

調査設計とデータの特性に関する基礎分析

藤井竜哉・上野春実・熊夢詩
(東北大学大学院教育学研究科)

1 はじめに

本稿の目的は、令和3年度(2021年度)の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)」において実施した調査設計の提示と得られたデータの基礎的な特性(学歴や性別などの基本属性により回答の有無に偏りがあるかないか)を分析することである。本実習では、社会調査の手続きにおける、調査企画・調査票設計・標本抽出・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・報告書執筆の過程を、受講者が実際に体得・会得できるように授業が編成されている。近年の教育学実習の報告書では受講者は2つのグループに分かれていた(藤井奈々子ほか 2017; 伊藤・廣谷 2018; 米田・松野 2019; 藤井竜哉ほか 2020)。だが、昨年度からはグループを分けず、より標本規模を大きくして調査を行っている(神山ほか 2021)。今年度の調査についても昨年度同様の標本規模となっている。本稿では実施した調査の設計について示すとともに、調査から得られたデータに偏りがあるか、あるとすればどのような偏りがあるかを示す基礎的な分析を行う。

2 調査の設計について

まず調査の設計を示す(概要は表1)。本実習における調査は若年層を対象としてその職業や最終学歴に関するデータ収集を目的としている。その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されている。また、一昨年度まで2つのグループに分かれ2種類の調査票を300部ずつ計600部の調査票を送付していたが、昨年度から1種類の調査票を600部送付しており、今年度も昨年度と同様のやりかたとなっている。

表1に示した調査の設計の調査項目について述べる。調査項目は年齢・性別といった基本属性項目、職業に関する項目、現在の暮らしに関する項目、過去の大人とのかかわりに関する項目、結婚相手や結婚生活に関する項目が盛り込まれている。

有効回収率については表1に示した通り90.2%となっており、昨年度よりも高いものとなっていた。ちなみに過去の通年での実習における回収率を付記しておくと、2016年度が70.3%(藤井奈々子ほか 2017)、2017年度が78.1%(伊藤・廣谷 2018)、2018年度が90.7%(米田・松野 2019)、2019年度が91.3%(藤井竜哉ほか 2020) 2020年度が89.5%(神山ほか 2021)である。先までに述べた通り2種類のグループには昨年度から分かれておらず、2016年度から2019年度までと違っており、単純な比較はできない。

次に本実習の調査スケジュールについて述べる。表2に示したのが調査スケジュールである。本実習においては調査を実施するにあたって前期までに調査の実施に関する内容、

後期においては実施した調査の分析に関する内容を授業内にて行っている。その中で前期においては調査企画，調査票設計などの1節に示した内容を実際に体験してもらっている。後期において学生たちは分析方法を学ぶとともに，収集したデータを分析している。

表1 調査の設計

調査名称	若者のライフスタイルと意識に関する調査
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：600 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者から年齢，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	2021年8月1日～9月1日
調査協力機関	楽天インサイト株式会社
調査項目	<ul style="list-style-type: none"> ・基本属性項目 ・職業に関する項目 ・現在の暮らしに関する項目 ・過去における大人とのかかわりに関する項目 ・結婚相手や結婚生活に関する項目
有効回収票(率)	541 (90.2%)

表2 調査スケジュール

年月	授業・調査実習の内容	
2021年	4月8日	実習授業開始
	6月23日	調査会社へ依頼
	7月12日	事前インターネット調査開始
	7月19日	事前インターネット調査終了
	7月20日	調査協力者リスト納品
	7月28日	調査票最終確認
	7月30日	調査票封入・投函
	7月31日	質問項目締切
	9月1日	調査票回答締切
	9月6日	データ入力・クリーニング開始
2022年	1月13日	最終報告会
	2月15日	最終レポート締切

3 欠票の分析

ここでは，本調査において，調査対象者600名のうちの返答者の属性の特徴を分析する。なお，ここでは，調査対象者の属性については，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使用する。

まず，今回の調査で対象となったサンプル全体の記述統計量と，そのうち質問紙に対し返送があったものの記述統計量は以下の通りとなった(表3)。

表3 記述統計

	全体				返送あり			
	mean	min	max	sd	mean	min	max	Sd
年齢	31.2	20	39	5.2	31.39	20	39	5.2
	%				%			
性別								
男性	50.0				49.7			
女性	50.0				50.3			
学歴								
中学校	5.7				5.4			
高等学校	23.3				22.7			
専修・高専・短大	38.7				39.2			
大学・大学院	32.3				32.7			
従業上の地位・雇用形態								
正規雇用	65.2				65.1			
非正規雇用	18.0				17.6			
自営業・家族従業者・内職	3.7				3.7			
無職・その他	13.2				13.7			

次に、調査対象者の属性の違いによって返送の有無に偏りがあるのかを検討する。そのためにもまず返送の有無と「年齢」、「性別」、「学歴」、「従業上の地位・雇用形態」のそれぞれの関連を確認する。

まず返送の有無と「年齢」の関連を確認するため、相関係数を確認すると相関係数は.10であり、有意水準5%で有意な結果となっていた。相関係数の値はそれほど高くはないものの、有意な結果となっており、年齢が高いほど返送するといえる。

次に返送の有無と「性別」、「学歴」、「従業上の地位・雇用形態」について、カイ二乗分析によって関連を確認する。結果を先に述べれば、「年齢」と違い、「性別」、「学歴」、「従業上の地位・雇用形態」の中に有意な関連を示したものはなかった。以下、詳述する。

返送の有無と「性別」については、以下の結果を得た(表4)。男性の場合、返送なしは10.3%、返送ありは89.7%となっており、女性の場合には返送なしが9.3%、返送ありが90.7%となっていた。カイ二乗検定の結果は先にも述べた通り、10%水準でも有意な結果となっておらず、返送の有無と「性別」との間に関連は見られなかった。

次に返送の有無と「学歴」についての結果を確認する(表5)。最終学歴が「中学校」の場合、返送なしは14.7%、返送ありが85.3%、「高等学校」の場合、返送なしが12.1%、返送あり87.9%、「専修・高専・短大」の場合、返送なしが8.6%、返送ありが91.4%、「大学・大学院」の場合、返送なしが8.8%、返送ありが91.2%となっていた。カイ二乗検定の結果は有意な結果となっておらず、返送の有無と「学歴」との間に関連は見られなかった。

次に、返送の有無と「就業上の地位・雇用形態」の関連についての結果を確認する(表6)。就業上の地位・雇用形態が「正規雇用」の場合、返送なしは10.0%、返送ありが90.0%、「非正規雇用」の場合、返送なしが12.0%、返送あり88.0%、「自営業・家族従業者・内職」

の場合、返送なしが9.1%、返送ありが90.9%、「無職・その他」の場合、返送なしが6.3%、返送ありが93.7%となっていた。カイ二乗検定の結果は有意な結果となっておらず、返送の有無と「就業上の地位・雇用形態」との間に関連は見られなかった。

以上の分析から、年齢を除き、性別、学歴、従業上の地位・雇用形態についていずれも、返送の有無とは関係が見られなかった。しかし、以上の分析は単独の変数のみに着目しており、その他の変数を考慮していない。そこで、その他の変数を考慮した場合に同様の結果になるかどうかを検証するため、従属変数を返送の有無（返送あり=0、返送なし=1）とした二項ロジスティック回帰分析を行った。結果は、表7の通りである。

表4 性別と返送の有無の関連

	返送の有無	
	なし	あり
男性	31 (10.3)	269 (89.7)
女性	28 (9.3)	272 (90.7)

$\chi^2 = .17$

表5 学歴と返送の有無の関連

	返送の有無	
	なし	あり
中学校	5 (14.7)	29 (85.3)
高等学校	17 (12.1)	123 (87.9)
専修・高専・短大	20 (8.6)	212 (91.4)
大学・大学院	17 (8.8)	177 (91.2)

$\chi^2 = 2.39$

表6 従業上の地位・雇用形態と返送の有無の関連¹⁾

	返送の有無	
	なし	あり
正規雇用	39 (10.0)	352 (90.0)
非正規雇用	13 (22.0)	95 (88.0)
自営業・家族従業者・内職	2 (3.4)	20 (90.9)
無職・その他	5 (6.3)	74 (93.7)

$\chi^2 = 1.71$

表7 二項ロジスティック回帰分析の結果

	Coef.	S. E	Exp
切片	.33	1.02	2.22
年齢	.06 *	.03	1.07
性別 (ref. 男性)			
女性	.10	.30	1.11
学歴 (ref. 大学・大学院)			
中学校	-.57	.59	.57
高等学校	-.32	.38	.73
専修学校・高専・短大	.00	.36	1.00
従業上の地位・雇用形態 (ref. 無職・その他)			
正規雇用	-.47	.52	.62
非正規雇用	-.63	.55	.53
自営業・家族従業者・内職	-.37	.88	.69
-2 対数尤度	375.76		
N	600		

*: p<.05

結果を確認すると、年齢において統計的に正の有意な結果が見られた。ほかの変数を統制しても返送の有無に影響を与えていることがわかった。つまり、本調査においては返送の有無に年齢による偏りが生じているということがわかった。

4 さいごに

本実習の欠票の分析を行ったところ返送の有無に年齢による偏りが生じていることが分かった。ただし、この結果は近年の調査に関する流れの中では、至極当然のことなのかもしれない。埴淵・山内(2019)は、近年、国勢調査の「不詳」サンプルが増加していることから、インターネット調査によって、どのような属性の場合に国勢調査に対して未提出となりやすいか分析したところ、「若年層や未婚者、単身世帯、短期居住者などが特に未提出になりやすい傾向が確認された。中でも、年齢は未提出に関連する重要な要素である」ことを明らかにしている。すなわち国勢調査ですら年齢による回答の偏りが生じているのである。そのことを考えれば、本調査においても年齢による返送の偏りが生じていることはなんら不思議なことではないのかもしれない。

最後に本実習の課題を挙げる。本実習の課題としては昨年度の報告書(神山ほか 2021)で述べられた調査方法の再考、過去データの活用、回収率に関する次年度以降へのフィードバックといった3点は今年度も挙げられるが、今年度の基礎分析の結果から考えられる課題に絞って述べる。年齢による偏りは近年の調査動向と関連しているかもしれないこ

とを思えば、サンプルがどう偏りやすいかといった調査のトレンドを踏まえて調査対象者のサンプリングを行うことによって偏りのないデータ収集が可能となるであろう。そのため調査のトレンドを踏まえ、サンプリングを工夫することが今年度の基礎分析から浮かび上がった来年度以降の本実習における課題ではないかと考えられる。昨年度からの3点に加え、この点も本実習の課題とし、今後改善していきたい。

[注]

1)セル度数5以下のセルがあったため、フィッシャーの正確確率検定も行ったもののその結果でも有意ではなかった。

[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「社会調査の理論と実践」報告書』:1-7.

藤井竜哉・遊佐賢・伊藤愛莉, 2020, 「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『令和元年度東北大学教育学部・教育学実習「社会調査の理論と実践」報告書』:1-7.

埴淵知哉・山内昌和, 2019, 「国勢調査『不詳』発生に関連要因——インターネット調査を用いた未提出者の分析」『E-journal GEO』14(1):14-29.

伊藤愛莉・廣谷貴明, 2018, 「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29年度東北大学教育学部・教育学実習「社会調査の理論と実践」報告書』:1-8.

神山真由・菅原純一・藤井竜哉, 2021, 「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『令和2年度東北大学教育学部・教育学実習「社会調査の理論と実践」報告書』:4-13.

米田佑・松野広, 2019, 「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30年度東北大学教育学部・教育学実習「社会調査の理論と実践」報告書』:1-6.