

項目無回答比率の属性による違い

——紛争経験調査 (2017) データによる分析——

お茶の水女子大学 杉野勇

1 ユニット無回答とアイテム無回答

サーヴェイの質の評価には、回収率（ユニット回答率）と、ユニット回答者の基本的属性についての代表性が用いられるのが一般的である。しかし、回収率はユニット無回答（の少なさ）だけを評価しているが、実際のデータ分析ではアイテム無回答（以下、項目無回答）も欠損値に違いなく、データの質の重要な目安である。しかも、学術調査によく見られる複雑な調査票では、一見した回収率よりもはるかに「欠損値」が多くなり得る。「欠損値」には、答えなくてよい／答えるべきではないので回答がない「非該当」と、答えて欲しいのに回答が得られなかった「無回答」があり、データの質を左右するのは、通常は后者である（但し、回答者の satisficing 行動の結果として「非該当」になるとしたらそれも実はデータの質に関わる）。各変数ごとの非該当の数・割合や無回答の数・割合は、データがきちんとして入力されてさえいれば、度数分布表で分かる。しかし、調査会社から納品されるデータファイルはしばしば非該当や無回答の区別が無頓着で一貫性が無い事がある。そして、各回答者毎に、どの程度の非該当質問（数・割合）や項目無回答（数・割合）があるのかは、通常はなかなか確認しない。本報告では、或程度明確な形で、項目無回答が人口学的或いは社会経済的属性によってどの程度異なるのかを示す。

2 用いるデータとエディティング処理

用いるのは、紛争経験全国調査（2017年）のデータである（基盤研究S「超高齢社会における紛争経験と司法政策」）。全国20歳以上男女12,000人に郵送・ウェブ併用方式で、過去5年間の日常生活における様々なトラブル経験をきき、特に重要なもの一つについて対応行動を詳細に尋ねた調査である。回収率は4,693人（回収率39.1%）、質問総数は（数え方によるが）データフレームの変数の数にして387個程度。中には、複数の質問からなる大きなブロックもあり、多くの学術調査同様条件分岐も多い。

本報告ではウェブ回答者706名は除外して郵送回答者3,987名のみを分析する。また、使用するデータファイルは、かなりデータ・エディティング（データ・クリーニング）を行ったものである（そうでないと非該当と無回答が正しく区別されていない為）。

3 属性による項目無回答比率の違い

非該当質問数は平均225変数（範囲11～298）、要回答質問数は89～376（平均162）変数であった。各回答者毎に項目無回答の変数の数を要回答の変数の数で割って項目無回答比率とした（変数の重要性の差は考慮していない）。重大な分岐質問の無回答パターンで5つのグループに分け、どの非該当パターンになるかを多項ロジットで、項目無回答比率を重回帰分析で分析した。その結果、予想通り高年齢層で項目無回答比率が高く、教育年数は無回答パターンと関連していた。また、女性において項目無回答比率が高い傾向も示唆された。職業や収入との関連はやや解釈が難しい。なお、調査への関心度を表す代理変数として、後日改めて（半構造化）面接に応じてくれるか否かと、結果の速報版を希望するかの回答も投入したが、後者の関連は余り明瞭でないものの、前者はやはり項目無回答比率の低さとはっきりと関連していた。

本研究はJSPS科研費16H06321, 18H03649の助成を受けたものです。