

調査設計とデータに関する基礎分析

藤井奈々子・中島日向子・廣谷貴明
(東北大学大学院 教育学研究科)

はじめに

本稿では、平成 28 年度東北大学教育学部開講科目「教育学実習（統計的調査実習）」において実施された調査の設計と、得られたデータの基本特性を整理する。「統計的調査実習（以下、本実習）」では調査企画・調査票設計・エディティング・コーディング・データ入力・クリーニング・分析、そして報告書執筆を通して、受講者がこれら一連の過程を体得することを目的としている。

今年度は一年を通しての形態を採り、二つのグループを構成し、それぞれのグループが調査紙を作成した。ゆえに、本章では調査概要・調査スケジュールに加えてそれぞれの調査紙におけるデータの特性について整理、ならびに回答率・回答の傾向に関して基礎的な分析を行う。

1. 調査の設計

調査の設計について、概要を表 1 に示す。前項で述べた通り、本年度は二つのグループそれぞれで調査紙を作成したが、対象者は例年と同様 20 歳以上 40 歳未満の非学生に設定した。職業・結婚に関するデータ収集を目的としたため、学生は調査対象から除いた。二種類ある調査紙それぞれ 200、合わせて 400 の調査紙を発送した。

調査対象者は調査会社の協力により、上記の条件を満たす者をインターネット調査モニターの中から協力を依頼、依頼に応じていただいた方の中から、年齢・性別に関して層化無作為抽出を実施した。だが、今回の調査の集団が「学生を除く日本全国の 20 歳以上 40 歳以下の男女」であることから分かるように、本調査での理論母集団は通常为社会調査で想定される日本全体の縮図としての集団ではない。加えて、調査会社に登録している層であることも留意点としたい。本稿ではこれらの点を踏まえたうえで、国勢調査との比較・サンプルの特性に関する分析を実施する。

表 1 東北大学教育学部 平成 28 年度統計的調査実習 調査概要

| | | |
|-------|---|--|
| 調査名称 | 若年層の大輔スタイルと意識に関する調査<1>、<2> | |
| 調査目的 | 現代日本の若年層に対し、教育、労働、結婚に関する行動や意識について質問紙調査 | |
| 調査対象 | 1)母集団：学生を除く日本全国の20歳以上40歳未満の男女 2)標本規模：それぞれ200 3)標本抽出法：上記条件を満たすインターネット調査モニターのうち、協力依頼に応じたものから層化無作為抽出 | |
| 調査期間 | 2016年7月 | |
| 調査協力 | 楽天リサーチ株式会社 | |
| 調査項目 | <1> 1)基本属性項目 2)夫婦関係に関する項目 3)仕事に関する項目 4)ライフスタイル（特に睡眠）に関する項目 5)学生時代への意識に関する項目 | <2> 1)基本属性項目 2)結婚に関する項目 3)現在の結婚に対する意識に関する項目 |
| 有効回収率 | <1>136 (68%) | <2>145 (72.5%) |

本実習のスケジュールは表 2 に記した。本年度は通年での開講であったため、前期期間での調査紙発送を区切りとして間が空く形となった。後期期間は調査実施に関する内容に加え、具体的な分析手法についての講義を実施している。また、2つのグループでそれぞれ作業を進行する形で調査紙作成は勿論、データ入力・クリーニングに至るまでの作業をグループ単位で行った。

表 2 東北大学教育学部 平成 28 年度統計的調査実習 調査スケジュール

| | | |
|-------|--------------|--------------------------|
| 2016年 | 4月15日 | 実習授業開始 |
| | 5月16日 | 調査会社へと依頼 |
| | 6月24日 | 中間報告会 |
| | 7月7日 | 事前インターネット調査開始 |
| | 7月11日 | 事前インターネット調査終了、調査協力者リスト納品 |
| | 7月15日 | 質問項目締切・調整 |
| | 7月22日 | 調査票最終確認 |
| | 7月25日 | 調査票封入・投函 |
| | 8月25日 | 調査票調査締切 |
| | | 10月14日 |
| 2017年 | 1月6日・20日・27日 | 最終報告会 |
| | 2月15日 | 最終レポート締切 |

表3 回答者の分布

| | | H27国勢調査 | | H28年度実習 | | | | |
|-----|--------|---------|----|---------|----------|-----|------|----------|
| | | | | <1> | | <2> | | |
| | | % | 度数 | % | χ^2 | 度数 | % | χ^2 |
| 性別 | 男性 | 50.7 | 64 | 47.4 | 0.70 | 56 | 38.6 | 8.9** |
| | 女性 | 49.3 | 71 | 52.6 | | 89 | 61.4 | |
| 年齢層 | 20-24 | 21.3 | 14 | 10.4 | 11.4* | 13 | 9.0 | 16.2** |
| | 25-29 | 22.9 | 42 | 31.1 | | 47 | 32.6 | |
| | 30-34 | 26.0 | 35 | 25.9 | | 42 | 29.2 | |
| | 35-39 | 29.7 | 44 | 21.6 | | 42 | 29.2 | |
| 学歴 | 中学校 | 6.1 | 0 | 0 | 7.3 | 7 | 4.7 | 14.0** |
| | 高校 | 40.4 | 39 | 31.5 | | 58 | 38.7 | |
| | 短大・高専 | 24.6 | 41 | 33.1 | | 35 | 23.3 | |
| | 大学・大学院 | 28.9 | 44 | 35.5 | | 50 | 33.3 | |

*p<.05, **p<.01, ***p<.001

注)学歴のみ H22 国勢調査をもととした数値である

2. 回収サンプルの特性

ここからは、回収されたサンプルの基本属性、そしてその特性について述べていきたい。サンプルから得られた基本情報をまとめるだけでなく、特性を明らかにするにあたっては、基準となる他のデータが必要である。そこで今回は、わが国で最も重要かつ大規模な調査とされる国勢調査と回収サンプルのデータを比較することとした。国勢調査はわが国に住んでいるすべての人及び世帯を対象とし、5年ごとに行われる。西暦の末尾が0の年は「大規模調査」として20項目を調査し、末尾が5の年には「簡易調査」として17項目を調査される。今回は、できるだけ本調査の時期に近い平成27年国勢調査の数値を基準としている。ただし、平成27年度は「簡易調査」の年であり、学歴が調査項目に含まれていない。したがって学歴のみ、平成22年国勢調査をもととした数値となっている。

回収サンプルの基本属性の分布と平成27年国勢調査を比較したものが表3である。他年度の調査実習では従業上の地位・雇用形態や職業等が調査項目に含まれていたが、今回はすでに述べたように調査紙が2種に分かれており、双方に共通する調査項目としては、①性別、②年齢、③学歴の3つのみであった。

それぞれの調査項目について詳しく見ていくと、調査用紙<1><2>で共通して母集団との差が有意に見られる項目は、年齢層である。どちらも母集団と比較して、20-24歳の若年層の割合がかなり低くなっている。この原因の1つとして、すでに述べたように今回の調査対象者に学生が含まれていないのに対して、国勢調査の年代分布には学生が含まれていることが考えられる。

調査用紙<2>のみに着目すれば、すべての基本属性において母集団との乖離が見られる。このような乖離は、インターネット調査モニターの中からサンプルを抽出している以上、一定程度生じることは予測できる。しかし、調査の実施にあたって全く同じ段階を踏んだ調査紙<1>が<2>に比べて、母集団との乖離が少ないことは無視できない。可能性とし

では、調査項目の内容の違いや調査項目数の違いがこのような差を生んだと考えられる。

3. 欠票の分析

本実習で郵送した調査票の全体的な回収率は 70.3% (計画標本数: 400, 有効票数: 281) であった。グループごとの回収率を検討すると「調査票<1>」に関しては 68%であり「調査票<2>」に関しては 72.5%である。わずかではあるが、「調査票<2>」の回収率の方が高くなっている。ただし、この数値は昨年度の教育学実習の回収率と比較すると低い(下瀬川・池田 2015) ⁱ。以下では、調査対象者の基本的属性と返送の有無との関連性について検討する。

本稿では回答者が調査票を返送したかしなかったについて、返送した場合を 1、返送しなかった場合を 0 とした二値変数を従属変数として設定し、二項ロジスティック回帰分析を行う。回答者の基本的属性としては次の 4 つのものを設定した。第一に男性を 1、女性を 0 とした男性ダミー変数である。第二に 20~24 歳を 0、25~29 歳を 1、30~34 歳を 2、35~39 歳を 3 とした年齢に関する順序変数である。第三に雇用形態に関して、正規雇用を参照カテゴリとした、非正雇用ダミー、自営・家族・内職ダミー、無職・その他ダミーといった各ダミー変数である。第四に最終学歴に関して、高等学校を参照カテゴリとした、中学校ダミー、専修学校・高専ダミー、短期大学ダミー、大学ダミーといった各ダミー変数である。第三と第四の独立変数に関して、影響を観察したい変数を 1 とし、それ以外のカテゴリを 0 としている。

分析に際して、本実習での調査対象者全員を対象とした分析 (Model 1)、「調査票<1>」の調査対象者を対象とした分析 (Model 2)、「調査票<2>」の調査対象者を対象とした分析 (Model 3) の 3 つのものを実施する。このようにすることで、調査票の内容による回答者の基本的属性の偏りを導出できる可能性がある。

分析結果を示したものが次の表 4 である。まず Model 1 について検討すると、統計的有意差が認められた変数は無職・その他ダミーのみであった。無職あるいはその他のカテゴリに分類される職業に従事している人は、正規雇用の人に比べて 2.538 倍回答する傾向にある。モデルの当てはまりに関しても 10%水準での統計的な有意性が認められた。一方で Model 2 を検討すると、いずれの変数にも統計的有意差が認められなくなり、さらにモデル自体にも統計的な有意性が認められなくなった。すなわち「調査票<1>」は返送したか、しなかったかは回答者の属性によって左右されないことを示す。次に Model 3 について検討すると、Model 1 と同様に無職・その他ダミーに 1%水準で統計的有意差が認められた。さらに、男性ダミーに関しても 10%水準での統計的有意差が認められる結果となった。このことは無職あるいはその他のカテゴリに分類される職業に従事している人は、正規雇用の人に比べて 4.101 倍回答する傾向にあり、男性は女性に比べて 2.094 倍回答する傾向にあ

ということを示す。モデルの当てはまりに関しても 10%水準での統計的な有意性が認められた。

表 4 返送の有無を従属変数とした二項ロジスティック回帰分析の結果

| 変数名 | Model 1 | | | Model 2 | | | Model 3 | | |
|-----------------|----------|-----------|-------|---------|-----------|-------|-----------|-----------|-------|
| | Coef. | Std. Err. | オッズ比 | Coef. | Std. Err. | オッズ比 | Coef. | Std. Err. | オッズ比 |
| 男性ダミー | .372 | .251 | 1.450 | .144 | .350 | 1.155 | .739 * | .400 | 2.094 |
| 年齢 | -.175 | .113 | .839 | -.120 | .155 | .887 | -.238 | .177 | .754 |
| 雇用形態 (ref:正規雇用) | | | | | | | | | |
| 非正規雇用 | .004 | .285 | 1.004 | .276 | .419 | 1.318 | -.293 | .402 | .746 |
| 自営・家族・内職 | .194 | .430 | 1.214 | .009 | .639 | 1.009 | .287 | .595 | 1.332 |
| 無職・その他 | .931 *** | .931 | 2.538 | .526 | .467 | 1.692 | 1.411 *** | .534 | 4.101 |
| 最終学歴 (ref:高等学校) | | | | | | | | | |
| 中学校 | -.603 | .491 | .547 | -.724 | .645 | .485 | -.459 | .801 | .632 |
| 専修学校・高専 | .270 | .349 | 1.310 | .576 | .473 | 1.779 | -.055 | .542 | .946 |
| 短期大学 | .599 | .406 | 1.821 | .616 | .656 | 1.851 | .675 | .539 | 1.964 |
| 大学 | .440 | .279 | 1.553 | .458 | .388 | 1.582 | .534 | .414 | 1.706 |
| 定数項 | .641 * | .320 | 1.899 | .530 | .478 | 1.698 | .792 * | .444 | 2.208 |
| -2対数尤度 | 472.977 | | | 244.482 | | | 220.708 | | |
| N | 400 | | | 200 | | | 200 | | |
| Chi sq | 15.715 * | | | 6.266 | | | 16.473 * | | |

(注) *は 10%水準, ***は 1%水準でそれぞれ統計的有意差があることを示す。

以上のことから「調査票<2>」の方が回答者の基本的な属性に偏りがあつたと判断することができる。「調査票<2>」は調査対象者の結婚意識について尋ねたものであり、雇用形態が結婚意識に何らかの影響を及ぼして、このような結果が出たと推察することもできる。例えば、厚生労働省の平成 25 年版の『厚生労働白書——若者の意識を探る』では、正規雇用の人に比べて無職の人が、結婚を「絶対したい」と回答する確率について男性に関しては 12.5%ポイント、女性に関して 7.5%ポイント、それぞれ低いことが示されているⁱⁱ。結婚願望が低い分、無職である方が回答の際に熟慮をする必要がないために、回答をしやすかつた可能性があるⁱⁱⁱ。ただし、これはあくまで推察に過ぎないために、このような偏りが調査票の内容によるものであるのか、単なる偶然であるのかについては検討の余地があるであろう。

4. さいごに

本実習では調査会社の協力を得て、インターネット調査モニターの中から依頼に応じていただいた方の中から、年齢・性別に関して層化無作為抽出を実施し、400 人の調査対象者を抽出した。国勢調査と比較した結果、回答者の属性に大きな偏りはなかった。さらに返送したか否かについて、属性に着目して分析を行った結果、非正規雇用である人の方が返送する傾向にあつたが、全体としてそれ以外の属性が返送するか否かに影響をもたらしていなかった。以上のことから、本実習で得たデータはおおむね良質なデータであると考えられる。しかし、本実習で残された課題もある。

第一に例年の実習と比較した場合の回収率が低くなってしまったことである。本年度の実習の回収率は70.3%であったが、2015年度の実習の回収率は92.3%(下瀬川・池田 2015)、同様に2014年度は92.0%(濱本 2015)、2013年度は94.0%(濱本 2014)、2012年度は93.0%(石田 2013)、2011年度は89.4%(三輪 2012)であった。郵送調査で実査を行う際、母集団の特性を導出するためには、なるべく多くの標本数を設定し、回収率の向上に努めなければならない。回収率の向上の方策として難解な言葉がないか、質問の配列が適切か等、調査票の内容を再検討したり、謝礼金や謝礼品をつける等といった方策があるが(林 2010)、授業の一環として実施するのであれば前者の点に特に力を入れて、今後の回収率の向上に努めるのが望ましいと思われる^{iv}。

第二に本実習を通じて、受講生は調査票の封入、データ入力、クリーニング、データ分析等の社会調査に必要なスキルを習得してきたが、標本設計やサンプリングについて十分なスキルを習得していない。標本数の設定や無作為抽出の実施は社会調査の実施上、非常に重要なものであるが、時間的・予算的制約上の問題で実施することが難しく、講義の一環として概要が説明されるにとどまらざるをえない。

第三に本調査で得られたデータの今後の使用可能性についてである。これは例年の教育学実習でも指摘されている点であるが(石田 2013: 濱本 2015: 下瀬川・池田 2015)、本実習の調査項目は各年度の実習の関心に応じて設定されるために、年度を越えた分析、すなわち時系列的な分析がなされない。ただし、本調査で得たデータを二次分析に活用して、受講生が行っていない分析を行うことのできる可能性は十分に残されているであろう。

以上の課題は抱えているものの、受講生にとっては社会調査の理論や実践に関する知識を習得するだけでなく、自ら主体的に問いや仮説を設定し、分析結果を報告するといった幅広いスキルを身に付けるきっかけになったとも考えられる。来年度の本実習に向けて、さらなる社会調査の理論、実践の知識を習得していく環境整備が必要となってくるであろう。

[注]

ⁱ 昨年度の実習も本年度と同様に郵送調査によって実査を行ったが、その回収率は92.3%であった(計画標本数: 400, 有効票数: 369)(下瀬川・池田 2015: 4)。

ⁱⁱ 厚生労働省, 2013, 『平成 25 年版厚生労働白書——若者の意識を探る』: 79.

ⁱⁱⁱ この想定通りであるとすると、回答の信憑性は低下してしまうということになる。

^{iv} だからといって謝礼金や謝礼品による回収率への効果を見逃して良いわけではない。

[文献]

石田賢示, 2013, 「調査データの特性に関する基礎分析」東北大学教育学部・教育政策科学

- 研究室編『平成 24 年度東北大学教育学部・教育学実習「統計的調査実習」報告書』：1-7.
- 下瀬川陽・池田岳大，2015，「調査設計とデータの特性に関する基礎分析」東北大学教育学部・教育政策科学研究室編『平成 27 年度東北大学教育学部・教育学実習「統計的調査実習」報告書』：1-6.
- 濱本真一，2014，「調査設計とデータ特性に関する基礎分析」東北大学教育学部・教育政策科学研究室編『平成 25 年度東北大学教育学部・教育学実習「統計的調査実習」報告書』：1-8.
- ，2015，「調査設計とデータ特性——若年者のライフスタイルと意識に関する調査の基礎分析」東北大学教育学部・教育政策科学研究室編『平成 26 年度東北大学教育学部・教育学実習「統計的調査実習」報告書』：1-8.
- 林英夫，2010，「郵送調査法の再評価と今後の課題」『行動計量学』37(2): 127-145.
- 三輪哲，2012，「調査設計とデータの特性」東北大学教育学部・教育政策科学研究室編『平成 23 年度東北大学教育学部・教育学実習「統計的調査実習」報告書』：1-7.