

# 複数WEB調査におけるシングルソース集計の 効用と事例

○佐藤寧  
株式会社日経リサーチ

## 1. はじめに

WEB調査において、サンプリングフレームとして公募型のWEB調査モニターを対象とした場合には、標本は国民全体などの母集団から無作為に抽出されたとはいえず、世論調査など統計的な調査として活用することはできない。(佐藤 2011)しかし、WEB調査の活用は 2011 年の現在にあって、市場調査の分野において非常に高いシェアを占めるまでになっている。これは、WEB調査にコストが安いこと、調査プロジェクト期間を短縮できること、特定の限られた層に対する調査が容易であること、調査票に様々な工夫ができることといった様々なメリットがあり、これらが有用に活用されていることが大きい。しかし、WEBモニター調査の活用においては更なる工夫の余地があり、従来の調査法では為し得なかった多角的な情報として集計することができる。このような取り組みの1つが、今回提唱する「複数WEB調査におけるシングルソース集計」である。

「シングルソース」または「シングルソースデータ」という言葉は、広告効果測定などの分野において、同一対象者から、購買や広告接触、ライフスタイルなど複数の情報を複数のソースから入手して広告と購買の関係などを分析する手法として既に活用されている。今回提唱する手法は、複数のWEB調査における同一対象者の回答を紐付けしたうえで集計・分析する手法として一般化したものである。

## 2. 活用の前提条件

前述したように、WEB調査は理論的に統計的な調査として用いることが出来ない。特にWEB調査のモニターは、特定の意識、価値観に偏った集団であると考えられるため、意識調査においては特に注意して用いるべきである。このようなテーマの場合には、「時系列的な傾向の把握」「クロス集計レベルでの傾向の把握」といった傾向分析に限定して用いるべきである。(佐藤 2011)今回の報告テーマも、WEBモニター調査の活用を前提としたものであり、活用場面はこれらに留めなくてはならない。

## 3. 従来型WEBモニター調査における課題

### 課題1: 同じ設問を、同一モニターに対して何度も設けることの非効率性

WEB調査では、その他の調査手法と同様に、集計分析に必要な項目をそのつど調査票に盛り込んでいる。しかしながら、WEB調査モニターの立場から見ると、依頼されるいくつもの調査で、全く同じ設問を何度も調査されることになる。例えば、性別や年代、職業、家族構成、ライフスタイルやライフステージといった項目については調査の度に回答することになってしまう。ライフスタイルを判別するための設問などについては時に数十問単位となることもあり、回答者への負担が大きいだけでなく、調査のコストにも影響する問題である。

### 課題2: 求められる情報量と、1回あたりの調査における設問の長さの限界にギャップ

多角的に調査結果を分析したい場合、必然的に1つの調査あたりで設問数が多く設ける必要がある。情報量が多いほど様々な分析が実行できるが、設問数が多すぎると、回答者への負担が大きいため、1つの調査での設問数には限界があると考えられる必要がある。

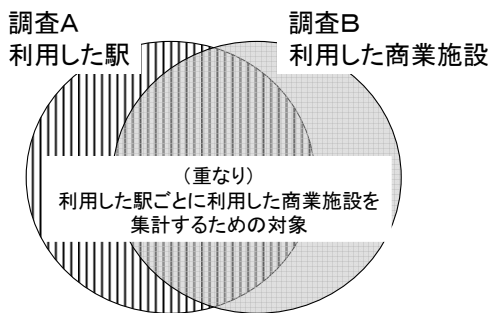
この2つの課題を解決する方法として「複数のWEB調査のシングルソース集計」を提起する。

#### 4. 複数WEB調査におけるシングルソース集計の定義と特性

複数WEB調査におけるシングルソース集計は、複数のWEB調査のうち、同一の回答者を特定したうえで、データを紐付けて集計することと定義する。例えば、調査Aで利用している駅を、調査Bで利用している商業施設を調査した場合、調査Aと調査Bをシングルソース集計することで、利用している駅と商業施設利用のクロス集計をすることができる。同様の理屈により、3つ以上の調査であっても相互にクロス集計することもできる。逆にとらえると、設問量の多い調査を複数の調査に分割することが出来るということで、1回あたりの調査設問数を少なくすることができる。これは、前述した課題1、課題2を解決するものである。

異なる調査の回答者は同一ではないため、集計対象となる設問の両方に回答した人のみを集計の対象とする。[イメージ:図1]このことから、シングルソース集計を行なった結果は、通常の調査における集計表と異なり、集計表ごとに集計対象となる回答者数が異なることになる。[例示:表 1] 集計対象とできる回答者数は、個々の調査における回答者数から小さくなるが、それでも後述する工夫によってこの重なりは個々の調査回答者数のうち60%~70%程度にまで高めることができる。

[図1]



[表1]

1つの調査内でのクロス集計		回答サンプル数	男性20代以下	男性30代	男性40代	男性50代以上	...
分析対象	伊勢丹 新宿店	1832	7.0	10.0	6.7	11.5	
比較対象	高島屋タイムスクエア 新宿店	1689	9.0	11.4	8.7	10.7	
	ルミネ新宿	1871	10.1	8.1	5.6	7.8	
	新宿三越アルコト店	402	9.3	9.1	5.2	4.9	

別調査Aとのクロス集計		回答サンプル数	流行、はやりもの好き	ファッション・デザインにこだわり	高級ブランド好き	ファミリーを大切に	...
分析対象	伊勢丹 新宿店	1086	13.2	16.7	12.9	9.0	
比較対象	高島屋タイムスクエア 新宿店	993	9.9	14.6	10.8	11.4	
	ルミネ新宿	978	12.9	16.7	10.6	9.0	
	新宿三越アルコト店	221	13.8	23.9	17.6	9.5	

別調査Bとのクロス集計		回答サンプル数	JR山手線	JR中央線(快速)・中央本線	東京メトロ丸の内線	JR中央・総武線	...
分析対象	伊勢丹 新宿店	1168	38.9	18.6	18.6	15.5	
比較対象	高島屋タイムスクエア 新宿店	1072	43.1	22.9	15.6	17.3	
	ルミネ新宿	1088	43.0	21.6	14.8	17.8	
	新宿三越アルコト店	250	45.8	25.6	29.5	23.2	

#### 5. シングルソース集計の実践に必要なことと、その検証

調査Aと調査Bの同一回答者を集計対象とする特性上、できるだけこの重なり(両方の調査に回答した人)を大きくすることが望ましい。このために実施すべきことは以下の(1)~(3)となる。

- (1) 2つの調査にできるだけ同一対象者を含むように調査の依頼をする。ただし、全く同一の対象者である必要はない。例えば、日経リサーチの事例では、日経リサーチのWEB調査モニターのうち首都圏の1都3県に居住している全数を調査の対象とした調査を複数回実施している。調査の時期が異なると、モニターを辞める人や新たにモニター登録する人がおり、調査対象は全く同じではないが、大部分が同一の対象者である。
- (2) 回収数が一定数に達しても調査終了しない。WEB調査では、回答数設定型(例えば回収数が1000に達したらサイトを閉じ調査を終了する)のケースが多い。このようにしてしまうと、調査依頼対象者がほぼ同一であっても、回答者が異なる人が多くなり、結果的に2つの調査の同一回答者が少なくなってしまう。
- (3) 調査依頼対象者からの回収率を出来る限り高める。これは(2)と同様の理由による。回収率を高める手段として最も有効なのは、調査の実施期間を長めに設定することと、未回答者に対して定期的に回答の督促を行なうことである。

調査の督促による回答の状況と、督促の回数による回答の重なり(割合)についての具体的な事例を、当日の発表で紹介する。

## 6. シングルソース集計の効用

複数WEB調査におけるシングルソース集計は、通常のWEB調査と比較して以下の(1)～(3)の効用がある。(1)(2)は既に述べたWEB調査の問題点への解決にあたるが、これ以外にも(3)という副次的な効用が得られる。

- (1) 同一の設問内容を複数の調査に盛り込む必要がなく無駄がない。
- (2) 1つの調査では設問数が多すぎるような場合に、複数の調査に分割することができる。
- (3) 1つの調査が終わった後に新たな調査課題が発生したり、分析したい項目が増えたりした場合に、新たな調査にて分析項目を追加できる。

調査の設計をいくら慎重に行なっても、分析の過程で設問に盛り込めていなかった課題が浮上することは良くあることである。(あの設問を入れておけば・・・選択肢をこうしておけば・・・といった反省は調査設計者であれば誰でも経験したことがあるのではないだろうか)しかし、複数調査をシングルソース集計するのであれば、次の調査機会に希望する設問を盛り込めば、前回調査データとのクロス集計が可能である。また、調査を実施するたびに新たな情報が追加されることとなり、調査回答の膨大なデータベースが構築されることとなる。この大きな情報は、従来では実現できなかった多彩な分析を可能とする。

## 7. シングルソース集計の実例

日経リサーチでは、シングルソース集計を全面的に活用した調査商品「首都圏センサス」をASPで提供している。本商品の主な目的は、「駅や鉄道の利用」「商業施設の利用」が、どのような人に、どのような内容でされているのかを集計するためのものである。様々な切り口での分析ニーズに応えるため、シングルソース集計を導入している。2011年6月時点で集計可能な項目を[表2]に記す。59の項目での集計が可能である。設問数は59だが、中には、選択肢が100を超えるようなものもあり、非常に膨大な項目が集計対象となっている。集計表の実例を[図2]に示す。神田駅の各線(山手線、京浜東北線、中央快速線、銀座線)利用者が、どの商業施設を利用しているのかを集計したものである。普段神田駅を利用している人(2010年6月、10月、2011年2月いずれかの調査で神田駅を利用したと回答した人)が、2010年9月調査時点で利用していた(調査時点から3ヶ月以内で利用した)商業施設を明らかにしている。同じ神田駅の利用者でも、利用する線によって、利用している商業施設が異なることがわかる。このような情報は、商業施設に関連する交通広告の掲出駅選定などに活用されている。

[表2]首都圏センサスで集計が可能な項目と、データソース

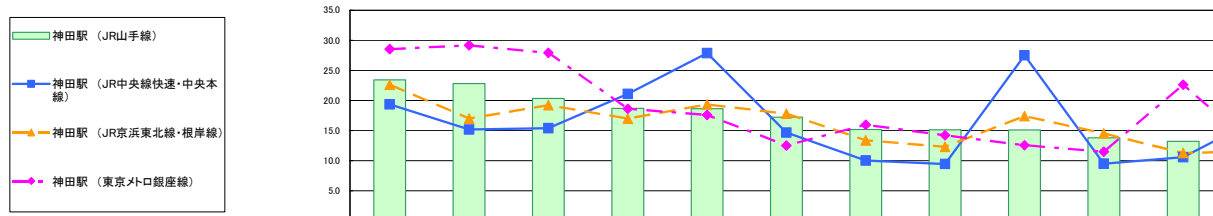
設問	データソース	設問	データソース
1 商業エリア・商業施設利用目的	2010年9月商圏調査	31 1ヶ月あたり自由に使えるお金・小遣い	最新の調査(駅または商圏調査)
2 商業エリア・商業施設利用頻度	2010年9月商圏調査	32 ライフスタイル・意識(先端層)	2010年3月日経リサーチ自主企画調査
3 商業エリア・商業施設来場手段	2010年9月商圏調査	33 ライフスタイル・意識	2010年3月日経リサーチ自主企画調査
4 商業施設への来場距離	モニター登録情報	34 利用している衣料品ブランド	2010年2月日経リサーチ自主企画調査
5 商業エリア・商業施設評価(リピート意向)	2010年9月商圏調査	35 利用している化粧品ブランド	2010年2月日経リサーチ自主企画調査
6 商業エリア・施設での利用金額	2010年9月商圏調査	36 利用しているバッグ・アクセサリブランド	2010年2月日経リサーチ自主企画調査
7 商業エリア・施設の平日・休日利用状況	2010年9月商圏調査	37 興味・関心ごと	2010年2月日経リサーチ自主企画調査
8 商業エリア・商業施設に誰と行っているか	2010年9月商圏調査	38 興味・関心ごと(最近の興味関心ごと)	2010年2月日経リサーチ自主企画調査
9 商業エリア・商業施設滞在時間	2010年9月商圏調査	39 読んでいる雑誌	2010年2月日経リサーチ自主企画調査
10 駅の利用目的	2010年6月～2011年2月駅調査	40 利用しているコンビニエンスストア・駅売店	2010年10月駅調査
11 路線のイメージ	2010年6月～2011年2月駅調査	41 利用しているスーパー・専門店	2010年10月駅調査
12 駅でしていること	2009年6月駅調査	42 利用しているファミレス・ファーストフード・カフェ・その他飲食店	2010年10月駅調査
13 駅や路線の平日・休日利用状況	2010年6月～2011年2月駅調査	43 利用している駅売店	2010年10月駅調査
14 電車の中でしていること(201102駅調査)	2011年2月駅調査	44 利用している施設やスポット	2009年6月駅調査
15 駅でしていること(201102駅調査)	2011年2月駅調査	45 情報を得るきっかけになっているメディアや広告	2011年2月駅調査
16 鉄道利用頻度	2010年3月商圏調査	46 利用しているオンライン情報サービス・コミュニケーションサイト	2010年10月駅調査
17 鉄道利用目的	2010年3月商圏調査	47 自動車での外出頻度	2010年2月駅調査
18 性年代10歳刻み	最新の調査(駅または商圏調査)	48 居住年数	2010年2月駅調査
19 性年代5歳刻み	最新の調査(駅または商圏調査)	49 今後の居住意向	2010年2月駅調査
20 性年代FMI・3分類	最新の調査(駅または商圏調査)	50 よく飲んでいる缶コーヒー	2010年10月駅調査
21 全体・性別・年代	最新の調査(駅または商圏調査)	51 利用したネットショッピングサイト	2011年2月駅調査
22 職業区分	最新の調査(駅または商圏調査)	52 居住の最寄駅	最新の調査(駅または商圏調査)
23 家族構成	最新の調査(駅または商圏調査)	53 利用した駅	2010年6月～2011年2月駅調査
24 同居家族	最新の調査(駅または商圏調査)	54 居住の最寄路線	最新の調査(駅または商圏調査)
25 住居形態	最新の調査(駅または商圏調査)	55 利用した路線	2010年6月～2011年2月駅調査
26 居住地エリア(大分類)	モニター登録情報	56 利用した商業施設	2010年9月商圏調査
27 居住地エリア(市区町村ベース)	モニター登録情報	57 利用した商業エリア	2010年9月商圏調査
28 居住地CAMIEコード	モニター登録情報	58 通勤通学の最寄駅	2010年6月～2011年2月駅調査
29 居住者プロフィールデータ(首都圏版)	モニター登録情報	59 通勤通学の最寄路線	2010年6月～2011年2月駅調査
30 世帯の年収	最新の調査(駅または商圏調査)		

[図 2]シングルソース集計結果の一例(神田駅各線利用者が利用している商業施設)

駅・路線分析「利用した商業施設」

集計内容: 利用した商業施設  
 集計対象: 全体  
 集計方法: 利用者の分布 (%)  
 ブレイクダウン: 全体

調査データ(表頭): 2010年9月商圏調査  
 調査データ(表側): 2010年6月~2011年2月駅調査



		回答 サンプル 数	ヨドバシAkiba ビル(マルチ メディア)	ビックカメラ有 楽町店本館	銀座三越	ルミネ新宿	高島屋タイム ズスクエア 新宿店	伊勢丹 新宿 店	丸の内オアゾ	八重洲地下 街	ヨドバシカメラ 新宿西口本 店	有楽町 ITOCIA(有楽 町マルイ)	松屋 銀座店
分析対象	神田駅 (JR山手線)	483	23.4	22.8	20.3	18.7	18.6	17.2	15.2	15.1	15.1	13.8	13.2
比較対象	神田駅 (JR中央線快速・中央本線)	310	19.4	15.2	15.4	21.1	27.9	14.7	10.0	9.4	27.6	9.5	10.6
	神田駅 (JR京浜東北線・根岸線)	157	22.6	17.0	19.2	17.0	19.4	17.8	13.4	12.3	17.4	14.5	11.3
	神田駅 (東京メトロ銀座線)	133	28.5	29.2	27.9	18.6	17.6	12.5	16.0	14.2	12.6	11.5	22.6

8. シングルソース集計の課題

シングルソース集計の手法、考えとも極めてシンプルだが、まだほとんど活用されていないのが現状である。シングルソース集計がより活用されるためには、以下のような課題を乗り越える必要がある。

- 大規模データの集計法の確立  
 仮に回答データを全て連結して1つの大きなデータを構築してしまうと、非常にファイルサイズが大きな情報となり集計作業には不向きである。ファイルを分割して保管したうえで、集計に必要な情報を都度マージして利用することが効率的であり、これに対応することが求められる。大規模データで速やかにこれに対応できる仕組みが必要である。
- 回答結果が矛盾するデータにおける補正  
 WEB調査では通常、矛盾のある回答が合った場合には調査回答中に回答者に対して注意を促し、矛盾のない回答を入力してもらうようにしている。しかし、複数の調査間では矛盾を系統的に排除するのは極めて難しい。しかし、矛盾のある回答を矛盾なきように機械的に修正したり、矛盾のある回答を全て排除したりすることは、逆に調査の品質を低下させる場合もある。このようなケースでどのように対応すべきか、基準を明確にする必要がある。
- データの活用  
 WEBモニター調査におけるシングルソース集計によって、従来は実施した調査が個々に切り離されていた複数のソースであったものが、モニター全体で1つのソースとして取り扱われる。ここから得られるものは、「調査結果」というよりは、調査という手段を通じて集められた「情報」と考えたほうがよい。この情報から作成可能なクロス集計だけでも膨大なパターンとなる。情報を最大限に活用するためにはデータマイニングの手法などを活用して効率的に有益な情報を抽出することが求められる。

参考文献

[1]佐藤寧 (2010) 「市場調査における調査の品質とWEBモニター調査での取り組み」 日本行動計量学会 35 回大会抄録集  
 [2]佐藤寧 (2011) 「WEB調査を活用するにあたって その特性と課題」 世論調査協会 よろん第107号  
 [3]佐藤寧 (2011) 「日本におけるWEB調査の現状と課題」 早稲田大学・復旦大学合同シンポジウム 日本と中国の世論調査とその技法 抄録集

連絡先 yasushi\_sato@nikkei-r.co.jp