

調査設計とデータに関する基礎分析

米田佑・松野広

(東北大学大学院教育学研究科)

1 はじめに

調査を伴う研究では通常、問いの設定、調査の企画、調査の実施、データ入力、クリーニング、分析、報告書や論文の執筆が行われる。2018（平成 30）年度東北大学教育学部開講科目「教育学実習（社会調査の理論と実践）（以下、本実習）」は、受講者がこの一連の過程を学ぶことができるように編成されている。それに加えて、統計的分析手法や統計ソフトの使い方を体得することも本実習の目的の 1 つである。受講生は授業を通して調査の一連の流れや統計的分析手法および統計ソフトの使い方を学びつつ、授業外において各自が関心のある分野の先行研究の読み込み、問いの設定、質問項目の設定、データ入力、クリーニング、分析、報告書執筆を行うことになっている。なお、受講者はランダムに 2 つのグループに分けられ、グループで 1 つの調査票を作成する。ただし、各受講生が自身の興味・関心に沿って上記のプロセスをこなすことになっている。したがって、グループの調査票には各メンバーが自らの興味・関心に沿う質問が組み込まれる。

本稿の目的は、このような流れで実施された本実習の調査の設計と、得られたデータの特徴を整理することである。

2 調査の設計

調査の設計の概要を表 1 に示す。若年層の職業や最終学歴に関するデータ収集を目的としたため、対象者は学生を除く 20 歳以上 40 歳未満の男女に設定した。また、上述したように本実習では 2 つのグループに分かれて調査紙を作成したため、2 種類の調査票をそれぞれ 300、あわせて 600 の調査票を発送した。

調査対象者については、調査会社の協力のもと、前述した条件に当てはまる対象者を層化無作為抽出した。具体的には、インターネット調査モニターに登録しており、協力の依頼に応じてくださった者の中から、年齢、性別、学歴にかんして層化無作為抽出を行った。このような制約のため、本実習の調査対象は「学生を除く 20 歳以上 40 歳未満の日本全国の男女」から全くランダムに得られたサンプルではないことには留意されたい。

(1) の調査の質問項目は、基本属性項目、欲しい子供の数に関する項目、子育て支援に関する項目、幸福・健康に関する項目、非認知能力に関する項目、相対的学歴に関する項目、政治に関する項目、部活・習い事に関する項目、研修に関する項目となっている。(2) の調

査の質問項目は、基本属性項目、非認知能力に関する項目、職場環境に関する項目、飲酒に関する項目、インターネットに関する項目、政治に関する項目、夫婦別姓に関する項目、習い事に関する項目となっている。項目数からも分かるように、(1) の調査の方が (2) の調査よりも回答者が答える質問数は多くなっている。実際、前者の調査のページ数は 15、後者の調査のページ数は 11 となった。

表 1 調査の概要

調査名称	若年層のライフスタイルと意識に関する調査 (1), (2)	
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：各 300 3) 標本抽出法：上記の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年齢、性別、学歴による層化無作為抽出。年齢と学歴は最新の国勢調査を用いて層化。	
調査期間	2018 年 8 月 1 日～8 月 31 日	
調査協力機関	楽天インサイト株式会社【(旧) 楽天リサーチ株式会社】	
調査項目	(1) 基本属性項目 欲しい子供の数に関する項目 子育て支援に関する項目 幸福・健康に関する項目 非認知能力に関する項目 相対的学歴に関する項目 政治に関する項目 部活・習い事に関する項目 文化的体験に関する項目 研修に関する項目	(2) 基本属性項目 非認知能力に関する項目 職場環境に関する項目 飲酒に関する項目 インターネットに関する項目 政治に関する項目 夫婦別姓に関する項目 習い事に関する項目
有効回収票数 (率)	(1) 274 (91.3%)	(2) 270 (90.0%)

表 2 調査スケジュール

年月	授業・調査実習の内容
2018 年	4 月 7 日 実習授業開始
	4 月 20 日 調査会社へ依頼
	6 月 1 日 事前インターネット調査開始
	6 月 20 日 事前インターネット調査終了
	6 月 30 日 調査協力者リスト納品
	7 月 6 日 質問項目締切・調整
	7 月 10 日 調査票最終確認
	7 月 27 日 調査票封入・投函
	8 月 31 日 調査票回答締切
	9 月 4 日 データ入力・クリーニング開始
2019 年	1 月 18 日, 25 日 最終報告会
	2 月 14 日 最終レポート締切

回収率についてみると、各調査のそれは(1) 91.3% (2) 90.0%であり、合わせると 90.7%であった。これは 2017 年度の回収率である 78.1% (伊藤・廣谷 2018) や 2016 年度の回収率である 70.3% (藤井・廣谷・中島 2017) に比べると大幅に改善したと言える。また、若干ではあるが、(1) の調査の回収率が (2) の調査の回収率と比べて高い結果となった。

なお、本実習のスケジュールは表 2 の通りである。

3 欠票の分析

ここでは、本調査においてどのような者が調査を返送しているのかを分析する。なお、ここでは事前のインターネット調査によって集計されたデータを用いている。

まず、それぞれの調査票の対象となったサンプルの記述統計を表 3 に示す。結果を確認すると、ほとんどの項目で調査 (1) と調査 (2) で大きな違いは確認されなかったものの、「従業上の地位・雇用形態」における「非正規雇用」で 4.0%、「自営業・家族従業者・内職」で 2.0%、「無職・その他」で 6.4%の違いがあることが分かる。

次に、返送の有無と「年齢」「性別」「学歴」「従業上の地位・雇用形態」のそれぞれの 2 変数間の関連を分析する。

まず、返送の有無と「年齢」の相関係数を確認すると、全体では 0.031、調査 (1) のサンプルで 0.019、調査 (2) のサンプルで 0.039 であった。係数はプラスであり年齢が上がれば返送しやすいと言えるが、いずれの結果も 10%水準でも統計的に有意ではなかった。

返送の有無と「性別」「学歴」「従業上の地位・雇用形態」については、カイ二乗分析によって関連を確認する。その結果は表 4、表 5、表 6 の通りである⁽¹⁾。

結果を確認すると、有意な関連を示したのは全体および調査 (1) の「学歴」と返送の有無の関連であった。まず、全体を対象とした分析を確認する。「学歴」が中学校の者で返送ありの者は 83.7%、高等学校の者で返送ありの者は 87.3%、専修・高専・短大の者で返送ありの者は 92.1%、大学・大学院の者で返送ありの者は 93.5%と、学歴が高くなるにつれて返送する者の割合が高くなっていることがわかる。調査 (1) についても同じ傾向を示しており、「学歴」が中学校の者で返送ありの者は 77.3%、高等学校の者で返送ありの者は 88.6%、専修・高専・短大の者で返送ありの者は 93.5%、大学・大学院の者で返送ありの者は 94.5%と、学歴が高くなるにつれて返送する者の割合が高くなっている。

以上の結果から、学歴が高い者ほど調査に返送をしやすい傾向にある。ただし、ここでの分析ではその他の変数を考慮していない。したがって、その他の変数を考慮しても同様の傾向が得られるのかを、従属変数を返送の有無 (返送あり=1, 返送なし=0) とした二項ロジスティック回帰分析を行う。結果は表 7 の通りである。なお、Model1 が全体を対象とした分析、Model2 が調査 (1) を対象とした分析、Model3 が調査 (2) を対象とした分析である。

結果を確認すると、有意な値を示したのは Model1 における「大学・大学院」のみであった。その係数は 0.706 であり、基準である「高等学校」に比べて「大学・大学院」である場合に返送する確率が 2.03 倍 ($\equiv \exp(0.706)$) であると言える。

一方で、上述のカイ二乗分析では調査（1）において「学歴」と返送の有無は有意に関連していたものの、ここでの分析では有意な関連は確認できなかった。

表 3 記述統計

	全体				調査（1）				調査（2）			
	min	mean	max	sd	min	mean	max	sd	min	mean	max	sd
年齢	20.0	31.5	39.0	5.3	21.0	31.9	39.0	5.1	20.0	31.1	39.0	5.5
	%				%				%			
返送あり	90.7				91.3				90.0			
性別												
男性	50.0				50.0				50.0			
学歴												
中学校	7.2				7.3				7.0			
高等学校	26.3				26.3				26.3			
専修・高専・短大	35.8				36.0				35.7			
大学・大学院	30.7				30.3				31.0			
従業上の地位・雇用形態												
正規雇用	52.2				52.0				52.3			
非正規雇用	21.0				19.0				23.0			
自営業・家族従業者・内職	6.3				5.3				7.3			
無職・その他	20.5				23.7				17.3			

表 4 性別と返送の有無のカイ二乗分析の結果

	全体			調査票 1			調査票 2		
	返送の有無		χ^2	返送の有無		χ^2	返送の有無		χ^2
	なし	あり		なし	あり		なし	あり	
女性	24 (8.0)	276 (92.0)	0.97	13 (8.7)	137 (91.3)	0.00	11 (7.3)	139 (92.7)	1.81
男性	32 (10.7)	268 (89.3)		13 (8.7)	137 (91.3)		19 (12.7)	131 (87.3)	

表 5 学歴と返送の有無のカイ二乗分析の結果

	全体			調査票 1			調査票 2		
	返送の有無		χ^2	返送の有無		χ^2	返送の有無		χ^2
	なし	あり		なし	あり		なし	あり	
中学校	7 (16.3)	36 (83.7)	6.75 †	5 (22.7)	17 (77.3)	8.04*	2 (9.5)	19 (90.5)	2.04
高等学校	20 (12.7)	138 (87.3)		9 (11.4)	70 (88.6)		11 (13.9)	68 (86.1)	
専修・高専・短大	17 (7.9)	198 (92.1)		7 (6.5)	101 (93.5)		10 (9.3)	97 (90.7)	
大学・大学院	12 (6.5)	172 (93.5)		5 (5.5)	86 (94.5)		7 (7.5)	86 (92.5)	

† p<0.1, *p<0.05

表 6 従業上の地位・雇用形態と返送の有無のカイ二乗分析の結果

	全体		χ^2	調査票 1		χ^2	調査票 2		χ^2
	返送の有無			返送の有無			返送の有無		
	なし	あり		なし	あり		なし	あり	
正規雇用	27 (8.6)	286 (91.4)	0.77	11 (7.1)	145 (92.9)	4.85	16 (10.2)	141 (89.8)	6.19
非正規雇用	14 (11.1)	112 (88.9)		3 (5.3)	54 (94.7)		11 (15.9)	58 (84.1)	
自営業・ 家族 従業者・ 内職	3 (7.9)	35 (92.1)		3 (18.8)	13 (81.3)		0 (0.0)	22 (100.0)	
無職・ その他	12 (9.8)	111 (90.2)		9 (12.7)	62 (87.3)		3 (5.8)	49 (94.2)	

表 7 二項ロジスティック回帰分析の結果

	Model 1		Model 2		Model 3	
	Coef.	S. E	Coef.	S. E	Coef.	S. E
切片	1.837 *	0.898	1.986	1.397	1.929	1.243
年齢	0.012	0.026	0.002	0.041	0.014	0.035
性別 (ref. 女性)						
男性	-0.363	0.309	0.000	0.448	-0.727	0.449
学歴 (ref. 高等学校)						
中学校	-0.256	0.499	-0.534	0.667	0.573	0.842
専修学校・高専・短大	0.445	0.355	0.747	0.548	0.232	0.479
大学・大学院	0.706 †	0.391	0.863	0.595	0.656	0.531
従業上の地位・雇用形態 (ref. 正規雇用)						
非正規雇用	-0.235	0.370	0.516	0.696	-0.649	0.470
自営業・家族・内職	0.055	0.645	-1.027	0.744	16.316	1369.127
無職・その他	-0.083	0.408	-0.300	0.545	0.362	0.698
-2LL	363.92		167.09		182.12	
N	600		300		300	

† p<0.1, *p<0.05

4 さいごに

本実習では、過去2年と比較して大きく回収率が改善した。これは、プリテストなどを通して受講生が調査票の内容を丁寧に検討したことが背景にあると考えられる。特に、今年は受講生として大学院生が各グループに1人ずつ所属していた。このことも、調査票の内容をより丁寧に精査する機会を提供したとも言える。

本実習に残された課題としては過去のデータの活用が挙げられる。これまでも、本実習に残された課題として過去のデータの活用は指摘されてきた（石田 2013, 濱本 2015, 下瀬川・池田 2016, 藤井・中島・廣谷 2017, 伊藤・廣谷 2018）。伊藤愛莉・廣谷貴明 (2018)

も述べているように、毎年受講生が各自の関心に沿って調査票を設計するため、時系列的な分析は難しい。ただし、毎年 TA が主導で同じ質問を尋ねておくなど、何らかの定点観測としての機能として本実習の調査を活かすことができると考えられる。例えば、「あなたは東北大学に対してどのような印象を持っていますか」という質問項目を設定しておけば、東北大学の大学改革のための貴重なデータの一つとして今後活かすことができる可能性もある。また、これまでの回収率の時系列分析などは、個票データが存在しなくても過去の報告書から行える余地がある。今後は、TA の主導のもと工夫して過去のデータを活用していく必要がある。

[注]

(1) 念の為、Fisher の直接確率検定を行なったが 10%水準以下ではカイ二乗検定と結果は変わらなかった。

[文献]

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。

藤井奈々子・中島日向子・廣谷貴明，2017，「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 28 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。

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。

伊藤愛莉・廣谷貴明，2018，「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 29 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。

下瀬川陽・池田岳大，2015，「調査設計とデータの特定に関する基礎分析」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。