無回答の発生の有無を正しく予測できる確率は約76%であることが示された。これによって、本テーマにおける予測モデルの性能に関して、一定の評価基準を提案した。今後の類似研究では、個人差変数から各回答者の無回答の発生を予測するモデルの予測精度が取りうる範囲を50%~100%ではなく、50%~76%と設定し、この範囲に照らし合わせてモデルの予測精度を評価する方が妥当であると言える。⁽⁴⁾ただし、この上限値は300項目中における無回答の発生の予測精度に限定される。よって、これ以外の項目数における無回答については一般化出来ない点には注意すべきである。

本研究の第二の目的は、質問紙調査における無回答の発生と、回答者のパーソナリティ特性の関連を検討することであった。そのために、Johnson(2014)による大規模データにおける無回答の発生の有無を、回答者のパーソナリティ特性および属性から予測するモデルを構築した。本研究で構築したモデルのテストデータにおける予測精度は、最大で約54%であった。無回答の発生の予測において、回答者のパーソナリティ特性および属性のうちいずれかは、一定の重要度を持つ個人差変数であると言える。しかし、本研究のモデルの予測精度である54%は、モデルが取りうる予測精度の上限値である76%に近いとは言えない。このことから、無回答の発生は、回答者のパーソナリティ以外の何らかの個人差変数にも依存すると示唆される。よって、無回答の発生を規定する個人差変数としては、回答者の年齢・性別、パーソナリティ以外の要因も考慮する必要があると言える。

ロジスティック回帰モデルの偏回帰係数から、無回答項目の発生は神経症傾向と外向性と正の関連を、開放性は無回答の発生と負の関連を持つと示唆される。この結果から、神経症傾向が無回答の発生と正の関連を持つという仮説は支持された。また、開放性は無回答の発生と正の関連を持つという仮説も支持された。一方、仮説とは反対に、外向性は無回答の発生と正の関連を持つということを示唆する結果が得られた。また、協調性・誠実性と無回答の発生との関連を支持するエビデンスは得られなかった。

パーソナリティ特性と無回答の発生に関する検討の結果から、質問紙調査における欠測データを不適切に処理すると、研究者が興味を持つパラメータの推定のバイアスにつながる可能性が高いと言える。具体的には、質問紙調査においてリストワイズ除去を行うと、外向性および神経症傾向が高い回答者や、開放性が低い回答者の情報を無視することにつながる危険があると言える。

5因子モデルによる各パーソナリティ特性は、様々な個人の特性や状態と関連することが確認されている。例えば、神経症傾向は心身の健康状態を予測することが示されている (Lahey, 2009)。他にも、EQ (emotional intelligence quotient) は神経症傾向と負の関連を、外向性および誠実性と正の関連を持つと報告されている (Brackett & Mayer, 2003)。このように、行動科学で測定される様々な個人差変数はパーソナリティ特性と関連を持つと示唆される。そのため、個人差変数を扱う研究において、無回答の発生は研究で設定するアウトカムや、説明変数、共変量に依存する可能性がある。これらのことから、本研究の結果は5因子モデルに基づくパーソナリティ特性にとどまらず、個人差変数一般 (特性、状態など) について示唆を提供するだろう。

本研究の課題を五点述べる。第一に、本研究で使用したデータセットには、11項目以上で無回答を示した回答者のデータが記録されていない。これは、データを収集・公開した Johnson (2014) のデータ除外基準による結果である。実際の質問紙調査では、高い割合で無回答を発生させた回答者のデータが得られることもある。本研究では、300項目のうち多くても10項目で無回答を示す回答者のデータを分析対象としたが、これ以上の無回答項目数を含む回答者について、本研究の知見が一般化できるかどうかについては不明である。

第二の課題点は、元データの収集の手続きに関するものである。Johnson (2014) は、回答者にパーソナリティの5因子モデルについて教育することを目的として、調査を実施した。その目的を達成するため、調査回答前にこの調査目的が回答者に伝えられ、回答後に回答者は自身のパーソナリティに関するフィードバックを受けた。そのため、本調査の回答者には、自分のパーソナリティに興味がある個人や、パーソナリティ心理学を学ぶ学生が多く含まれている可能性がある。実際、回答者の平均年齢は25歳 ($SD = 10.0$) であり、回答者の年齢が大学生の付近に集中していると読み取れる。そのため、このデータにおける回答者の属性や興味などの変数には偏りがあると示唆される。よって、本研究で用いたサンプルが母集団を代表しているかどうか、さらなる検討が必要である。

第三に、本研究ではパーソナリティ特性を測定する項目内における欠測データを補完し、その補完データをもとにパーソナリティ特性と無回答の発生を関連を検討した。しかし、本研究における従属変数が無回答の発生そのものである点を考えると、本研究で構築した予測モデルにおけるパラメータにもまたバイアスが存在する可能性がある。そのため、パーソナリティ特性を測定する尺度において欠測値が全く存在しない回答データ (完全デー

タ) を用いることが望ましい。しかし、5因子モデルに基づいてパーソナリティ特性を測定する質問紙は、60項目の NEO-FFI (Costa & McCrae, 1992) や、120項目の IPIP-NEO-120 (Johnson, 2014)、240項目の NEO PI-R (Costa & McCrae, 1992)、本研究で用いた300項目の IPIP-NEO-300 (Goldberg, 1999) など、多数の項目が含まれるものが主流である。また、項目数が多くなると無回答の発生率は高まると予想される。そのため、通常の方法では、パーソナリティに関する質問紙調査において欠測値を全く含まない回答データを収集することは難しいと言える。これに関しては、パーソナリティ特性を測定する項目において回答を強制するなど、オンライン調査におけるオプションを活用することで部分的に解消されると期待される。

第四に、本研究では各回答者において、無回答が項目間で独立に発生し、その発生確率が一定であると仮定した。この仮定の下で、Johnson (2014) のパラメータの範囲における各回答者の無回答項目数はポアソン分布に従うことが示された (3.2を参照)。しかし、実際の無回答の発生過程がこの仮定を満たしているかどうかについては不明である。特に、個人内において、無回答の発生確率がそれ以前における無回答の発生に依存する場合 (e.g. 自己相関)、本研究で置いた仮定は脅かされる。この場合、各回答者における無回答項目数はポアソン分布に従うという仮定は妥当でない。無回答が項目間で独立に生起するかに関しては、回答者個人レベルでのさらなる検討が必要である。

最後に、本研究では文化差の効果を考慮に入れていない。回答者の集団主義傾向や居住地域の人口密度といった文化的要因が無回答の発生率と関連があるという報告 (Johnson & Cho, 2010) から、無回答の発生を検討するにあたって文化的要因は無視することが出来ないと考えられる。回答者の文化的背景や属性、パーソナリティ特性と無回答の発生に関連について、さらなる研究が必要と考えられる。

注

- (1) データセットはデータ収集者によって公開されている (<https://osf.io/tbmh5/>)。
- (2) 300項目よりも少ない項目数において各回答者における無回答の発生を予測するモデルが持ち得る予測精度の理論的な上限値についても、同様の手続きのシミュレーションによって求められる。この結果、予測精度の理論的な上限値は、200項目で74.54%、100項目で75.56%、50項目で83.96%であった。
- (3) 全体のデータの内、半分の参加者のデータ ($N = 153,657$) を訓練データとし、訓練データにおいて10分割交差検証を行った。10分割交差検証においては、モデルの構築に全体の9/10のデータを使用するため、解釈対象のロジスティック回帰モデルにおけるサンプルサイズは $153,657 \times 0.9 = 138,291$ である。
- (4) 二値判別問題に関する予測モデルの構築において予測精度が50%を下回る場合、一方の予測ラベル (e.g. 無回答あり) を他方のそれ (e.g. 完全回答) と入れ替えることによって、予測精度は50%を上回る。したがっ

て、予測精度は必ず 50 % を上回ると言える。そのため、ここでは無回答の発生の有無を予測するモデルが持ち得る予測精度の下限値を 50 % とした。ただし、予測モデルがデータに対して過剰適合していない場合においてのみ、これは成り立つ。

引用文献

- Brackett, M. A. & Mayer, J. D. (2003). Convergent, discriminant, and incremental validity of competing measures of emotional intelligence. *Personality and Social Psychology Bulletin*, 29, 1147-1158.
- Caan, B., Hiatt, R. A., & Owen, A. M. (1991). Mailed dietary surveys: Response rates, error rates, and the effect of omitted food items on nutrient values. *Epidemiology (Cambridge, Mass.)*, 2, 430-436.
- Costa, P. T. & McCrae, R. R. (1992). *NEO-PI-R professional manual: Revised NEO personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI)*. Odessa, FL: Psychological Assessment Resources.
- Denscombe, M. (2009). Item non-response rates: A comparison of online and paper questionnaires. *International Journal of Social Research Methodology*, 12, 281-291.
- De Rada, V. D. & Domínguez-Álvarez, J. A. (2014). Response quality of self-administered questionnaires: A comparison between paper and web questionnaires. *Social Science Computer Review*, 32, 256-269.
- Eekhout, I., de Boer, M. R., Twisk, J. W., de Vet, H. C. W., & Heijmans, M. W. (2012). Missing data: A systematic review of how they are reported and handled. *Epidemiology (Cambridge, Mass.)*, 23, 729-732.
- Fisher, S. E., Burgio, L. D., Thorn, B. E., & Hardin, J. M. (2006). Obtaining self-report data from cognitively impaired elders: Methodological issues and clinical implications for nursing home pain assessment. *The Gerontologist*, 46, 81-88.
- Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe: Vol. 7* (pp. 7-28). Tilburg, The Netherlands: Tilburg University Press.
- Guadagnoli, E. & Cleary, P. D. (1992). Age-related item nonresponse in surveys of recently discharged patients. *Journal of Gerontology*, 47, 206-212.
- Johnson, J. A. (2014). Measuring thirty facets of the five factor model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality*, 51, 78-89.
- Johnson, J. A. (2005). Ascertaining the validity of individual protocols from Web-based personality inventories. *Journal of Research in Personality*, 39, 103-129.
- Johnson, T., Lee, G., & Cho, Y. (2010). Examining the association between cultural environments and survey nonresponse. *Survey Practice*, 3, 3006.
- Krosnick, J., Holbrook, A., Berent, M., Carson, R., Hanemann, M., Kopp, R. J., Mitchell, R., Presser, S., Ruud, P., Smith, V., Moody, W., Green, M. C., & Conaway, M. (2002). The impact of "no opinion" response options on data quality: Non-attitude reduction or an invitation to satisfice? *Public Opinion Quarterly*, 66, 371-403.
- Kutschar, P., Weichbold, M., & Osterbrink, J. (2019). Effects of age and cognitive function on data quality of standardized surveys in nursing home populations. *BMC Geriatrics*, 19, 244.
- Lahey, B. B. (2009). Public health significance of neuroticism. *The American Psychologist*, 64, 241-256.
- Little, R. J. & Rubin, D. B. (2019). *Statistical analysis with missing data (Vol. 793)*. John Wiley & Sons.
- McCrae, R. R. & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52, 81-90.
- 三浦麻子・小林哲郎 (2016). オンライン調査における努力の最小限化 (Satisfice) を検出する技法—大学生サンプルを用いた検討—. *社会心理学研究*, 32, 123-132.
- Okada, K., Vandekerckhove, J., & Lee, M. D. (2018). Modeling when people quit: Bayesian censored geometric models with hierarchical and latent-mixture extensions. *Behavior Research Methods*, 50, 406-415.
- 尾崎幸謙・鈴木貴士 (2019). 機械学習による不適切回答者の予測. *行動計量学*, 46, 39-52.
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Smith, S. M., Roster, C. A., Golden, L. L., & Albaum, G. S. (2016). A multi-group analysis of online survey respondent data quality: Comparing a regular USA consumer panel to MTurk samples. *Journal of Business Research*, 69 (8), 3139-3148.
- 高橋将宜・阿部穂日・野呂竜夫 (2015). 公的統計における欠測値補定の研究—多重代入法と単一代入法—. *製表技術参考資料*, 30, 1-95.
- 谷伊織 (2015). パーソナリティに関する研究の動向と課題. *教育心理学年報*, 54, 30-44.
- Tsiampalis, T. & Panagiotakos, D. B. (2020). Missing-data analysis: Socio-demographic, clinical and lifestyle determinants of low response rate on self-reported psychological and nutrition related multi-item instruments in the context of the ATTICA epidemiological study. *BMC Medical Research Methodology*, 20, 148.
- van Ginkel, J. R., Sijtsma, K., van der Ark, L. A., & Vermunt, J. K. (2010). Incidence of missing item scores in personality measurement, and simple item-score imputation. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 6, 17-30.
- West, B. & Blom, A. (2016). Explaining interviewer effects: A research synthesis. *Journal of Survey Statistics and Methodology*, 5, 175-211.
- West, B. T. & Olson, K. M. (2010). How much of interviewer variance is really nonresponse error variance. *Public Opinion Quarterly*, 74, 1004-1026.

(受稿：2020年12月22日 受理：2021年1月29日)