

調査設計とデータに関する基礎分析

下瀬川陽・池田岳大
(東北大学大学院 教育学研究科)

1 はじめに

本稿の目的は、平成 27 年度の東北大学教育学部開講科目「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ことである。統計的調査の手続きは、調査企画・調査表設計・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・報告書刊行と進むのが一般的である。本実習は受講者がこれら一連の過程を体験・会得できるよう編成されている。こうした背景から、時間および予算の制約は免れない。以下ではそのような制約下で行われた調査にかんして、その設計を詳らかにし、また得られた調査データの特性および無回答者の傾向をまとめる。

2 調査の設計

調査の設計について、まずは表 1 に基本的事項を調査概要としてまとめた。

表 1. 東北大学教育学部 平成 27 年度統計的調査実習 調査概要

調査名称	若年者のライフスタイルと意識に関する調査
調査目的	現代日本の若年層に対し、教育や労働に関する経験・意見を収集し、それに関わる科学的命題を検証する
調査項目	(1) 基本属性項目 (2) 仕事に関する項目 (3) 祖父母との経験に関する項目群 (4) 資格・学校時代の経験に関する項目群
調査対象	(1) 母集団 学生を除く日本全国の 20 歳以上 40 歳以下の男女 (2) 標本規模 400 (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	2015 年 6 月
調査方法	郵送調査法 (事前の協力伺いと一部質問はインターネット調査)
調査実施協力機関	楽天リサーチ株式会社
回収結果	有効回収票数 (率) 369 (92.3%)

調査の対象を若年に限ったため、対象者の年齢範囲を20歳以上40歳以下と設定した。た

だし、職業やキャリアの情報を収集することを重要視したゆえに、学生は調査対象から除いた。

調査会社の協力により、インターネット調査モニターのなかで前述の対象者としての属性条件に合致する方々へと協力を依頼した。それに応じたなかから、年齢・性別にかんして層化したうえで無作為抽出をおこない、調査対象者を 400 名に絞った。もちろん、この作業における無作為抽出は、通常の世界調査における無作為抽出とは意味が異なる点に留意すべきである。ここでの無作為抽出は、日本の縮図としての標本をつくりだすことをまったく保証しえない。なぜなら、抽出するためのリストが、そもそもインターネット調査モニターであるし、さらにその中で我々の調査依頼に素早く応じてくれた人たちに限定されているからである。

なお、実査および実習全体の詳細なスケジュールについては、以下の表 2 を参照されたい。

表 2. 東北大学教育学部 平成 27 年度統計的調査実習スケジュール

2015 年 4 月 10 日	実習授業開始
2015 年 5 月 22 日	中間報告会（於 東北大学）
2015 年 6 月 1 日	調査会社へと依頼
2015 年 6 月 5 日	事前インターネット調査開始
2015 年 6 月 8 日	質問項目締切・調整
2015 年 6 月 9 日	事前インターネット調査終了，調査協力者リスト納品
2015 年 6 月 10 日	調査票最終確認
2015 年 6 月 12 日	調査票封入・投函
2015 年 6 月 25 日	郵送調査締切
2015 年 6 月 26 日	データ入力・クリーニング
2015 年 7 月 5 日	データクリーニング終了
2015 年 8 月 9 日	最終報告会（於 山形大学）
2015 年 8 月 19 日	実習レポート締切

3 データの特性

以下では平成 22 年度国勢調査と本調査との比較を通して、本調査で抽出されたサンプルの分布の特性を確認していく。具体的には「性別」「年齢層」「(居住) 地域」「学歴」「従業上の地位・雇用形態」「職業」の回答比率について、国勢調査と本調査とのズレをカイ 2 乗検定によって検討した。その結果、「学歴」「従業上の地位・雇用形態」「職業」の 3 つの分布に関して 2 つの調査の間で有意な違いがみられた。まず「学歴」の分布の特徴は、国勢調査に比べて本調査のほうが「大学以上」の比率がより高く、「高校」の比率が低い。次に「従業上の地位・雇用形態」は、本調査のほうが「パート・アルバイト」のようないわゆる非正

規雇用労働に当てはまるものの比率がより高く、逆に「常時雇用の正社員・正規職員」のような正規雇用労働者の比率は低い。3つ目の「職業」は、本調査のほうが「専門・技術職」や「事務職」の比率がより高くなっており、逆に「販売職」や「サービス職」、「生産現場・技能職」の比率が低くなっている。

表 3. 国勢調査と本調査の比較

		H22国勢調査 H27年度実習			
		%	度数	%	χ^2
性別	男性	50.6	183	49.6	0.2
	女性	49.4	186	50.4	
年齢層	20-24	20.2	52	14.1	8.7
	25-29	22.9	89	24.1	
	30-34	26.2	102	27.6	
	35-39	30.7	126	34.2	
地域	北海道・東北	10.5	31	8.5	8.2
	関東	36.5	147	40.4	
	中部	16.5	60	16.5	
	近畿	17.8	70	19.2	
	中国・四国	8.1	31	8.5	
	九州・沖縄	10.6	25	6.9	
学歴	中学校	4.8	10	2.7	37 ***
	高校	34.7	125	34	
	短大・高専	20.5	89	24.2	
	大学以上	24.7	144	39.1	
従業上の地位 ・雇用形態	経営者・役員	1.7	3	0.8	33 ***
	常時雇用の正社員・正規職員	66.5	207	56.7	
	パート・アルバイト・その他	22.6	117	32.1	
	派遣	4.0	13	3.6	
	自営業主	3.1	20	5.5	
	家族従業者	2.0	3	0.8	
	内職	0.2	2	0.6	
職業	専門・技術	20.3	92	26.1	28 ***
	管理	0.6	6	1.7	
	事務	23.0	98	27.8	
	販売	17.2	41	11.6	
	サービス	14.1	44	12.5	
	運輸・保安	5.3	18	5.1	
	生産現場・技能	18.2	51	14.5	
	農業	1.4	3	0.9	

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

こうした分布の違いが現れる理由として、先にも述べたとおり、本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て郵送調査を行っていることが挙げられる。サ

サンプルはモニター登録を行っている集団からのみ選ばれるため、モニターに登録する可能性が高い集団が選ばれやすい。ただし、国勢調査と本調査との間には 5 年のずれがあることにも注意されたい。学校基本調査や労働力調査などより近年の調査データと平成 22 年度国勢調査とを比較してみると、この 5 年の間で「学歴」構成と「職業」構成には大きな変化はみられないものの、「従業上の地位・雇用形態」構成は特に若年層で変化がみられる。つまりここ 5 年の間に非正規雇用労働者の比率が（さらに）増加していることが、本調査回答者における非正規雇用労働者の比率が国勢調査に比して高い要因のひとつであろう。こうした非正規雇用労働者比率の増加という社会構造の変化（周辺分布の変化）を踏まえるならば、本調査は平成 22 年度国勢調査よりもやや高階層に偏った調査であることが、学歴や職業の分布から考えられる。

4 欠票の分析

本年度調査の回収率は 92.3%（計画サンプル数 400、有効票数 369）であった。以下では、回収状況と基本的な属性との関連について検討する。

表 4 は調査表の返送の有無を従属変数とした 2 項ロジスティック回帰分析の結果である。ここでは返送した場合を 0、しなかった場合を 1 としているので、係数の値が大きいほど返送に消極的であると考えることができる。独立変数となる基本属性は性別・年齢・雇用形態・学歴を用いた。これらの変数は、郵送調査に先立って行われたインターネット調査での回答をもとに作成したものである（注 A）。

推定の結果が表 4 である。ほとんどすべての変数の影響が統計的に有意ではなく、返送の有無の傾向が基本的な属性に左右されているとは言い難い。年齢が有意であり 1 歳上がるごとに返送の可能性は 0.91 倍になる（すなわち年齢が高くなるほど返送に消極的）との結果を得たが、この差異がそれほど大きな意味を持つとは考えにくい（注 B）。モデル自体も有意ではないため、今回のサンプルにおいては属性による回収状況の偏りは見られないと判断してよいだろう。これは、インターネットモニターに登録している人の中でも郵送調査へ回答することに前向きであった人をサンプルとして抽出しているためであると考えられる。

5 さいごに

ここまで本実習における調査設計および得たデータの特性を確認した。平成 27 年度の実習では、調査会社の協力のもと、インターネット調査モニターに事前に協力依頼をして、応諾した方々から層化無作為抽出をすることで 400 名からなる標本を得た。標本に含まれた調査対象者からは、郵送調査法で回答が収集された。回収率は 9 割を超えるほど高かった。平成 22 年国勢調査と比較すると「学歴」「従業上の地位・雇用形態」「職業」にやや偏りが見られるものの、回収（欠票）傾向には属性による顕著な違いはなく、おおむね良質なデータであることが推測される。

表 4. 返送の有無にかんする 2 項ロジスティック回帰分析結果

	係数	オッズ比	標準偏差
男性	-.336	.714	.287
年齢	-.092	.912 *	.033
従業上の地位 (基準は正規雇用)			
非正規雇用	-.541	.582	.270
自営・家族・内職	-.880	.415	.440
無職・その他	-.557	.573	.615
学歴 (基準は高校卒)			
中学校卒	1.002	2.724	1.975
専修・高専	-.063	.939	.540
短大	-.314	.730	.592
大学以上	-.360	.698	.328
定数	.750	2.118	2.367
-2対数尤度(切片のみ)		218.097	
-2対数尤度(モデル)		206.206	
N		400	

* p<.05

残された課題としては、次の 2 点があろう。第 1 に、標本設計と抽出である。これは平成 23 年度の実習からたびたび指摘されている(三輪 2012; 濱本 2014)。無作為抽出による標本設計は社会調査の特徴でもあり重要だが、講義の一貫として行われる以上冒頭で述べたとおり時間的・予算的制約があり受講生がそれをする機会を得るのは難しい。何か、標本抽出を直接に体験する方策が求められる。第 2 に、過年度の同実習で得たデータの活用である。この点についても既に石田(2013)および濱本(2014)が言及しているものの、解決には及んでいない。受講生の関心に基づき調査項目が作成される本データは、年度間での共通性が低く、過年度データとの比較といった使い方をするのが難しい。ゆえに講義の終了とともに調査データの分析もなされなくなってしまう。本年度と同様の形式の調査実習は 6 年前から行われている。過年度からの蓄積をうまく活用する術を検討する時期に差し掛かっているといえよう。

[注]

- 1) 事前調査における選択肢のうち、「経営者・役員」「常時雇用の一般従業者」を正規雇用として 1 つのカテゴリにまとめた。また「臨時雇用・パート・アルバイト」「派遣社員」「契約社員・嘱託」を非正規雇用、「自営業主・自由業者」「家族従業者」「内職」を自営・家族・内職として 1 つのカテゴリにまとめた。なお、無職のケースはすべて返送があり、推定上問題があったため、「その他」と回答したケースとまとめた。
- 2) 30 代ダミーを作成し年齢の代わりに投入したが、効果は見られなかった。

[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.

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.

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.