

調査設計とデータに関する基礎分析

伊藤愛莉・廣谷貴明
(東北大学大学院 教育学研究科)

はじめに

本稿の目的は、2017（平成 29）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）」で実施した調査の設計と、得られたデータの特性を整理することである。「教育学実習（以下、本実習）」では、調査企画、調査設計、実査、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できるよう編成されている。同時に社会科学的な仮説の構築方法、統計手法を学ぶことも目的の 1 つである。今年度は通年の開講であり、受講者は 2 つのグループに分かれ各グループで調査票を作成した。

このような背景から時間及び予算の制約は免れない。ゆえに、以下ではこの制約下の調査について、その設計を示し、得られたデータの特性及び、無回答者の分析を行う。

1 調査の設計

調査の設計について概要を表 1 に示す。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 20 歳以上 40 歳未満の男女に設定した。また、前項で述べ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 種類の調査表をそれぞれ 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中から、年齢、性別にかんして層化無作為抽出を行った。したがって、本調査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。さらに、調査会社への登録を行っており、調査以来に応じてくださった方々に限定されている点に留意が必要である。

本実習のスケジュールは表 2 の通りである。本年度は通年での開講であったため、前期の調査紙発送からデータの入力に間が空く形となった。前期では調査の実施に関する内容を、後期では具体的な分析手法についての講義を行った。また、受講生はテーマ設定、仮説構築、質問項目、分析結果に関するプレゼンテーションを行うのと同時に、時期に応じて各グループで調査票の作成、データ入力・クリーニングといった実査作業を行った。そのため、授業外の時間に各グループが自主的に作業を行うことが求められた。

なお、今年度も授業内で調査票のプリテストを行った。その際、受講生同士が各グループの調査票を互いに回答し合い、議論した。さらに授業を受けていない学生にも協力をしてもらった。これらは、より回答のしやすい調査票の作成に寄与したと思われる。

表 1 東北大学教育学部平成 29 年度統計的調査実習 調査概要¹

調査名称	若者のライフスタイルと意識に関する調査< 1 >, < A >	
調査目的	現代日本の若年層に対し, 教育や労働, 政治に関する経験や意見を調査すること	
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と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	2017 年 7 月 25 日~8 月 25 日	
調査協力機関	楽天リサーチ株式会社	
調査項目	< 1 > ・基本属性項目 ・インターネットの利用に関する項目 ・大学の大衆化に関する項目 ・キャリア教育に関する項目 ・税金に関する項目	< A > ・基本属性項目 ・地域愛着に関する項目 ・きょうだいに関する項目 ・労働に関する項目 ・コミュニケーションに関する項目 ・部活動に関する項目 ・喫煙に関する項目
有効回収数 (率)	< 1 > 243 (81.0%)	< A > 226 (75.3%)

表 2 東北大学教育学部 平成 28 年後統計的調査実習 調査スケジュール

年月	授業・調査実習の内容
2017 年	4 月 7 日 実習授業開始
	4 月 20 日 調査会社へ依頼
	6 月 16 日 中間報告会
	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
	6 月 30 日 質問項目締切・調整
	7 月 10 日 調査票最終確認
	7 月 25 日 調査票封入・投函
	8 月 25 日 調査票回答締切
	10 月 13 日 データ入力・クリーニング開始
	1 月 19 日, 26 日 最終報告会
2018 年	2 月 13 日 最終レポート締切

2 回収サンプルの特性

ここでは2015（平成27）年度国勢調査と本実習の調査との比較を通じて、本実習で回収されたサンプル属性の特性を検討する。具体的に検討するのは「性別」「年齢」「居住地域」「学歴」「従業上の地位・雇用形態」「職業」の6項目である。このうち「学歴」のみ、平成22年度の国勢調査との比較を行う。国勢調査は日本の人口や世帯構成、就業構造等を明らかにすることを目的として、1920年から5年ごとに行われる調査であるが、学歴や人口移動を含む大規模調査が行われるのは「0」が末尾となる年である。すなわち、2015（平成27）年度の調査では学歴に関する項目が含まれていない。そのため、年は離れてしまうが、2018年3月時点で最新のデータである2010（平成22）年度時点のデータと比較する。分析に際しては、カイ二乗検定を用いる。この分析を通じて、本実習により回収されたサンプルが母集団を代表するものかどうかを検討する。

表3 カイ二乗検定による回収サンプルの特性分析

	国勢調査	平成29年度教育学実習								
	%	全体			調査票<I>			調査票<A>		
		度数	%	χ^2	度数	%	χ^2	度数	%	χ^2
[性別]										
男性	48.7%	182	39.0%	1.454 *	92	38.3%	1.665 *	90	39.8%	1.312 *
女性	51.3%	287	61.2%		151	61.7%		136	60.2%	
[年齢層]										
20～24歳	21.3%	49	10.7%	9.609 **	24	9.9%	9.854 **	25	11.1%	9.48 **
25～29歳	22.9%	190	40.5%		97	39.9%		93	41.2%	
30～34歳	26.1%	112	23.9%		61	25.1%		51	22.6%	
35～39歳	29.7%	118	25.2%		61	25.1%		57	25.2%	
[地域]										
北海道・東北	11.3%	49	10.4%	1.255	28	11.5%	2.296	21	9.3%	1.895
関東	33.8%	149	32.2%		76	31.3%		74	32.7%	
中部	18.3%	105	22.4%		47	19.3%		58	25.7%	
近畿	16.3%	93	19.8%		56	23.0%		37	16.4%	
中国・四国	8.9%	33	7.0%		17	7.0%		16	7.1%	
九州・沖縄	11.4%	39	8.3%		19	7.8%		20	8.8%	
[学歴]										
中学校	18.8%	0	0.0%	26.224 ***	0	0.0%	27.231 ***	0	0.00%	24.802 ***
高等学校	46.5%	239	51.0%		121	49.8%		118	52.2%	
短大・高専	14.8%	45	9.6%		22	9.1%		23	10.2%	
大学以上	19.9%	185	39.7%		100	41.2%		85	37.6%	
[従業上の地位・雇用形態]										
経営者・役員	2.0%	0	0.0%	10.397	0	0.0%	9.258	0	0.0%	12.795 **
常時雇用の正社員・正規職員	58.9%	231	60.3%		118	50.6%		113	50.2%	
パート・アルバイト・その他	25.0%	85	22.1%		46	19.6%		39	17.3%	
派遣	3.0%	45	11.7%		17	7.2%		28	12.4%	
自営業主	7.6%	13	3.4%		5	2.1%		8	3.6%	
家族従業者	3.3%	4	1.0%		4	1.7%		0	0.0%	
内職	0.2%	7	1.8%	5	2.1%	2	0.9%			
[職業]										
専門・技術	17.4%	63	19.9%	7.458	24	17.6%	10.754 *	39	21.5%	6.2
管理	2.6%	5	1.6%		1	0.7%		4	2.2%	
事務	20.8%	96	30.3%		41	30.1%		55	30.4%	
販売	13.8%	42	13.2%		19	14.0%		23	12.7%	
サービス	12.7%	50	15.8%		25	18.4%		25	13.8%	
運輸・保安	13.1%	17	5.4%		6	4.4%		11	6.1%	
生産現場・技能	19.6%	44	13.9%		20	14.7%		24	13.3%	

*** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.10$

分析の結果を示したものが表3である。国勢調査との比較に際して、「調査票<1>」と「調査票<A>」を合わせた回収サンプルの属性の分布、「調査票<1>」のみの回収サンプルの分布、「調査票<A>」のみの回収サンプルの分布、合計3つの分布との比較を行った。全体として、性別と年齢層、学歴に母集団分布からの偏りが見受けられた。母集団と比較して、性別に関しては女性がやや多く抽出される結果となった。年齢層に関しては「20～24歳」が過小になっており、「25～29歳」が過大になっている。さらに学歴に関しては、最終学歴が中学校の回答者が1人もおらず、大卒以上の者が多くなっている。

調査票ごとに詳細に検討すると、「調査票<1>」に関しては、職業に対してもやや偏りが見受けられた。「運輸・保安」に関する職業に従事するサンプルが小さく、一方で「事務」に関する職業に従事するサンプルが多い。「調査票<A>」に関しては「従業上の地位・雇用形態」にやや偏りが見受けられた。「常時雇用の正社員・正規職員」「パート・アルバイト・その他」のサンプルが過小になっており、「派遣」のサンプルが多くなっている。

このような分布の差異が出る理由として、本実習の調査がリサーチ会社にモニター登録している会員を無作為抽出した上で、郵送調査を行っているということが考えられる。サンプルはモニター登録を行っている集団からのみ抽出されるため、モニターに登録しやすい属性をもつサンプルが選ばれやすくなる。本実習では母集団と比較して、女性がやや多めに、学歴が高めに、年齢構成がやや高めに出る結果となった。モニター調査の実施の際に女性の方が抽出されやすいことは本多(2006)でも指摘されている²。中学校が最終学歴である回答者が1人もいなかったのは、比較の基準とする国勢調査が2010(平成22)年度時点のものであり、大学進学率が上昇する昨今、調査時点からさらに高学歴化が進んでいることが1つの背景として考えられる(広田ほか編2013)。年齢構成に関して高めに出たことは、本調査が20～39歳の「非学生」を調査対象としたことが考えられる。すなわち、大学生が調査対象の中に含まれないことで、その年齢層のサンプルが少なくなったという理由である。本調査データの解釈の際には、このような回答サンプルの基本的属性の分布の、母集団からのズレを考慮する必要があるであろう。

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78.2%(計画標本数:600,有効票数:469)であった。調査票別に回収率を検討すると調査票<1>は81.0%(計画標本数:300,有効票数:243)、調査票<A>は75.3%(計画標本数:300,有効票数:226)であった。わずかに調査票<1>の方の回収率が高かった。これらの回収率の数値は、いずれも昨年度の本実習のものを上回る(藤井・中島・廣谷2017)³。以下では、どのような属性をもつ回答者が調査票を返送したのかについて検証する。

本稿では回答者が調査票を返送した場合を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た、二項ロジスティック回帰分析を実施する。回答者の属性としては大きく次の4つを設定した。第1に性別である。本稿では男性の場合に1、女性の場合に0をとる男性ダミー変数を設定する。第2に年齢である。年齢に関しては回答者の年齢をそのまま用いた。第3に雇用形態である。雇用形態に関する変数として、正規雇用を参照基準として、非正規雇用ダミー、自営・家族・内職ダミー、無職・その他ダミーの3つを設定した。第4に最終学歴である。最終学歴に関する変数として、高等学校卒業を参照基準とした、

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.	S.E.	オッズ比	Coef.	S.E.	オッズ比	Coef.	S.E.	オッズ比
男性ダミー	0.172	0.213	1.188	-0.023	0.306	0.977	0.388	0.329	1.474
年齢	-0.021	0.021	0.979	0.024	0.032	1.024	-0.060 **	0.029	0.941
雇用形態 (ref: 正規雇用)									
非正規雇用	0.550 **	0.267	1.733	-0.603	0.629	0.547	0.635 *	0.383	1.887
自営・家族・内職	0.004	0.434	1.004	0.590	1.092	1.804	0.178	0.641	1.195
無職・その他	0.180	0.280	1.197	-0.275	0.321	0.759	0.017	0.396	1.017
最終学歴 (ref: 高等学校)									
専修学校・高専	0.204	0.477	1.227	0.498	0.403	1.646	0.820	0.791	2.271
短期大学	0.199	0.579	1.220	-0.035	0.606	0.966	0.063	0.722	1.065
大学	0.113	0.217	1.120	0.419	0.446	1.521	0.484	0.312	1.623
定数項	1.642 **	0.673	5.168	0.729	1.018	2.072	2.461 **	0.964	11.711
-2対数尤度	620.967			283.144			323.732		
N	600			300			300		
Chi_sq	6.221			5.668			11.454		

** : $p < 0.05$, * : $p < 0.10$

専修学校・高専ダミー，短期大学ダミー，大学ダミーの3つを設定した⁴。第3と第4の独立変数に関して，影響を観察したいカテゴリーを1として，それ以外のカテゴリーを0としている。

分析に際しては分析対象に「調査票<1>」「調査票<A>」の調査対象者をどちらも含めた Model 1，「調査票<1>」の調査対象者のみを分析対象とした Model 2，「調査票<A>」の調査対象者を分析対象とした Model 3 の3つのモデルに基づく分析を行う。このように分析対象を分割することで，それぞれの調査票に係る回答者の属性の偏りを検証できる可能性がある。

分析結果を示したものが表4である。まず Model 1 を検討すると，非正規雇用ダミーに5%水準での統計的有意差が認められた。符号がプラスであることから，雇用形態が非正規雇用の方は正規雇用の方に比べて，調査票を1.773倍，返送する傾向にあることが示される。ただし，モデルの当てはまりは統計的に有意とならなかった。次に Model 2 を検討すると，いずれの変数の統計的有意差も認められなかった。さらに，モデルの当てはまりについても統計的な有意性が認められなかった。これらの結果は，「調査票<1>」を調査対象者が返送するか否かは，調査対象者の属性によって左右されないことを示す。次に Model 3 を検討すると，年齢に5%水準での統計的有意性，非正規雇用ダミーに10%水準での統計的有意性がそれぞれ認められた。年齢の符号がマイナスであることから，年齢が高くなるほど調査票を返送しなくなる傾向にあることを示す。具体的には1歳上昇するごとに0.941倍，調査票を返送しなくなる傾向にある。非正規雇用ダミーの符号がプラスであることは，Model 1 の分析結果と同様に，雇用形態が非正規雇用の方は正規雇用の方に比べて1.887倍，調査票を返送する傾向にあることを示す。ただし，Model 3 についてもモデルの統計的有意性は認められなかった。

以上のことから「調査票<A>」に関して，調査対象者が調査票を返送するか否かは回答者の年齢，雇用形態に左右されるということが明らかになった。「調査票<A>」は労働環境に関する質問項目が多く含まれているのが「調査票<1>」と比較した特徴である。厚生

労働省の2016年の「パートタイム労働者総合実態調査」によると、パートタイム労働者のうち会社や仕事に対して「不安・不満がある」と回答した者は50.1%いる。「パートタイム労働者総合実態調査」は、これまで1995（平成7）年から実施され、2001年、2006年、2011年、2016年の計5回実施されている。各調査のパートタイム労働者の会社や仕事に対する不満を過去の数値から順に検討すると、41.2%、54.3%、62.1%、54.9%、50.1%となっている。1995年の調査を除いて、仕事や会社に不安や不満を感じるパートタイム労働者は5割以上存在することがデータから示されている。不安や不満を感じることから、労働問題への関心が高く、労働に関する設問が多く含まれる「調査票<A>」に、非正規労働者からの回答が相対的に多く集まったと考えることもできる⁵。しかし、モデルの当てはまりの統計的有意性が認められていないことには注意が必要である。より詳細な調査対象者の属性を分析に組み込むことで、調査票を返送するか否かに関して、より精緻な検証を行うことができ、社会調査設計上の有用な情報を提供できると考えられる⁶。

4 さいごに

2017（平成29）年度の本実習は昨年度から計画標本数を増やし、600人を計画標本とした。その結果、昨年度より高い回収率を得ることができた。本調査の回収サンプルの特性を2010（平成22）年度、及び2015（平成27）年度の国勢調査との比較を通じて分析したところ、全体として性別や年齢構成、学歴に母集団とのズレがあった。さらに調査票によっては雇用形態や職業に関するズレも見受けられた。ただし、調査票を返送したか否かに関しては、属性による大きな違いは検討されなかった。母集団からの多少のズレはあるが、本実習により回収されたデータはおおむね良質なデータであると考えられる。

本実習に残された課題としては次の2点が考えられる。第1に、回収率の問題である。本年度の回収率は78.1%であり、昨年度の回収率70.3%（藤井・廣谷・中島2017）は上回った。この背景には、プリテストを通しての、質問項目の選抜、難解な語句の回避、配置順の検討が考えられる。しかし、2015（平成27）年度の回収率92.3%（下瀬川・池田2015）、2014（平成26）年度の回収率92.0%（濱本2014）、2013年度の94.0%（濱本2014）といった過去の回収率と比較すると、低い回収率と言わざるを得ない。今後、調査票の内容の再検討をより丁寧に行う必要がある。

第2に、本調査で得られた過去のデータの活用についてである。この点については過去の実習でも述べられている点である（石田2013、濱本2015、下瀬川・池田2016、藤井・中島・廣谷2017）。本調査は各受講生の関心に基づいて行われるため、時系列的な分析は難しい。しかし、本年度と同様の形式の実習は8年前から行われており、過去のデータの活用法を検討する必要があるといえる。

以上のような課題はあるものの、本実習では、受講生は社会調査の理論、統計的手法の習得だけでなく、自らテーマ設定や仮説構築、データ分析、報告書執筆、さらには、調査票の作成、封入・投函作業、データ入力といった調査の実務も行うことができた。今後も、社会調査の理論と実践を体得するよりよい実習となるように、課題を克服していく必要がある。

[注]

- ¹ 「調査票<1>」の有効回答数(率)に関して、回収率は81.3%(計画標本数:300,回収数:244)であったが、1通のみ年齢が20~39歳から外れているものがあったため、そのような調査票は有効回答数から除外している。
- ² ただし、本多(2006)が想定しているのはインターネット調査を前提としたモニター調査であることには留意が必要である。
- ³ 昨年度の回収率は70.3%(計画標本数:400,有効票数:281)であった。昨年度は『調査票<1>』『調査票<2>』の2種類の調査票を郵送し、1種の調査票につき、200人を計画標本数とした。それぞれ回収率は順に68%,72.5%であった(藤井・中島・廣谷2017)。
- ⁴ 例年、調査対象者の中に最終学歴が中学校の者がいたが、今年度の調査実習では最終学歴が中学校の者が含まれていなかったために、分析から外れている。
- ⁵ 奥西(2008)によると、非正社員は自身の賃金を正社員の賃金と第一に比較する場合に賃金格差への納得度が下がることを指摘している。その一方で、正社員と非正社員間の仕事満足度には大差がないことを、独自のWeb調査を用いた分析から指摘している。このことも踏まえて、非正規雇用者が労働に関してどのような点に不満を感じているかは、検討の余地がある。
- ⁶ 例えば、どのような設問項目を含む調査票であれば、どのような属性をもつ調査対象者からの回答が得られやすいかといったことの、ある程度の予見が可能になると考えられる。

[文献]

- 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.
- 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.
- 広田照幸・吉田文・小林傳司・上山隆大・濱中淳子・白河優治編, 2013, 『大衆化する大学——学生の多様化をどうみるか』岩波書店.
- 本多則恵, 2006, 「インターネット調査・モニター調査の特質——モニター型インターネット調査を活用するための課題」労働政策研究・研修機構編『日本労働研究雑誌』48(6):32-41.
- 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.
- 奥西好夫, 2008, 「正社員および非正社員の賃金と仕事に関する意識」労働政策研究・研修機構編『日本労働研究雑誌』50(7):54-69.
- 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.