

RAGGIUNGIBILITÀ DELLE AZIENDE AGRICOLE E DIMENSIONE AMBIENTALE E
TERRITORIALE: EVIDENZE EMPIRICHE DALLE INDAGINI PILOTA PER IL
CENSIMENTO 2020

Francesco BORRELLI¹, Luana DE FELICI², Roberto GISMONDI³, Chiara GNESI⁴

SOMMARIO

In vista del 7° Censimento Generale dell'Agricoltura, l'Istituto Nazionale di Statistica (ISTAT) ha condotto, nei mesi di novembre e dicembre 2018, tre indagini sperimentali con l'obiettivo di testare la complessità del questionario in relazione alle tecniche di raccolta dati che si intendono utilizzare nel prossimo Censimento, come alternativa al metodo PAPI (*Paper And Pencil Interviewing*) impiegato nel Censimento precedente.

L'obiettivo del presente lavoro è quello di analizzare i risultati delle indagini pilota, in termini di raggiungibilità delle aziende in funzione di variabili, non solo relative alla *governance* e alla struttura dell'azienda, ma anche legate alle caratteristiche dell'ambiente e del territorio in cui l'azienda opera. A tale scopo, l'analisi si compone di due parti. Nella prima si procede a un'analisi descrittiva dei risultati dell'indagine in funzione delle principali caratteristiche individuali, aziendali e territoriali, evidenziando anche l'impatto di ciascuna tecnica di raccolta dati. Successivamente, si è applicato un modello logistico per individuare le variabili più significative nel determinare la raggiungibilità delle aziende in termini di tassi di risposta e di esiti delle rilevazioni.

¹ Istituto Nazionale di Statistica, Viale della Civiltà del Lavoro 50, 00144, Roma, e-mail: borrelli@istat.it

² Istituto Nazionale di Statistica, Viale della Civiltà del Lavoro 50, 00144, Roma, e-mail: defelici@istat.it

³ Istituto Nazionale di Statistica, Viale della Civiltà del Lavoro 50, 00144, Roma, e-mail: gismondi@istat.it

⁴ Istituto Nazionale di Statistica, Viale della Civiltà del Lavoro 50, 00144, Roma, e-mail: gnesi@istat.it

1. Introduzione⁵

L' Istat è attualmente impegnata nell'organizzazione del prossimo Censimento dell'agricoltura 2020 con l'obiettivo di introdurre elementi di innovazione rispetto alle passate tornate censuarie. Tre sono i fattori che stanno guidando le principali scelte strategiche che sta effettuando l'Istituto:

- a) maggiore sostenibilità del modello in ottica di censimento permanente;
- b) contenimento dei costi;
- c) salvaguardia della qualità dei dati censuari.

Quello del 2020, infatti, sarà l'ultimo Censimento "tradizionale". Successivamente all'operazione censuaria, si lascerà il campo alla stagione dei censimenti permanenti, basati sui registri statistici e sull'integrazione tra fonti amministrative e indagini statistiche.

Nel 7° Censimento Generale dell'Agricoltura, che si svolgerà dal 5 ottobre 2020 al 31 marzo 2021, si adotterà un unico modello di questionario, indipendentemente dalla dimensione aziendale, per tutte le aziende che superano una prefissata soglia fisica⁶. La raccolta dei dati avverrà utilizzando una strategia di rilevazione multi-tecnica, che prevede la totale eliminazione dei supporti cartacei utilizzati nella precedente edizione censuaria. La stima delle aziende per ciascun canale di rilevazione è stata effettuata sulla base dei risultati delle indagini pilota che si sono svolte nel periodo novembre - dicembre 2018, al fine di testare un approccio multi-tecnica di rilevazione da adottare nel prossimo censimento. Le sperimentazioni hanno riguardato le tecniche di raccolta dati Computer Assisted Web Interviewing (CAWI), Computer Assisted Telephone Interviewing (CATI) e Computer Assisted Personal Interviewing (CAPI); quest'ultima indagine è stata svolta mediante intervista diretta dai tre maggiori Centri di Assistenza Agricola (CAA), Cia, Coldiretti e Confagricoltura. Sulla base dei risultati raggiunti si è stimato che circa il 15% di aziende autocompilerà il questionario via web, il 22% utilizzerà il canale telefonico (nella duplice forma *inbound* e *outbound*⁷) e il restante 63% sarà intervistato direttamente da un CAA.

Nel presente lavoro si sono analizzati i risultati ottenuti dalle indagini pilota con l'obiettivo di individuare i fattori più significativi nel determinare il raggiungimento delle aziende agricole. A tal fine si sono considerate, oltre alle variabili relative alla *governance* e alla struttura dell'azienda, anche quelle legate alle caratteristiche dell'ambiente e del territorio in cui è ubicato il centro aziendale.

Dopo una descrizione delle indagini pilota – sotto il profilo organizzativo e di strategia campionaria (paragrafo 3) e dei principali risultati ottenuti (paragrafo 4), si è svolta una breve trattazione metodologica (paragrafo 5) e una descrizione dell'applicazione del modello lineare ai dati delle indagini pilota (paragrafo 6). Successivamente sono riportati i risultati delle analisi, descrittiva ed econometrica, svolte (paragrafo 7). Il lavoro si conclude con una simulazione dei possibili risultati del prossimo Censimento, attraverso l'applicazione delle stime ottenute dal modello alla lista di aziende agricole, costruita con una metodologia simile a quella censuaria. Infine, alcune considerazioni finali sui principali risultati ottenuti e alcune prospettive di lavoro futuro (paragrafo 8).

2. Le indagini pilota

I dati rilevati dalle indagini pilota hanno permesso di calcolare, per ogni canale di raccolta, i tassi di risposta e qualità dei dati, oltre ad eventuali problematiche di carattere tematico, organizzativo o tecnologico. La sperimentazione CATI, in particolare, ha avuto l'obiettivo di testare la tecnica di raccolta dati telefonica,

⁵ Ai fini dell'attribuzione delle singole parti del lavoro si fa presente che: il sommario e il paragrafo 7 sono interamente redatti da R. Gismondi, i paragrafi 2, 5, 6.2 da C. Gnesi, i paragrafi 3 e 6.1 da L. De Felici, i paragrafi 1, 4 e 6.3 sono redatti da F. Borrelli.

⁶ La soglia prevista comprende tutte le aziende che possiedono almeno 20 are di Superficie Agricola Utilizzata (SAU) oppure 10 are di vite o almeno 1 Unità di Bestiame Adulto (UBA).

⁷ Si prevede di attivare, per tutto il periodo della rilevazione, un canale *inbound*, attraverso cui le aziende potranno partecipare alla rilevazione chiamando direttamente il Contact Center per procedere alla compilazione del questionario in modalità assistita da operatore, affiancato, nella fase successiva, da canale *outbound*, che rappresenta la classica modalità di somministrazione del questionario attraverso il canale telefonico.

anche in relazione alla bontà dei contatti presenti nella lista oltre che alle strategie utilizzate per la gestione dei contatti telefonici.

La numerosità campionaria è stata di oltre 30.000 aziende agricole (19.816 aziende agricole per la CAWI, 10.986 per la CATI⁸ e 2.016 per la CAPI), estratte secondo un disegno di campionamento stratificato dal Registro statistico delle Aziende Agricole (FR) 2015, composto da 1.659.928 aziende agricole gestito dall'Istat. Per la CAPI, il Registro è stato integrato con un database fornito da Agea e relativo alle 959.772 aziende che hanno aperto un fascicolo aziendale nel 2015, trattandosi di una indagine progettata con l'obiettivo di valutare l'efficienza organizzativa della rete dei CAA anche alla luce delle informazioni sulle aziende già in loro possesso.

La sperimentazione della rete CAA è stata svolta dai CAA delle 3 organizzazioni professionali agricole Coldiretti, Confagricoltura e CIA. Data l'esigua numerosità campionaria e la breve fase di rilevazione (dal 3 al 14 dicembre 2018), oltre che la natura volontaria della partecipazione dei CAA, si è deciso di effettuare la sperimentazione concentrandosi su 4 province. Per la selezione delle 4 province sono stati adottati i seguenti vincoli:

- area geografica: Nord, Centro, Sud, Isole;
- presenza di almeno 12 sedi dei tre CAA principali nella Provincia.

Pertanto, le province scelte sono state: Bologna, Viterbo, Bari e Catania.

Nelle 4 Province selezionate non sono stati coinvolti tutti gli uffici CAA delle 3 organizzazioni professionali ma solo alcuni di essi (12 uffici) poiché tra gli obiettivi della sperimentazione c'era quello di verificare la sostenibilità dei carichi di lavoro per singolo operatore e la tenuta organizzativa degli uffici a fronte di un carico di lavoro aggiuntivo derivante dall'impegno della raccolta dati. Pertanto, ad ogni ufficio sono state assegnate 42 aziende da intervistare nell'arco di due settimane, ovvero 4 interviste al giorno, che è indicativamente il carico di lavoro minimo che l'Istat stima per ogni rilevatore per il censimento a regime.

L'universo di riferimento era costituito dalle aziende presenti nel Farm Register (FR) 2015 senza fascicolo e da quelle inserite nel database fornito da Agea relativo al 2017. Il disegno campionario ha previsto una doppia stratificazione: tipologia di legame CAA-CUAA (azienda agricola) e soglie dimensionali delle aziende. Per la tipologia di legame CAA-CUAA sono stati considerati 3 strati, che sulla base delle informazioni disponibili hanno identificato 3 profili diversi di azienda agricola:

1. aziende che hanno aggiornato il fascicolo 2017 presso i CAA delle 3 organizzazioni professionali agricole;
2. aziende che hanno aggiornato il fascicolo 2017 presso CAA diversi da quelli del punto 1;
3. aziende del FR 2015 che non hanno fascicolo.

Per le soglie dimensionali si considerano due strati, aziende sopra e sotto soglia definita nel Regolamento UE, stabilita rispetto al raggiungimento del 98% di superficie in termini di Superficie Agricola Totale (SAT). Per ogni provincia sono stati previsti circa 500 contatti, l'allocazione delle unità del campione tra gli strati è stata fatta tenendo conto di due fattori:

- il peso demografico delle aziende agricole dello strato sul totale delle aziende agricole;
- un indicatore "di rischio" associabile allo strato, che si reputa essere crescente dallo strato 1 al 3.

L'allocazione delle unità è stata quindi effettuata con il criterio di allocazione ottimale di Neyman (in cui alle deviazioni standard sono stati sostituiti i livelli di rischio, assimilabili a diversi gradi di incertezza attribuibili all'esito delle interviste al variare dello strato). Le aziende dello strato 2 e 3 sono state assegnate casualmente agli uffici CAA all'interno della Provincia dove ha sede l'ufficio stesso. Le indagini pilota CATI e CAWI si sono concluse il 15 dicembre dopo un mese di raccolta dati. La CATI si è svolta in due fasi: dopo le prime due settimane di rilevazione esclusivamente *inbound* (in cui si è data la possibilità alle

⁸ Nel capitolato di gara era previsto un numero di interviste pari a 2.200 aziende. Tuttavia, alla ditta che ha condotto l'indagine è stato fornito un campione esteso di circa 11.000 unità in modo da avere un serbatoio di aziende da cui attingere per arrivare alla numerosità campionaria prefissata.

aziende di contattare il *contact center* ed effettuare l'intervista), è stata affiancata la fase *outbound* nella quale le aziende presenti nel campione sono state contattate da rilevatori formati ad hoc per la raccolta dati. Per la CAWI, solamente alle aziende i cui questionari risultavano in lavorazione alla data di chiusura dell'indagine è stata data la possibilità di terminare l'intervista successivamente, in ogni caso entro e non oltre il 21 gennaio 2019. Il disegno di campionamento delle due indagini è stato di tipo probabilistico a uno stadio stratificato. Le variabili di stratificazione hanno fatto riferimento alla dimensione territoriale e alla dimensione economica. Per la prima si è considerata l'ubicazione del centro aziendale, in termini di distribuzione regionale. Per la seconda, è stata utilizzata una soglia del 98% di SAU/UBA stabilita nei Regolamenti europei.

3. I principali risultati delle indagini pilota

Le indagini sperimentali hanno avuto un esito, in media, per il 38,3 per cento delle aziende agricole del campione.

Il risultato complessivo è stato fortemente influenzato dalla tecnica CAWI, che ha avuto un tasso di risposta pari all'11,2 per cento. L'indagine CAPI, svolta direttamente dai CAA, ha raggiunto, invece, oltre il 98 per cento delle aziende nelle due settimane previste per la rilevazione. La pilota CATI ha contattato il 77 per cento delle aziende presenti nel campione per raggiungere la quota di 2.200 interviste prevista dal capitolato di gara.

L'incidenza delle aziende attive, pari al 43,6 per cento, è fortemente differenziata tra le tre sperimentazioni (tavola 1). Infatti, mentre nella CAPI tale percentuale si attesta sul 60 per cento, per la CAWI essa è molto maggiore, pari ad oltre il 90 per cento, a dimostrazione della propensione ad autocompilare il questionario via web da parte delle aziende che si riconoscono come agricole e/o zootecniche. Il risultato dell'indagine CATI (26,6 per cento) sottostima il tasso di attività a causa del raggiungimento della quota di interviste prevista dal capitolato. Nel complesso, dunque, le aziende raggiunte dalle pilota sono pari a circa il 48 per cento, considerando anche le aziende che hanno rifiutato e quelle che, nel caso del CAWI, non disponevano di un accesso Web.

Tavola 1. Questionari inviati per esito, tecnica di rilevazione e soglia proposta per Censimento 2020 – valori assoluti e percentuali*

CAA	Totale aziende		Sotto soglia		Sopra soglia	
Inviati	1.984	100,0	58	100,0	1.926	100,0
Azienda attiva	1.171	59,0	16	27,6	1.155	60,0
Azienda non più esistente, doppiata o fuori campo di osservazione	168	8,5	13	22,4	155	8,0
Azienda temporaneamente inattiva	26	1,3	1	1,7	25	1,3
Non rilevata	619	31,2	28	48,3	591	30,7
anagrafica errata	4	0,2	0	0,0	4	0,2
irreperibile	487	24,5	26	44,8	461	23,9
inesitato	0	0,0	0	0,0	0	0,0
mancata disponibilità web	15	0,8	0	0,0	15	0,8
rifiuto	113	5,7	2	3,4	111	5,8
CATI	Totale aziende		Sotto soglia		Sopra soglia	
Inviati	8.268	100,0	555	100,0	7.713	100,0
Azienda attiva	2.200	26,6	141	25,4	2.059	26,7
Azienda non più esistente, doppiata o fuori campo di osservazione	772	9,3	115	20,7	657	8,5
Azienda temporaneamente inattiva	94	1,1	16	2,9	78	1,0
Non rilevata	5.202	62,9	283	51,0	4.919	63,8
anagrafica errata	1.668	20,2	118	21,3	1.550	20,1
irreperibile	3.210	38,8	136	24,5	3.074	39,9
inesitato	10	0,1	2	0,4	8	0,1
mancata disponibilità	0	0,0	0	0,0	0	0,0

rifiuto	314	3,8	27	4,9	287	3,7
CAWI	<i>Totale aziende</i>		<i>Sotto soglia</i>		<i>Sopra soglia</i>	
Inviati	2.217	100,0	216	100,0	2.001	100,0
Azienda attiva	2.063	93,1	183	84,7	1.880	94,0
Azienda non più esistente, doppione o fuori campo di osservazione	84	3,8	26	12,0	58	2,9
Azienda temporaneamente inattiva	0	0,0	0	0,0	0	0,0
Non rilevata	70	3,2	7	3,2	63	3,1
anagrafica errata	9	0,4	3	1,4	6	0,3
irreperibile	0	0,0	0	0,0	0	0,0
inesitato	0	0,0	0	0,0	0	0,0
mancata disponibilità web	57	2,6	4	1,9	53	2,6
rifiuto	4	0,2	0	0,0	4	0,2
TOTALE	<i>Totale aziende</i>		<i>Sotto soglia</i>		<i>Sopra soglia</i>	
Inviati	12.469	100,0	829	100,0	11.640	100,0
Azienda attiva	5.434	43,6	340	41,0	5.094	43,8
Azienda non più esistente, doppione o fuori campo di osservazione	1.024	8,2	154	18,6	870	7,5
Azienda temporaneamente inattiva	120	1,0	17	2,1	103	0,9
Non rilevata	5.891	47,2	318	38,4	5.573	47,9
anagrafica errata	1.681	13,5	121	14,6	1.560	13,4
irreperibile	3.697	29,6	162	19,5	3.535	30,4
inesitato	10	0,1	2	0,2	8	0,1
mancata disponibilità web	72	0,6	4	0,5	68	0,6
rifiuto	431	3,5	29	3,5	402	3,5

*Sono considerate sotto la soglia le aziende con SAU<0,20 are e Vite < 0,10 are e Uba <1

Fonte: indagini pilota del Censimento Agricoltura 2018

L'insieme delle aziende non agricole (non esistenti, fuori campo di osservazione, doppione o temporaneamente inattive) è pari al 9,2 per cento. Tale quota è molto simile sia che si consideri la CATI (in media pari al 10,1 per cento) sia la CAPI (9,8 per cento). Nel caso della CAWI tale incidenza, pari al 3,8 per cento, presenta una distorsione derivante dall'assenza di informazioni dovuta alla mancata risposta totale.

Il tasso di rifiuti, pari al 3,5 per cento, è in linea con quanto risulta dalle altre indagini correnti agricole. Sotto questo aspetto il risultato della CAWI (0,2 per cento), molto più basso rispetto a quello della CATI (3,8 per cento) e della CAPI (5,7 per cento), è distorto dalla tecnica di rilevazione utilizzata.

Poco meno della metà delle aziende del campione (47,2) non sono state rilevate nelle sperimentazioni. Anche sotto questo profilo la tecnica di raccolta dati ha prodotto risultati molto differenziati. Mentre nella CAPI la quota di non rilevate è pari a circa un terzo del campione (31,2), essa raddoppia nella CATI (63,2) mentre con la CAWI è solo del 3,2 per cento, in quanto oltre il 90 per cento delle aziende che hanno autocompilato via web si sono dichiarate attive.

La principale causa di non rilevazione risiede nell'irreperibilità del conduttore (29,6 per cento) riconducibile, tuttavia, anche a lacune informative all'interno delle fonti amministrative e statistiche utilizzate per i dati di contatto. Nel caso della CAPI l'incidenza dell'irreperibilità (24,5 per cento) è inferiore rispetto a quanto riscontrato nella CATI (38,8 per cento) in quanto i Centri di Assistenza Agricola (CAA) hanno avuto un ruolo attivo nell'utilizzare le proprie reti, formali ed informali, per rintracciare le aziende del campione. I tassi di risposta sono molto diversificati tra le tre tecniche: nel caso della CAPI è stato inviato il 98,4 per cento dei questionari, il tasso di risposta della CAWI si è attestato all' 11,2 per cento, mentre per la CATI il numero di aziende da intervistare è stato fissato da capitolato (2.200 aziende rilevate).

Riguardo ai principali esiti:

- l'incidenza delle aziende inattive, non esistenti e fuori dal campo di osservazione è circa il 10 per cento;

- le aziende che si *presumono attive* (considerando sia quelle che hanno dichiarato di essere attive nel questionario, sia quelle che hanno rifiutato o hanno manifestato l'impossibilità di rispondere tramite web) sono state il 48 per cento; le aziende raggiungibili sono circa il 56 %, in questo caso sono state considerate come raggiungibili, oltre alle aziende che si *presumono attive*, anche quelle inattive e fuori dal campo di osservazione dell'indagine. La principale causa di non rilevazione è l'irreperibilità del conduttore, che ha riguardato circa un terzo del campione totale.

4. Metodologia

I modelli di regressione logistica si usano quando si vuole modellare la probabilità di un evento in relazione ad una serie di variabili esplicative. In questi modelli la variabile dipendente definisce l'appartenenza a uno dei due gruppi in cui si esplica il fenomeno oggetto di studio⁹ e, pertanto, ha carattere qualitativo. Le variabili indipendenti possono essere sia di tipo quantitativo che qualitativo.

Se Y è una variabile binaria si può ipotizzare che segua una distribuzione bernoulliana (o binomiale con $n=1$) con media pari alla probabilità di successo relativo al verificarsi dell'evento:

$$Y_i \sim \text{Bernoulli}(\mu_i) \quad [1]$$

con $\mu_i = \Pr(Y_i=1)$

Il modello non può essere definito con l'equazione di regressione lineare in quanto il valore della Y dato il set di predittori, cioè $E(Y|X)$, non può assumere un qualsiasi valore da meno infinito a più infinito, bensì deve necessariamente essere pari ad un valore compreso nell'intervallo $[0,1]$, trattandosi di una probabilità.

Il modello logistico, nella sua formulazione lineare, ha la forma:

$$\text{logit}(p) = \log(p/(1-p)) = \alpha + \beta X \quad [2]$$

Esso modella la probabilità p che la variabile dipendente Y assuma, ad esempio valore 1, in presenza di un dato vettore X di variabili esplicative. Si avrà quindi come per il modello lineare classico una stima per l'intercetta α e per il vettore dei parametri β .

Utilizzando la funzione logistica, si dimostra che il logaritmo del rapporto tra la probabilità di successo e la probabilità di insuccesso, denominato *logit*, è una funzione *link* che lega la variabile di risposta alla combinazione lineare di predittori (per $i=1,2,\dots,n$) in maniera tale da assicurare che, per ogni valore assunto dalle variabili esplicative, la risposta Y sia sempre compresa nell'intervallo $[0,1]$.

Il parametro α è il valore del logit (p) che si ottiene per $X=0$. Il parametro β esprime la variazione prodotta sul logit (p) dall'incremento di un'unità di X .

Il modello precedente è lineare su scala logit, ma ovviamente non è lineare sulla scala delle probabilità:

$$p_i = e^{\alpha + \beta X} / (1 + e^{\alpha + \beta X}) \quad [3]$$

Il coefficiente α è quindi il logit della probabilità del gruppo in cui x vale 0:

$$\lambda_0 = \ln(p_0/(1-p_0)) \text{ e } p_0 = e^{\alpha} / (1 + e^{\alpha}) \quad [4]$$

Nel gruppo in cui x vale 1, avremo:

$$\lambda_1 = \ln(p_1/(1-p_1)) \text{ e } p_1 = e^{\alpha + \beta X} / (1 + e^{\alpha + \beta X}) \quad [5]$$

La differenza tra λ_0 e λ_1 è rappresentata da $\beta_1 = \lambda_1 - \lambda_0$:

$$\beta_1 = \ln[(p_1/(1-p_1)) / (p_0/(1-p_0))] \quad [6]$$

La quantità racchiusa tra le parentesi quadre è il rapporto tra gli *odds* delle probabilità nei due gruppi; questo rapporto prende il nome di *odds ratio* ed è una misura di associazione: vale 1 se e soltanto se $p_0 = p_1$. Il coefficiente β rappresenta quindi il logaritmo naturale dell'*odds ratio* e varrà 0 se e soltanto se $p_0 = p_1$. Nel modello logistico i confronti tra probabilità sono fatti eseguendo confronti tra logaritmi di *odds ratio*. Gli *odds* aumentano in maniera moltiplicativa di $\exp(\beta)$ per un cambiamento unitario della variabile esplicativa X . I due parametri α e β vengono stimati con il metodo di massima verosimiglianza, che consiste nello

⁹ La variabile dipendente (Y) descrive l'esito riguardante il verificarsi di un evento aleatorio. Pertanto assume solo 2 valori: evento presente ($y=1$) e evento assente ($y=0$).

scegliere quei valori di p_0 e p_1 per cui la funzione di verosimiglianza raggiunge il massimo, nell'ipotesi che tali valori esistano e siano unici.

La selezione delle variabili più significative può essere effettuata seguendo una molteplicità di approcci, tra cui i più ricorrenti sono i metodi *backward*¹⁰, *forward*¹¹, *stepwise*¹².

5. Applicazione

L'obiettivo dell'analisi svolta nel presente lavoro è quello di individuare le determinanti della probabilità di azienda agricola di essere *raggiunta* attraverso una delle tre tecniche di raccolta dati utilizzate nel prossimo Censimento dell'Agricoltura.

L'analisi è stata svolta in due fasi: nella prima si è svolta un'analisi descrittiva dei risultati delle indagini pilota in funzione delle principali caratteristiche individuali, strutturali e territoriali delle aziende agricole, evidenziando anche l'impatto di ciascuna tecnica di raccolta dati. Nella seconda fase si è utilizzato un modello logistico per individuare le variabili più significative nel determinare la raggiungibilità delle aziende in termini di tassi di risposta e di esiti delle rilevazioni. I risultati ottenuti sono stati anche utilizzati per effettuare una simulazione dei possibili risultati del prossimo Censimento, attraverso l'applicazione delle stime ottenute dal modello alla lista di aziende agricole, costruita con una metodologia simile a quella che si sta definendo per l'indagine censuaria.

Per ciascuna fase si sono utilizzati i dati relativi alle indagini pilota del Censimento Agricoltura, oltre ai dati presenti all'interno del Registro Statistico delle Aziende Agricole detenuto dall'Istat, relativo all'anno 2016.

La *variabile dipendente* è la *raggiungibilità* dell'azienda, che assume valore pari a 1, quando nelle indagini pilota il contatto con l'azienda ha dato esito certo, anche se non necessariamente sia stato compilato il questionario (come nel caso del rifiuto esplicito o di azienda non più attiva o fuori campo di osservazione). Pertanto non sono state considerate raggiunte le aziende irreperibili o con errata anagrafica. In questo modo, l'insieme delle unità è suddiviso in due gruppi: aziende agricole non raggiunte ($Y=0$) e aziende agricole raggiunte ($Y=1$).

Le variabili esplicative inserite nel modello fanno riferimento a tre dimensioni, individuale, strutturale e territoriale, che afferiscono all'azienda agricola.

Per la prima dimensione, quella *Individuale*, sono state considerate l'età e il genere del conduttore.

Per la componente *Strutturale* è stata considerata la Superficie Agricola Totale (SAT), come *proxy* della dimensione dell'azienda, la consistenza del bestiame, per caratterizzare il settore della zootecnia e la componente indipendente della manodopera aziendale. Inoltre, le aziende sono state caratterizzate anche attraverso la presenza di segnali di attività nelle principali fonti amministrative e statistiche che confluiscono nel Farm Register (tra queste, Agea, Catasto, Anagrafe Zootecnica, Camere di Commercio, Censimento dell'Agricoltura 2010).

La componente *Territoriale* è rappresentata dalle principali caratteristiche del territorio in cui è ubicato il centro aziendale con riferimento alla zona altimetrica, ai caratteri di montanità e litoraneità, al grado di urbanizzazione e alla ripartizione geografica.

Infine, anche la tecnica di raccolta dei dati nelle indagini pilota è stata inclusa tra le variabili indipendenti, essendo intrinsecamente correlata agli esiti raggiunti dalle sperimentazioni.

¹⁰ L'eliminazione *backward* inizia stimando il modello con l'insieme di variabili esplicative più vasto ed eliminando le variabili che risultano meno significative; la procedura viene iterata fino a che nessuna variabile possa essere più eliminata.

¹¹ Il metodo *forward* inizia con un modello con solo l'intercetta e aggiunge una ad una le variabili risultate significative con Likelihood Ratio o Statistica di Wald; la procedura termina quando nessuna delle variabili ancora da includere risulta significativa secondo i criteri precedenti.

¹² Il metodo *stepwise* è un metodo intermedio tra i due precedenti: è articolato in più stadi, ciascuno dei quali consiste nell'utilizzare una selezione *forward* per decidere quale variabile includere e un'eliminazione *backward* per decidere quale eliminare.

Nonostante inizialmente si sia costruito il modello inserendo tutte le variabili esplicative sopra esposte (modello *pieno*), si è poi utilizzato un modello *ridotto*, comprendente le sole variabili indipendenti significative. In effetti, essendo il modello *ridotto* di tipo *nested*, è stato possibile effettuare un'analisi della devianza con un test del Chi-quadro da cui la differenza in termini di capacità predittiva dei due modelli non è risultata apprezzabile. Inoltre, anche il confronto tra gli AIC, ha confermato la stessa evidenza. Pertanto, in virtù del principio di parsimonia, si è utilizzato il modello ridotto.

6. Risultati

6.1 Analisi descrittiva

L'analisi esplorativa di tipo descrittivo ha rivelato che circa il 23% di aziende del campione sono state raggiunte dalle indagini pilota, secondo la definizione di *raggiungibilità* utilizzata nel presente lavoro (Tavola 2). La tecnica CAPI è risultata la più efficace in termini di raggiungibilità delle aziende agricole, seguita dalla CATI che ha raggiunto circa il 40 per cento di aziende, mentre quella CAWI appena l'11 per cento delle aziende presenti nel campione. Si ricorda, tuttavia, che il risultato della CATI è sottostimato, essendo stato fissato *ex-ante* un numero di questionari completi - pari a 2.200 - oltre il quale la ditta non ha più contattato altre aziende agricole. Ciò che emerge con chiarezza è che la presenza di un rilevatore influenza in maniera significativa la raggiungibilità dell'azienda, sia mediante intervista diretta sia attraverso contatto telefonico.

Tavola 2 – Distribuzione delle aziende raggiunte per esito e tecnica di rilevazione– valori assoluti e percentuali

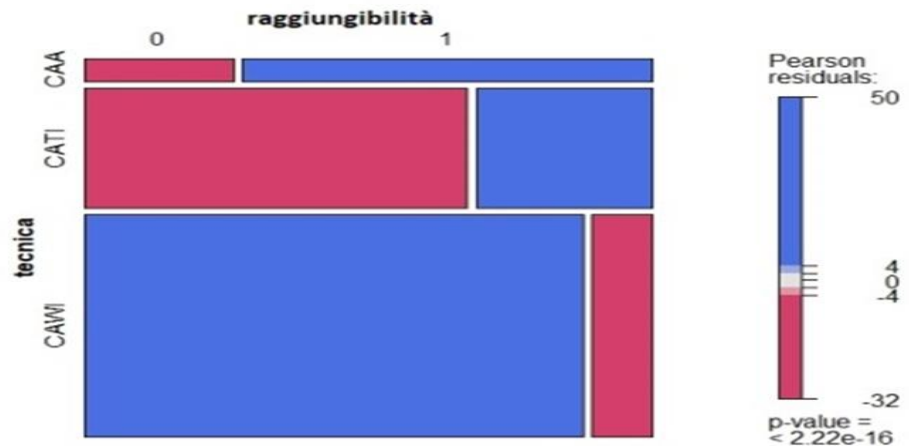
	CAPI		CATI		CAWI		TOTALE	
Totale aziende	2.016	100,0	8.268	100,0	19.816	100,0	30.100	100,0
Totale aziende raggiunte	1.478	73,3	3.380	40,9	2.151	10,9	7.009	23,3
Azienda attiva	1.171	58,1	2.200	26,6	2.063	10,4	5.434	18,1
Azienda non più esistente, doppiata o fuori campo di osservazione	168	8,3	772	9,3	84	0,4	1.024	3,4
Azienda temporaneamente inattiva	26	1,3	94	1,1	0	0,0	120	0,4
rifiuto	113	5,6	314	3,8	4	0,0	431	1,4

Fonte: indagini pilota del Censimento Agricoltura 2020

Utilizzando il grafico a mosaico¹³, particolarmente adatto ad analizzare l'associazione tra variabili qualitative, si evince che la raggiungibilità dell'azienda agricola è strettamente dipendente dalla tecnica di raccolta dati utilizzata. La tecnica CAPI è la più adatta a raggiungere le aziende agricole, come dimostra la distribuzione condizionata associata alle aziende raggiunte dall'indagine svolta dai CAA. Al contrario, nella CAWI la frequenza delle aziende non raggiunte è molto al di sopra di quanto ci si sarebbe aspettati sotto l'ipotesi di indipendenza tra le due variabili (Figura 1).

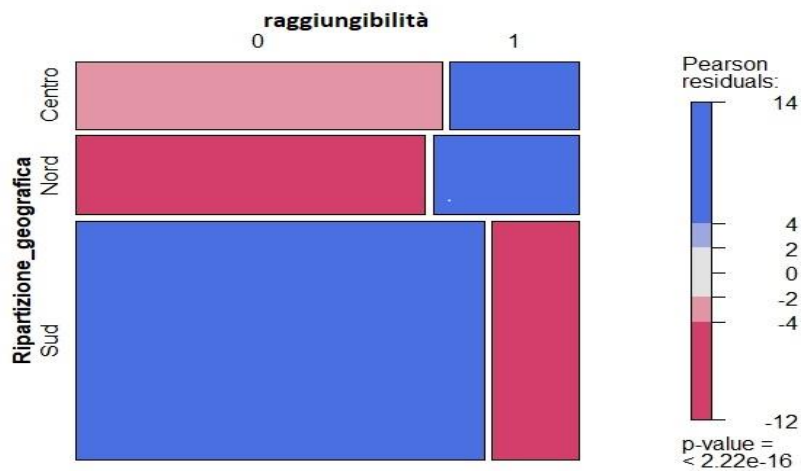
¹³ Nel grafico a mosaico l'altezza di ogni rettangolo rappresenta la frequenza della modalità della variabile indipendente, mentre l'area del rettangolo indica la frequenza dell'incrocio tra le modalità della dipendente e indipendente. Il colore rappresenta l'entità dei residui di Pearson rispetto a un modello di indipendenza. In breve una colorazione blu più o meno marcata indica una frequenza dell'incrocio maggiore rispetto a quella attesa in caso di indipendenza tra la variabile dipendente e quella indipendente, mentre una colorazione rossa indica una frequenza minore.

Figura 1 – Raggiungibilità delle aziende per tecnica di rilevazione



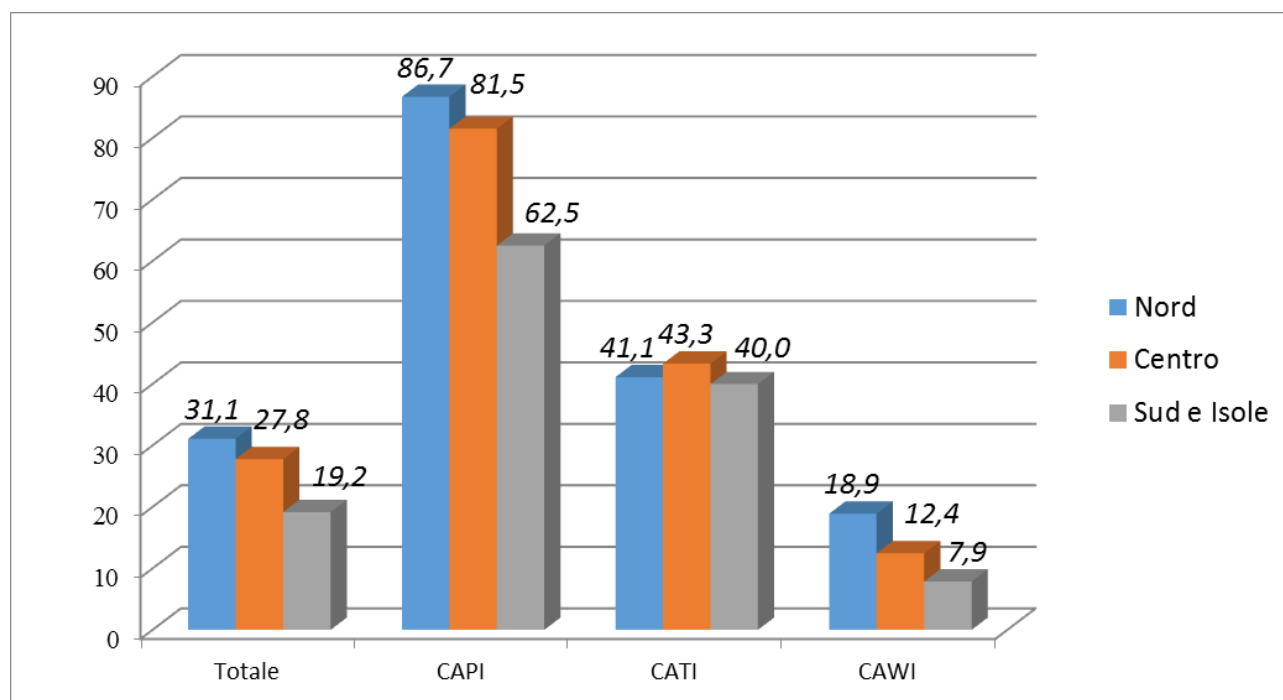
Rispetto alla dimensione territoriale, le aziende situate nel Nord del paese sono, in media, più raggiungibili rispetto a quelle localizzate nel Centro e soprattutto nel Sud. Dal grafico a mosaico si evince, infatti che nel Sud la raggiungibilità è di gran lunga inferiore a quella attesa in caso di indipendenza (figura 2).

Figura 2 – Raggiungibilità delle aziende per ripartizione geografica



Per la CATI la ripartizione geografica non sembra essere discriminante in termini di raggiungibilità, come dimostrano i valori registrati da tutte e tre le ripartizioni per cui la percentuale di aziende raggiunte si attesta intorno al 40%. Si osserva, inoltre, che la maggiore incidenza delle aziende raggiunte dal CATI si registra nel Centro, diversamente da quanto accade per le altre tecniche per le quali il settentrione registra il valore più alto. La tecnica CAWI, pur riproducendo la stessa tendenza che si manifesta per la tecnica CAPI, presenta valori decisamente inferiori alla media generale per tutte e tre le ripartizioni geografiche (figura 3).

Figura 3 – Distribuzione delle aziende raggiunte per tecnica di rilevazione e ripartizione geografica – valori percentuali



Fonte: indagini pilota del Censimento Agricoltura 2018

6.2 Il modello logistico

Il modello logistico utilizzato contiene le sole variabili rispetto al set incluso nel modello completo risultate statisticamente significative attraverso i valori del p-value (tavola 3). Tutte le variabili individuali (età e genere del conduttore) influenzano la raggiungibilità dell'azienda in maniera significativa. Per quanto attiene alle dimensioni strutturale e territoriale, la vocazione zootecnica e l'utilizzo di manodopera in forma indipendente non sembrano essere rilevanti, così come l'ubicazione dell'azienda all'interno di un comune montano e la zona altimetrica di riferimento. Parimenti, la presenza di segnali nelle fonti Anagrafica Agea, Anagrafe Zootecnica e Camere di Commercio non presentano valori significativi all'interno del modello.

Rispetto alla componente territoriale, si conferma la rilevanza della ripartizione geografica nell'influencare la raggiungibilità delle aziende: come emerso dall'analisi descrittiva le aziende del Nord hanno una maggiore propensione a essere raggiunte rispetto a quelle del Centro e del Sud, a parità di condizioni. Anche il grado di urbanizzazione del territorio contribuisce alla raggiungibilità: le aziende ubicate nei comuni con un elevato grado di urbanizzazione tendono ad avere una maggiore propensione ad essere raggiunte, al contrario di quelle localizzate nei comuni litoranei.

Le aziende condotte da uomini, così come quelle guidate da persone anziane, è più probabile che vengano raggiunte, il che dimostra che una maggiore stabilità nella conduzione migliora la capacità di rispondere alle indagini censuarie. La dimensione dell'azienda, al contrario, opera in direzione opposta: sono le micro aziende ad avere una maggiore probabilità di essere raggiunte rispetto a quelle con maggiore superficie agricola totale.

La tecnica, come atteso, ha influenzato fortemente i risultati ottenuti dalle indagini pilota, confermando la CAPI come la tecnica con una maggiore capacità di penetrazione delle aziende agricole. Tra le fonti amministrative e statistiche in cui le aziende risultano avere segnali di esistenza, sono AGEA e Catasto quelle che influenzano maggiormente la raggiungibilità, oltre a quelle rilevate dal precedente Censimento.

Tavola 3– Coefficienti stimati del modello logistico (log odds ratio) sulla raggiungibilità delle aziende agricole

<i>Variabili esplicative</i>	<i>Coefficiente stimato</i>
Intercetta	0,74***
Genere: Donne	-0,07*
Classe di età: 50<età <=70	0,15**
Classe di età: età>70	0,29***
Ripartizione geografica: Nord	0,26***
Ripartizione geografica: Sud	-0,42***
Comune litoraneo	-0,16***
Grado di urbanizzazione media	-0,15*
Grado di urbanizzazione bassa	-0,20**
Tecnica: CATI	-1,99***
Tecnica: CAWI	-3,35***
1<SAT<=10	-0,31***
SAT>10	-0,31***
Archivio: AGEA	0,45***
Archivio: Redditi Agrari	0,23***
Archivio: Catasto	0,09 ^Δ
Archivio: Aziende produttrici di prodotti DOP/IGP	0,18 ^Δ
Archivio: Aziende Agrituristiche	0,43**
Archivio: Censimento Agricoltura 2010	0,23***

* p<0.05 ** p<0.01 *** p<0.001 ^Δ p<0.1

Data la difficile interpretabilità dei risultati in termini di log-odds, si è proceduto al calcolo degli esponenziali dei coefficienti, in modo da ottenere direttamente i valori di odds ratio e i relativi intervalli di confidenza (tavola 4).

Al netto della tecnica, la variabile che in valore assoluto discrimina maggiormente la probabilità di un'azienda agricola di essere raggiunta rispetto a quella di non esserlo è la fonte amministrativa o statistica in cui l'azienda presenta dei segnali di esistenza in vita: l'appartenenza ad Agea e all'indagine sugli Agriturismi aumentano la probabilità di raggiungibilità dell'azienda, rispettivamente, del 56 e del 53 per cento.

Tavola 4 – Modello logistico - odds ratio e intervalli di confidenza

<i>Variabili esplicative</i>	<i>odds</i>	<i>2,5%</i>	<i>97,5%</i>
Intercetta	2,09	1,66	0,26
Genere: Donne	0,93	0,12	0,16
50<Età <=70	1,16	0,03	0,04
Età >70	1,33	0,87	0,99
Nord	1,29	0,78	0,93
Sud	0,66	0,75	1,00
Comune litoraneo	0,85	0,71	0,95
Grado di urbanizzazione media	0,86	1,17	1,42
Grado di urbanizzazione bassa	0,82	0,61	0,71
Tecnica CATI	0,14	1,05	1,29
Tecnica CAWI	0,03	1,20	1,48
1< SAT <= 10	0,73	0,68	0,78
SAT >10	0,74	0,66	0,82
Archivio: AGEA	1,56	1,39	1,75
Archivio: Redditi Agrari	1,26	1,17	1,36
Archivio: Catasto	1,09	0,99	1,21
Archivio: Aziende produttrici di prodotti DOP/IGP	1,20	0,98	1,46
Archivio: Aziende Agrituristiche	1,53	1,13	2,05
Archivio: Censimento Agricoltura 2010	1,26	1,15	1,38

La seconda variabile in termini di rilevanza esplicativa del modello è l'età del conduttore, in particolare nella fascia superiore ai 70 anni, che influenza la raggiungibilità dell'azienda molto più di quanto faccia il genere; le aziende condotte da una donna, infatti, hanno una probabilità di essere raggiunte più bassa del 7 per cento rispetto a un'azienda condotta da un uomo. Altrettanto significativa la localizzazione nelle ripartizioni geografiche: rispetto a un'azienda del Centro, una del nord o del sud ha una probabilità di essere raggiunta rispettivamente, maggiore e minore, di circa il 30 per cento. Anche l'estensione della superficie agricola risulta avere un effetto apprezzabile: in particolare, le aziende con una SAT superiore all'ettaro hanno una probabilità di essere raggiunte ridotta di circa un terzo rispetto a quella delle microaziende.

I risultati sopra esposti sono supportati anche dai valori relativi agli intervalli di confidenza di Odds Ratio, che confermano l'assenza di una differenza apprezzabile in termini di raggiungibilità tra i gruppi di aziende relativi al sesso del conduttore, all'ubicazione dell'azienda in un comune con urbanizzazione media, alle fonti di provenienza corrispondenti al Catasto e alle Aziende produttrici di prodotti DOP/IGP.

Le probabilità associate alla raggiungibilità dell'azienda sono state calcolate congiuntamente per tutti i profili di unità corrispondenti alle variabili esplicative contenute nel modello. In questo modo si è calcolata la probabilità condizionata a tutte le combinazioni possibili delle caratteristiche individuali, strutturali e territoriali dell'azienda agricola (tavola 5).

Tavola 5 – Probabilità associata ai profili di azienda agricola

Tecnica	Genere	Comune litoraneo	Urbanizzazione	Rip. geografica	Età classe	SAT classe	Agea	Redditi Agrari	Catasto	Dop/ Igp	Agriturismi	Cens 2010	P
CAWI	2	1	3	Sud	1	2	0	0	0	0	0	0	0,020
CAWI	2	1	3	Sud	1	3	0	0	0	0	0	0	0,023
CAWI	2	1	2	Sud	1	2	0	0	0	0	0	0	0,024
CAWI	2	1	2	Sud	1	3	0	0	0	0	0	0	0,024
CAWI	1	1	3	Sud	1	2	0	0	0	0	0	0	0,024
....												
CATI	2	0	1	Nord	3	3	1	0	0	0	1	1	0,505
CAA	2	0	1	Sud	1	2	0	0	1	0	0	0	0,505
CATI	1	0	1	Nord	2	1	1	1	0	1	0	0	0,505
CATI	1	1	1	Centro	1	1	1	1	1	0	1	1	0,505
CATI	1	1	2	Centro	3	1	1	1	0	1	1	0	0,505
...
CAA	1	0	2	Nord	3	1	1	1	1	1	1	1	0,939
CAA	1	0	1	Nord	2	1	1	1	1	1	1	1	0,940
CAA	1	0	1	Nord	3	1	1	1	0	1	1	1	0,942
CAA	2	0	1	Nord	3	1	1	1	1	1	1	1	0,943
CAA	1	0	1	Nord	3	1	1	1	1	1	1	1	0,947

L'ordinamento delle probabilità consente di classificare la raggiungibilità di tutti i profili di azienda agricola. La probabilità più bassa, pari al 2 per cento, si ottiene nel caso di un'azienda condotta da una donna con meno di 50 anni, localizzata in un comune a bassa urbanizzazione, litoraneo, del Sud Italia, con una superficie agricola di oltre 10 ettari e che non presenta segnali in nessuna delle fonti amministrative e statistiche più rilevanti (Agea, Redditi agrari, Catasto, Dop/Igp, Agriturismi e Censimento), intervistata via CAWI. Un'azienda intervistata dal CAA, condotta da un uomo anziano, localizzata in un comune non litoraneo ad alto grado di urbanizzazione del Nord, con una superficie inferiore all'ettaro, che presenta segnali di esistenza nelle principali fonti amministrative e statistiche, ha una probabilità di essere raggiunta del 95 per cento.

Per riuscire ad apprezzare il singolo impatto di ciascuna variabile esplicativa sulla raggiungibilità è stata calcolata le medie delle probabilità associate a ciascuna delle possibili combinazioni con tutte le altre variabili. I valori così ottenuti possono rappresentare una stima della probabilità associata a ciascuna variabile esplicativa, *ceteris paribus* (Tavola 6). Si confermano i risultati suggeriti dai coefficienti e dagli odds ratio del modello: le probabilità rimangono quasi costanti nel caso del genere del conduttore (0,40 per i maschi e 0,39 per le femmine), mentre si diversificano sensibilmente nel caso della tecnica di raccolta dati utilizzata (passando da 0,11 per la CAWI a 0,76 per la CAPI).

Tavola 6 – Probabilità di raggiungibilità per caratteristiche aziendali

		<i>Media</i>	<i>Minimo</i>	<i>Massimo</i>
Genere del conduttore	Maschi	0,40	0,02	0,95
	Femmine	0,39	0,02	0,94
Classe di età del conduttore	eta <=50	0,37	0,02	0,93
	50<eta <=70	0,40	0,03	0,94
	eta>70	0,42	0,03	0,95
Ripartizione Geografica	Centro	0,41	0,03	0,93
	Nord	0,45	0,04	0,95
	Sud	0,34	0,02	0,90
	Comune Litoraneo	0,38	0,02	0,94
Grado di urbanizzazione	Alta	0,42	0,03	0,95
	Media	0,39	0,02	0,94
	Bassa	0,38	0,02	0,94
tecnica	CAPI	0,76	0,40	0,95
	CATI	0,32	0,08	0,71
	CAWI	0,11	0,02	0,38
Classe dimensionale dell'azienda (espressa in SAT)	0=<sat<=1	0,43	0,03	0,95
	1<sat<=10	0,38	0,02	0,93
	sat>10	0,38	0,02	0,93
Fonti di provenienza nel FR15	AGEA	0,43	0,03	0,95
	Redditi Agrari	0,42	0,03	0,95
	Catasto	0,40	0,02	0,95
	Aziende produttrici di prodotti DOP/IGP	0,41	0,03	0,95
	Aziende Agrituristiche	0,43	0,03	0,95
	Censimento Agricoltura 2010	0,42	0,03	0,95

6.3 Simulazione della raggiungibilità per il Censimento del 2020

I risultati ottenuti con il modello sopra esposto sono stati utilizzati per effettuare una breve simulazione di quelli che potrebbero essere gli esiti del prossimo Censimento dell'Agricoltura 2020.

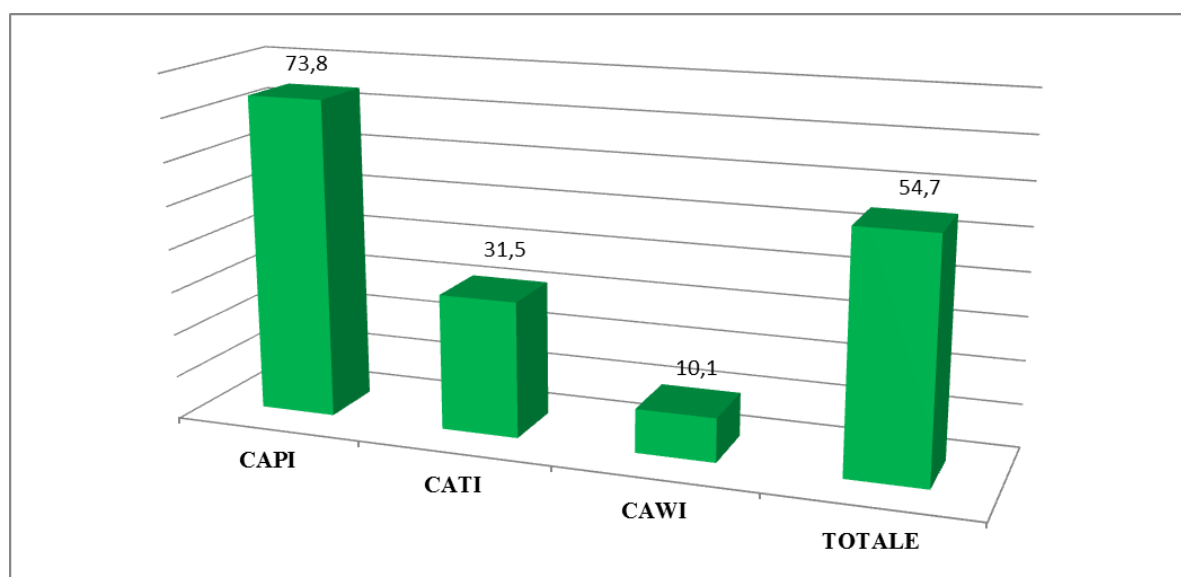
La rilevazione censuaria è stata progettata, anche sulla base dei tassi di risposta delle indagini pilota, stimando la partecipazione delle aziende ai tre canali di rilevazione nel seguente modo: 15% CAWI, 22% CATI e 63% CAPI. Pertanto, nella simulazione le aziende sono state attribuite ai tre canali rispettando le stesse quote. Per la costruzione della lista censuaria si è utilizzata l'ultima versione disponibile del Farm Register (*i.e.* 2016), applicando le stesse soglie dimensionali previste per il Censimento (*cfr.* par. 2).

Sfruttando le informazioni relative ai Centri di Assistenza Agricola disponibili¹⁴, si è effettuata un'assegnazione delle aziende alle tecniche di raccolta dati seguendo gli stessi criteri stabiliti per l'indagine censuaria. Postulando una partecipazione al canale web pari al 15 per cento, le aziende che nel periodo di riferimento non avranno autocompilato il questionario via web saranno assegnate ai CAA, insieme a quelle che hanno il centro aziendale nella stessa provincia del CAA di riferimento, oltre a tutte le altre per le quali non si dispone di contatti telefonici. Il complemento, pari a poco oltre un quinto delle aziende, dovrebbe essere raggiunto attraverso il canale telefonico. La lista è stata integrata con le informazioni relative alle variabili individuali, strutturali e territoriali inserite come esplicative nel modello logistico utilizzato.

A ciascuna azienda della lista censuaria *simulata* è stata attribuita, per ciascun profilo relativo alle modalità delle variabili indipendenti, la probabilità di essere raggiunta stimata dal modello. I risultati così ottenuti rappresentano una *proxy* di quelli che potrebbero essere i risultati del prossimo Censimento, al netto delle operazioni di ottimizzazione delle informazioni da lista che saranno implementate nei prossimi mesi.

La previsione, *ceteris paribus*, è di raggiungere il 55% delle aziende della lista censuaria (figura 4).

Figura 4 – Tassi di risposta attesi nel Censimento Agricoltura 2020 per tecnica – valori percentuali



Alle condizioni attuali, il tasso di risposta atteso per il CAWI, pari al 10 per cento, è in linea con quanto riscontrato nelle altre indagini, anche considerando il profilo medio dei conduttori delle aziende agricole italiane. Attraverso il canale CATI si raggiungerebbe un'azienda su tre mentre la rilevazione condotta dai CAA raggiungerebbe oltre il 70% di aziende, indipendentemente dall'assegnazione iniziale.

7. Conclusioni

La raggiungibilità delle aziende agricole è fortemente condizionata, oltre che dalla tecnica di raccolta dati, dalle variabili individuali, strutturali e territoriali dell'azienda agricola.

Le aziende più stabili, condotte da persone anziane, sono più facilmente raggiungibili, così come quelle a dimensioni ridotte situate in comuni ad alta urbanizzazione, non litoranei, preferibilmente nel Nord dell'Italia.

I risultati della simulazione vanno considerati al netto di una serie di attività, già progettate dall'Istituto, per incrementare l'efficacia delle indagini censuarie. Tra le altre, si annovera il potenziamento dei contatti telefonici sia per aumentare la quantità degli stessi integrando altre fonti informative nella lista sia per individuare un criterio gerarchico che consenta di selezionare, in caso di contatti multipli, quelli più affidabili.

¹⁴ Nella simulazione si è utilizzato il database fornito da Agea relativo all'anno 2017, essendo la versione più aggiornata di cui si dispone.

La tecnica CATI sembra essere meno influenzata dalle caratteristiche del territorio in cui si ubica l'azienda rispetto alla dimensione individuale e strutturale. Sarebbe particolarmente utile ai fini della riuscita del prossimo Censimento dell'Agricoltura 2020 individuare quali siano le principali caratteristiche individuali e strutturali che aumentano la raggiungibilità delle aziende mediante il canale telefonico in modo da massimizzare la probabilità di riuscita della tecnica Cati e ridurre, al contempo, il carico di lavoro dei CAA cui verranno riassegnati i residui del CATI. Pertanto uno sviluppo di questo lavoro è l'utilizzo del modello implementato al sottoinsieme delle aziende intervistate via CATI e l'individuazione dei profili di aziende più adatte, indipendentemente dal possesso del recapito telefonico, a rispondere a questo tipo di intervista.

Un'altra estensione di questo lavoro è rappresentata dallo studio dello stato di attività delle aziende agricole, utile al fine di stimare gli esiti delle indagini oltre che migliorare il modello probabilistico con cui si stabilisce l'esistenza in vita delle unità agricole.

ABSTRACT

In view of the Italian Census of Agriculture 2020 based on a multi-channel data collection method (Computer-Assisted web Interviewing — CAWI, Computer-Assisted Telephone Interviewing — CATI, Computer-assisted personal interviewing — CAPI), from November 2018 to January 2019 the Italian National Institute of Statistics has carried out three pretest surveys to check organization, methods and techniques to be used in the Agricultural Census 2020, that finally gives up the PAPI (Paper And Pencil Interviewing).

The pretest surveys involved a total of 30.100 agricultural holdings (19.816 farms for CAWI, 10.986 for CATI and 2.016 for CAPI), selected by stratified random sampling from the Statistical Farm Register (FR) 2015, composed of 1.659.928 farms. For CAPI survey the FR has been integrated with a database supplied by the Italian Agency for Agricultural Payments, related to 959.772 farms that opened an agricultural dossier in 2015.

The aim of this work is to analyze the pretest surveys results in terms of reachability by considering not only the characteristics of the holder and of the farm, but also territorial and geographical variables related to holding localization. To this end, the analysis has been conducted in two steps. In the descriptive analysis, the results have been analyzed in terms of individual, structural and territorial variables, also considering the different data collection method. In the econometric analysis, a logit model has been performed to point out the most relevant explicative variables for reachability that emerged from the pretest surveys.

Keywords: pretest surveys, Agricultural Census, territory, linear model

Bibliografia

- Kabacoff R.I. (2015) *R in Action: Data Analysis and Graphics with R* Manning Publications
- Fabbris L. (2011) *Statistica multivariata: analisi esplorativa dei dati* McGraw-Hill Education
- Gismondi, R. , Gnesi, C. (2019) *7° CENSIMENTO GENERALE DELL'AGRICOLTURA: Modello organizzativo e ruoli degli organi di Censimento* Istituto Nazionale di Statistica Direzione centrale per le statistiche ambientali e territoriali
- Gismondi, R., Greco, M., (2019) *7° CENSIMENTO GENERALE DELL'AGRICOLTURA: Indagine pilota con i Centri di Assistenza Agricola (CAA)* Istituto Nazionale di Statistica Direzione centrale per le statistiche ambientali e territoriali
- Gnesi, C. (2019) *La sperimentazione con la rete CAA* Comitato consultivo sul Censimento Agricoltura
- Legge 27 dicembre 2017, n. 205 recante il *Bilancio di previsione dello Stato per l'anno finanziario 2018 e Bilancio pluriennale per il triennio 2018-2020*
- Regolamento (UE) 2018/1091 del Parlamento Europeo e del Consiglio del 18 luglio 2018 relativo alle *Statistiche integrate sulle aziende agricole*
- Regolamento di Esecuzione (UE) 2018/1874 della commissione del 29 novembre 2018 sui dati da presentare per l'anno 2020 a norma del regolamento (UE) 2018/1091 del Parlamento europeo e del Consiglio relativo alle *Statistiche integrate sulle aziende agricole*