

**STATISTICAL DIGITAL ARCHIVE OF
AGRICULTURE, FORESTRY AND FISHERIES**



Working Paper Series No. 1

農業統計の高度利用に当たっての課題

島田 依佐央

WORKING PAPER SERIES

学術情報メディアセンター
食料・農業統計情報開発研究分野

Working Paper Series No. 1

農業統計の高度利用に当たっての課題

島田 依佐央

2014 年 3 月

京都大学農林水産統計デジタルアーカイブ講座

Academic Center for Computing and Media Studies,
Agricultural Economics and Information Laboratory

本 Working Paper は、京都大学寄附講座 農林水産統計デジタルアーカイブ講座に派遣された、農林水産省大臣官房統計部からの出向職員が、その研究支援業務に関連して行った成果を紹介するためのものである。ただし、本資料に示された見解は、執筆者の個人的見解である。

農業統計の高度利用に当たっての課題

1. はじめに

京都大学大学院農学研究科農林水産統計デジタルアーカイブ講座では、これまで農林水産省大臣官房統計部（以下、統計部という。）と連携して、同省が実施した統計調査の調査票情報（以下、個票データという。）の高度利用を推進するためのデータ基盤整備を進めてきた。この高度利用の推進に当たっての課題について、以下のとおり整理したので述べる。

2. 調査全般に渡る課題

1) データの保存期間

個票データの高度利用を行うためには、データが磁気化されている必要がある。そこで統計部から、平成 24 年 8 月に磁気媒体で存在する個票データの一覧について別添 1 のとおり提供を受けた。これをみると、一定年次（おおむね平成 17、18 年）以前の調査の個票データについては、磁気データが整備されておらず、また調査ごとに整備されている年次にばらつきがある。

磁気化された個票データの保存期間に関しては、『調査票情報等の管理及び情報漏えい等の対策に関するガイドライン』（平成 23 年 3 月 28 日改正（平成 21 年 2 月 6 日、総務省政策統括官（統計基準担当）決定））のなかで、「調査票情報及びドキュメントの保存期間は、基幹統計調査において施行規則等を省令等に規定している場合はその保存期間とし、それ以外については「常用」又は「無期限」として保存し続ける必要がある」とされている。

統計部においては、上記ガイドラインの改正を見越して、平成 22 年 3 月に個票データの保存期間を永年とするよう省内に通知がなされたようであるが、通知以前は個票データの保存期間に関する統一的な基準はなく、調査ごとにまちまちに保存期間が設定されていた。このため、先に指摘したとおり、磁気媒体で存在する個票データの年次にばらつきがみられるものと考えられる。通知以前においても、地方統計組織において磁気化された個票データが永年保存とされている調査（作物統計調査、木材統計調査など）も見受けられることから、別添 1 に掲げられた年次以外にも、さらに遡って過去の個票データを収集できる可能性がある。改めて個票データの残存状況を確認する必要がある。

現在統計部では、調査要領において、磁気データの個票データの保存期間は永年とされているが、調査対象リストについては依然として一定の保存期間が設定されている。このため、保存期間が経過すると調査対象リストに記載された個票データの属性情報（名称、住所、他統計調査とのリンクコードなど）が失われてしまうこととなる。

前出の『調査票情報等の管理及び情報漏えい等の対策に関するガイドライン』では、調査対象リスト等の「調査関係書類」は、「基幹統計調査において施行規則等を省令等によって規定している場合はその保存期間とし、その他については事務及び事業の性質、内容等に応じて調査実施者において行政文書管理規則に基づき定めること」とされている。例えば集落営農農態調査要領（平成 23 年 9 月）において、関係書類等の保存について表 1 のよ

うに整理されており、調査標本リスト（調査対象リスト）の保存期間は5年となっている。

調査対象に永久的に固定化された番号が付けられていれば、この番号を利用することで容易にパネルデータ化が可能である。しかしながら、年次によって調査対象の番号の付け替えが行われている場合、調査対象リストを用いた名寄せによる方法がパネルデータ化を行ううえで有効となる。こうしたことから、調査対象リストについても、個票データと一体的に永年保存される必要があると考えられる。

2) 調査ドキュメントの整備

調査票情報の高度利用を行うため必要となるのは、個票データを読み解くための関連情報の整備である。しかしながら、例えば平成6年以前に実施されていた農家経済調査については、調査要領、手引き類の調査の仕様を解説したドキュメントはほとんどが廃棄されており、一部資料しか保存されていない。

調査要領・手引き類の資料については、省内の文書管理規則において保存期間が定められており、保存期間の経過したものについては廃棄されている場合がある。調査要領・手引き類の資料は、個票データを理解するうえで重要な資料となるため、個票データと一体的に永年保存していく必要がある。

3) 符号表の整備

個票データの構造を理解するうえで必要な、データレイアウトの構造を示す符号表を整備することが重要である。

『統計調査等業務の業務・システム最適化計画』（平成18年3月31日、各府省情報化統括責任者（CIO）連絡会議決定）のなかで、「個票データ（個々の調査票の回答内容を記録したデータをいう。）のレイアウト構造（乗率や回答事項の配列、符号内容、データの型等）を示す記法及び符号表等のデータ形式について、『政府統計個票データレイアウト標準記法』に準拠するもの」とされている。

統計部においては、当該最適化計画に基づく符号表は基本的に作成されていないため、個票データの利用者は、符号内容等の個票データの構造を把握することなしに、調査票様式をみながら集計計画を立てることになる。例えば、農林業センサスにおいては、調査項目とは別に主副業別分類等の集計分類コードのデータが付与されているが、この事実は一般的には知られていない。利用者が統計法第33条に基づく個票データの申請に係る検討段階で集計分類コードの存在を知っていれば、より効率的に集計計画を立てることが可能になると考えられる。符号表を整備し、個票データ利用者に対して、集計計画検討段階での積極的な提供が求められる。

4) 推計倍率の提供

標本調査の場合、全数調査とは異なり、公表される集計結果は推計により求められる。このため、標本調査の個票データを利用する際には、集計結果を再現するための推計倍率（乗率）が必要となる。

推計による集計には大きく2つの方法（図1）があり、一つは個票データの調査結果に推計倍率を乗じて、その結果を積み上げて集計を行う方法であり、統計部の調査では、農

業経営統計調査が該当する。もう一つは個票データの調査結果を積み上げた結果に推計倍率を乗じて集計を行う方法であり、統計部のほとんどの調査がこの方法に該当するものと思われる。後者の集計方法の場合、統計部内に構築された統計システムの中で、システムに登録された母集団情報や標本調査結果から自動的に推計倍率が計算されて集計が行われていると考えられ、各個票データに推計倍率が付加されてデータが保存される仕組みとなっていない。このため、統計部において個票データの提供時に、あわせて推計倍率を提供することが困難な状況となっている。

なお、推計倍率が提供されないと、提供を受けた個票データを用いて公表値を復元させることができず、公表値と一致するかどうかの検証が行えない。また、組み替え集計をも行うことができないという問題がある。標本調査については、個票データとあわせて推計倍率が提供できるよう仕組みを講じる必要がある。

3. 個別調査における課題

これまで統計部から個票データの提供を受けた調査の一部について、調査ごとに課題を整理する。

1) 農林業センサス

(1) 個票データの保管状況

農林業センサスの磁気化されているデータの保管状況をみると、世帯・事業体に対する調査では、農家調査は 1975 年以降、農家以外の農業事業体調査は 1985 年以降、農業サービス事業体調査は 1995 年以降、林家調査は 1990 年以降、林家以外の林業事業体調査および林業サービス事業体調査は 2000 年以降の調査の個票データが保管されている。また、地域に対する調査では、農業集落調査が 1990 年以降、林業地域調査が 2000 年以降の調査の個票データが保管されている。このように調査の種類によって、個票データの残存状況は異なっているが、これは 2000 年センサス以前の調査要領（要綱）において、磁気化された個票データの保存期間は 10 年とされていたことが個票データ散逸の要因の一つとして考えられる。

また、1975 年以降残存している農家調査についても、1990 年以前の個票データは一部項目の欠落があるなど不完全な状態であるとの条件付きでデータが提供されている。1980 年、1985 年、1990 年の農家調査の個票データには、表 2（1980 年センサスの例）のように、基本指標コードの一部である「集計単位地域」と「調査区」のコードが欠落しており、基本指標コードがユニークなコードとならないことから、個票レベルでの年次間リンクに支障をきたす状況にある。また 1975 年の個票データは、データを 2 つに分割したファイルが提供されているが、表 3 のように、2 つのファイルで、都道府県コード別のレコード数が一部の県において一致していない。

農林業センサスについては、調査規模が大きくかつ継続的に調査されているため、パネルデータとして有用であることから、残存データを探索し、適正に保存された個票データの発掘が望まれる。

（２）集計論理の整備

農家調査（2005 年以降は農林業経営体調査）については、個々の農家（または農業経営体）の経営耕地面積規模や農産物販売金額規模等について、前回のセンサス結果と個票レベルでリンケージして動態的变化がわかるよう、相関表形式で集計された農業構造動態統計が作成されている。この農業構造動態統計を作成するための集計データは、「構造動態マスタ」と呼ばれ、個票データとあわせて提供されているが、このデータを集計するための作成仕様が統計部において体系的に整備されていない。

農林業センサスの集計表を整理した一覧表・結果表に関する集計論理については、毎回センサス時に作成される『電算処理仕様書』において、その仕様が整備されているが、農業構造動態統計の集計のために作成される構造動態マスタの作成仕様については、電算処理仕様書で整備されるよう位置付けられていない。例えば、2005-2010 年の構造動態マスタについては、電算処理仕様書とは別に「編集方針」、「作成フロー」、「2005 年と 2010 年個票データの接続論理」について、断片的にドキュメントは整備されている（別添 2「2005-2010 年構造動態マスタ編集方針等」）ものの、体系的な整備とはなっていない。2000-2005 年の構造動態マスタについては、2000 年と 2005 年個票データの接続論理のみドキュメントが整備されており、1995-2000 年以前の構造動態マスタについては、一切のドキュメントが残されていない。このため、提供された個票データを用いて構造動態マスタを復元することができない状況にある。

また、2005 年および 2010 年の個票データの世帯集約に係る集計論理についても整備されていない。

2005 年および 2010 年センサスは、農業経営体単位の個票データとともに、1 世帯複数経営体について世帯（販売農家）単位に集約した個票データが提供されており、2000 年調査以前の販売農家の集計結果と接続できるようになっている。しかしながら、1 世帯複数経営体の結果を世帯に集約を行うための集計論理に関する資料が整備（提供）されていないため、世帯単位に集約した個票データの作成方法について、対外的に説明できない状況にある。

（３）リンクコードの整備

2010 年農林業センサスでは、調査客体候補名簿において農林業経営体以外の自給的農家・土地持ち非農家等の前回センサスとのリンクコードが整備されなかったため、これらの客体の動態がとらえられなくなった。

2010 年センサスの調査客体候補名簿の「NUMBER」欄には、2005 年センサスの基本指標番号である市区町村（3 桁）、旧市区町村（2 桁）、農業集落（3 桁）、調査区（2 桁）、経営体（4 桁）のコードをつないだ 14 桁のコードがリンクコードとして収録されている。しかしながら、当該欄は、2005 年センサスにおいて農林業経営体に該当しなかった客体については、表 4 のように経営体コード（下 4 桁）が「0000」となっており、リンクコードとして機能していない。このため、2005 年センサスにおいて農林業経営体に該当しなかった客体については、2005 年から 2010 年への動態変化がとらえられない状況となっている。

リンクコードについては、農林業経営体のみならず、名簿に記載されたすべての客体について、リンクコードを整備する必要がある。

（４）客体名簿の精査

客体名簿（調査客体候補名簿、照査表）は、内容が精査されていない。

2010 年センサスの場合、調査客体の判定が正しく行われているか、調査客体と判定された数と調査票の枚数が一致するかについて審査が行われているが、それ以外の項目の審査が行われていない。例えば、前回の基本指標番号が入力されている NUMBER 欄は、基本的に個票データごとにユニークな番号となると考えられるが、同一県内で重複している場合が 920 客体あった。

（５）調査項目の簡素化・廃止

統計調査をめぐる環境は、調査対象者の負担軽減から、調査項目が簡素化・廃止が行われる傾向にある。このため、例えば 2010 年農林業センサスでは、図 2 のように工芸農作物、野菜、果樹等の作付面積の把握内容が従来の品目別面積から類別の総面積に変更されており、時系列で分析ができなくなっている。2015 年センサスでは、従来に戻って品目別による面積の把握となったが、調査項目の利活用を精査し、簡素化・廃止は必要最小限に留めることが求められる。

2) 集落営農実態調査、集落営農活動実態調査

（１）リンクコードの整備

集落営農実態調査は、毎年すべての集落営農組織を把握しており、パネルデータによる利用が有効であると考えられるが、集落営農組織のパネル化を行うためのリンクコードが整備されていない。

集落営農実態調査において、各集落営農組織に通し番号が附されており、例えば各年次で同一組織が同一番号となっていればリンクコードとして機能する。しかしながら、表 5 のように、ある都道府県では、平成 23 年調査は都道府県単位で通し番号が附される一方、平成 24 年調査は地域センター単位で通し番号が附されていると考えられるなど、年次によって通し番号の付け方は異なっている。このため、同一組織であっても通し番号は年次間で一致せず、リンクコードとして機能しない。

通し番号がリンクコードとして機能するよう、付け方に基準を設けるなど、工夫する必要がある。

また、集落営農活動実態調査は、集落営農実態調査で把握した集落営農組織を母集団とする標本調査であるため、両調査のリンケージは容易に行うことができる。しかしながら、平成 19 年の集落営農活動実態調査において、一部の客体で指標コードの付け替えが行われており、集落営農実態調査とのリンケージが行えない。

集落営農活動実態調査の指標コードである「調査経営体番号」は集落営農実態調査の「通し番号」（大半の県において県一本の通し番号）を転写することとなっているが、一部の県の客体（2,590 客体中 253 客体）において、通し番号にない番号が調査経営体番号となっている場合や、通し番号と調査経営体番号は一致しているものの市町村番号が一致しない場合が見受けられた。

（２）標本の選定

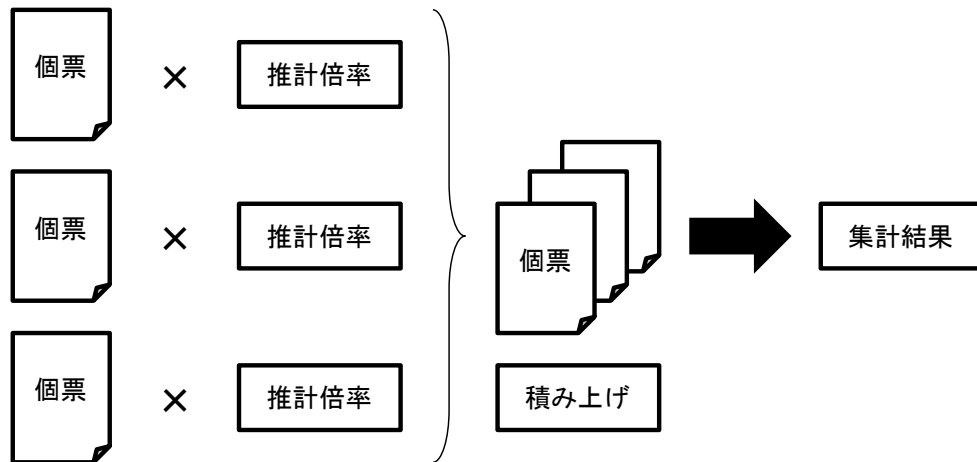
集落営農活動実態調査の標本は、固定標本ではなく、表６のように近年を除き毎年条件を変えて選定が行われ、調査が実施されている。このため、継続的に同一の調査客体が標本として選定される可能性は低く、結果的に一部の客体しかパネル化が行えない状況にある。標本の抽出条件を固定化し、一定年次の間隔を置いて、同一の調査客体を選定されるよう調査設計を行うことで、パネル化による動態分析が可能となる。

（３）データ入力のルール化

集落営農実態調査は、調査項目に集落営農の名称、代表者氏名、住所といった文字情報が含まれており、これらの調査結果のエクセル表への入力に際し、セル内で改行が行われている場合があった。このため、電算処理において支障をきたす場合があった。

文字情報は集計対象となっていないことから、調査要領において入力方法に関する規程はなされていないと考えられるが、高度利用に当たっては、文字情報も重要な情報となる場合があるので、入力ルールに対する指導を徹底する必要がある。

①個票データに推計倍率を乗じて、その結果を積み上げて集計を行う方法



②個票データの積み上げ結果に推計倍率を乗じて集計を行う方法

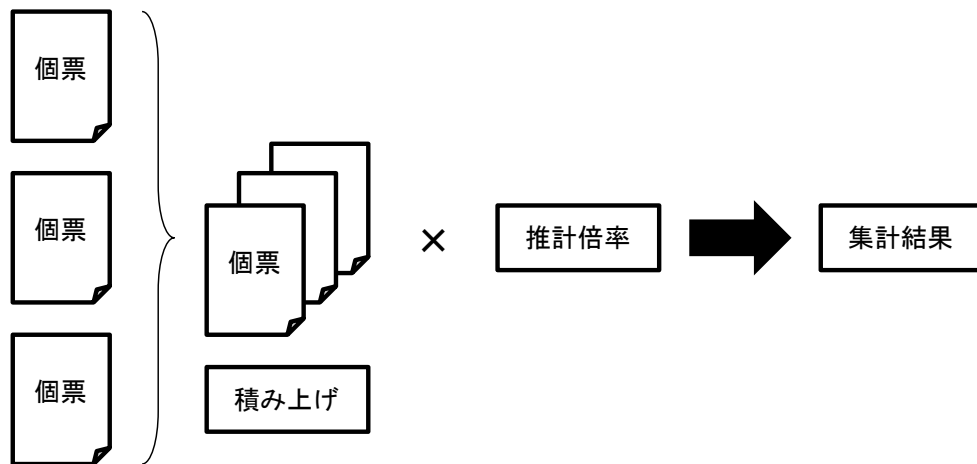


図 1 標本調査の集計方法

出所：筆者作成。

(2005 年)

(2010 年)

		露地 (ha) (町) (反) (畝)			施設 (a) m		
野	ト マ ト	433			434		
	な す	435			436		
	ピーマン	437			438		
	きゅうり	439			440		
	キャベツ	455			456		
	結球はくさい	457			458		
	レタス	461			462		
	ほうれんそう	463			464		
	ね ぎ	465			466		
	たまねぎ	467			468		

➡

		露地 (ha) (町) (反) (畝)			施設 (a) m		
野 菜 類	431				432		
花 き 類	481				482		
花 木	487				487		
その他の作物	489				490		

図 2 農林業センサス農林業経営体調査の調査項目変更の例

出所：2005 年及び 2010 年農林業センサス農林業経営体調査票より転載。

表 1 関係書類等の保存の例

関係書類	保存期間	保存責任者
集落営農実態調査 集落営農活動実態調査票	3 年	地域センター等の長
集落営農実態調査 集落営農活動実態調査票を収録した電磁的記録	永年	統計部長
集落営農活動実態調査の結果表を収録した電磁的記録	永年	統計部長
集落営農活動実態調査 調査標本リスト	5 年	統計部長
集落営農活動実態調査 実施結果報告表	1 年	統計部長

出所：農林水産省大臣官房統計部「集落営農実態調査要領」（平成 23 年 9 月）より転載。

表 2 1980 年センサスの符号表（一部抜粋）

政府統計コード		実施時期		作成日	5月21日	訂正日		8月7日	
統計調査名	1980年世界農林業センサス	集社区分		ファイル名	(照会先等)				
調査票名	農家調査票	(備考・補足事項)							
コード体系	京都大学入手版								
レコード集									

行番号	変数番号	項目名	階層	位置	項目番号	バイト数	繰返し	配置	型	小数点	種別	対象	符号	符号内容	備考
1	baa80_001	都道府県				2									
2	baa80_002	支庁				2									
3	baa80_003	市区町村				3									
4	baa80_004	旧市区町村				2									
5	baa80_005	農業集落				3									
6	baa80_006	全国農業地域コード				4									
7	baa80_007	農家				2									

出所：1980 年農林業センサス農家調査の調査票情報のデータレイアウトを基に、京都大学 農林水産統計デジタルアーカイブ講座において作成。

表 3 1975 年センサスの都道府県別レコード数

都道府県	データセット①	データセット②	差	都道府県	データセット①	データセット②	差
1	134,263	268,526	-134,263	25	82,723	82,723	0
2	109,872	219,744	-109,872	26	65,413	65,413	0
3	121,760	243,520	-121,760	27	56,119	56,119	0
4	117,263	234,526	-117,263	28	167,892	167,892	0
5	114,610	125,847	-11,237	29	52,555	52,555	0
6	106,353	106,353	0	30	61,504	61,504	0
7	153,017	153,017	0	31	53,582	53,582	0
8	181,074	181,074	0	32	79,275	79,275	0
9	108,789	108,789	0	33	142,400	142,400	0
10	109,456	109,456	0	34	138,409	138,409	0
11	133,656	133,656	0	35	96,006	96,006	0
12	146,552	146,552	0	36	65,065	65,065	0
13	31,019	31,019	0	37	74,577	74,577	0
14	51,661	51,661	0	38	105,727	105,727	0
15	178,981	178,981	0	39	58,608	58,608	0
16	74,761	74,761	0	40	138,992	138,992	0
17	66,481	66,481	0	41	67,684	67,684	0
18	56,950	56,950	0	42	83,689	83,689	0
19	68,479	68,479	0	43	135,489	75,352	60,137
20	200,114	200,114	0	44	101,071		101,071
21	121,520	121,520	0	45	90,205		90,205
22	140,127	140,127	0	46	195,084		195,084
23	157,713	157,713	0	47	48,018		48,018
24	108,513	108,513	0				

出所：筆者作成。

表 4 2010 年センサスの調査客体候補名簿データ（一部抜粋）

都道府県 CD	支庁CD	郡CD	市区町村 CD	旧市区町 村CD	農業集落 CD	調査区 CD	整理番号 CD1	整理番号 CD2	NUMBER	市区町村 名	旧市区町村 名	農業集落 名
1	1	0	101	1	1	1	1	0	10101001010000	中央区	札幌市6-1	幌南
1	1	0	101	1	1	1	2	0	10101001010000	中央区	札幌市6-1	幌南
1	1	0	101	1	1	1	3	0	10101001010000	中央区	札幌市6-1	幌南
1	1	0	101	1	1	1	4	0	10101001010000	中央区	札幌市6-1	幌南
1	1	0	101	1	1	1	5	0	10101001010000	中央区	札幌市6-1	幌南
1	1	0	101	1	1	1	6	0	10101001010000	中央区	札幌市6-1	幌南
1	1	0	101	1	1	1	7	0	10101001010000	中央区	札幌市6-1	幌南
1	1	0	101	1	1	1	8	0	10101001010010	中央区	札幌市6-1	幌南

出所：2010 年農林業センサス農林業経営体調査の調査票情報に基づき、筆者作成。

表 5 集落営農実態調査における通し番号の付け方の例

(平成 23 年)

(平成 24 年)

23_002通 し番号	23_003集落営農の名称	23_009市 町村コー ド	24_002通 し番号	24_003集落営農の名称	24_009市 町村コー ド
1	豊畑地区土地利用組合	100	1	豊畑地区土地利用組合	100
2	日の出生産組合	204	1	農事組合法人日の出生産組合	204
3	香川町牧野組合	205	1	香川町牧野組合	205
4	相内麦作生産組合	208	1	相内麦作生産組合	208
5	端野町コンバイン組合(麦)	208	2	豊田地区農用地管理組合	208
6	端野町水稲コンバイン組合	208	3	柏木地区農用地管理組合	208
7	温根湯麦類生産組合	208	4	端野町コンバイン組合(麦)	208
8	東浜麦作集団	208	5	端野町水稲コンバイン組合	208
9	日吉麦作集団	208	6	温根湯麦類生産組合	208
10	豊福第1麦作集団	208	2	(農)岐阜ライスセンター利用組合	210
11	共立川西コンバイン利用組合	208	3	(有)岐阜コントラクター	210
12	(農)岐阜ライスセンター利用組合	210	2	(有)ドリームアグリ	214
13	(有)岐阜コントラクター	210	4	一心米麦営農組合	215
14	(有)ドリームアグリ	214	5	沼の内営農組合	215
15	一心米麦営農組合	215	6	上美唄米麦営農組合	215
16	沼の内営農組合	215	7	元村米麦営農組合	215
17	上美唄米麦営農組合	215	8	中村共同施設利用組合	215
18	元村米麦営農組合	215	9	峰樺3号機械利用組合	215
19	中村共同施設利用組合	215	10	開発みどりふあーむ	215
20	峰樺3号機械利用組合	215	7	なぎさ飼料生産組合	219
21	開発みどりふあーむ	215	8	立牛自走ハーベスター利用組合	219
22	アグリ山形合同会社	215	9	上渚滑自走ハーベスター利用組合	219
23	なぎさ飼料生産組合	219	10	上東自走ハーベスター利用組合	219
24	立牛自走ハーベスター利用組合	219	11	中渚滑自走ハーベスター利用組合	219
25	上渚滑自走ハーベスター利用組合	219	12	下渚滑自走ハーベスター利用組合	219
26	上東自走ハーベスター利用組合	219	13	藻別自走ハーベスター利用組合	219
27	中渚滑自走ハーベスター利用組合	219	14	新中央自走ハーベスター利用組合	219
28	下渚滑自走ハーベスター利用組合	219	15	紋別フィードアンドクローブ利用組合	219
29	藻別自走ハーベスター利用組合	219	3	中士別第1トラクター利用組合	220
30	新中央自走ハーベスター利用組合	219	4	中士別第1営農組合	220
			5	中士別第3営農組合	220
			6	中士別第4営農組合	220

出所：平成 23 年及び 24 年集落営農実態調査の調査票情報に基づき、筆者作成。

表 6 集落営農活動実態調査の調査客体の抽出条件

年次	調査対象の条件
平成 19 年	集落営農の定義に該当する集落営農
平成 20 年	集落内の営農を一括管理・運営している集落営農
平成 21 年	水田・畑作経営所得安定対策に加入している集落営農
平成 22 年	集落内の営農を一括管理・運営している集落営農
平成 23、24、25 年	農産物の生産・販売を行っている集落営農

出所：集落営農活動実態調査要領及び集落営農実態調査要領より転載。