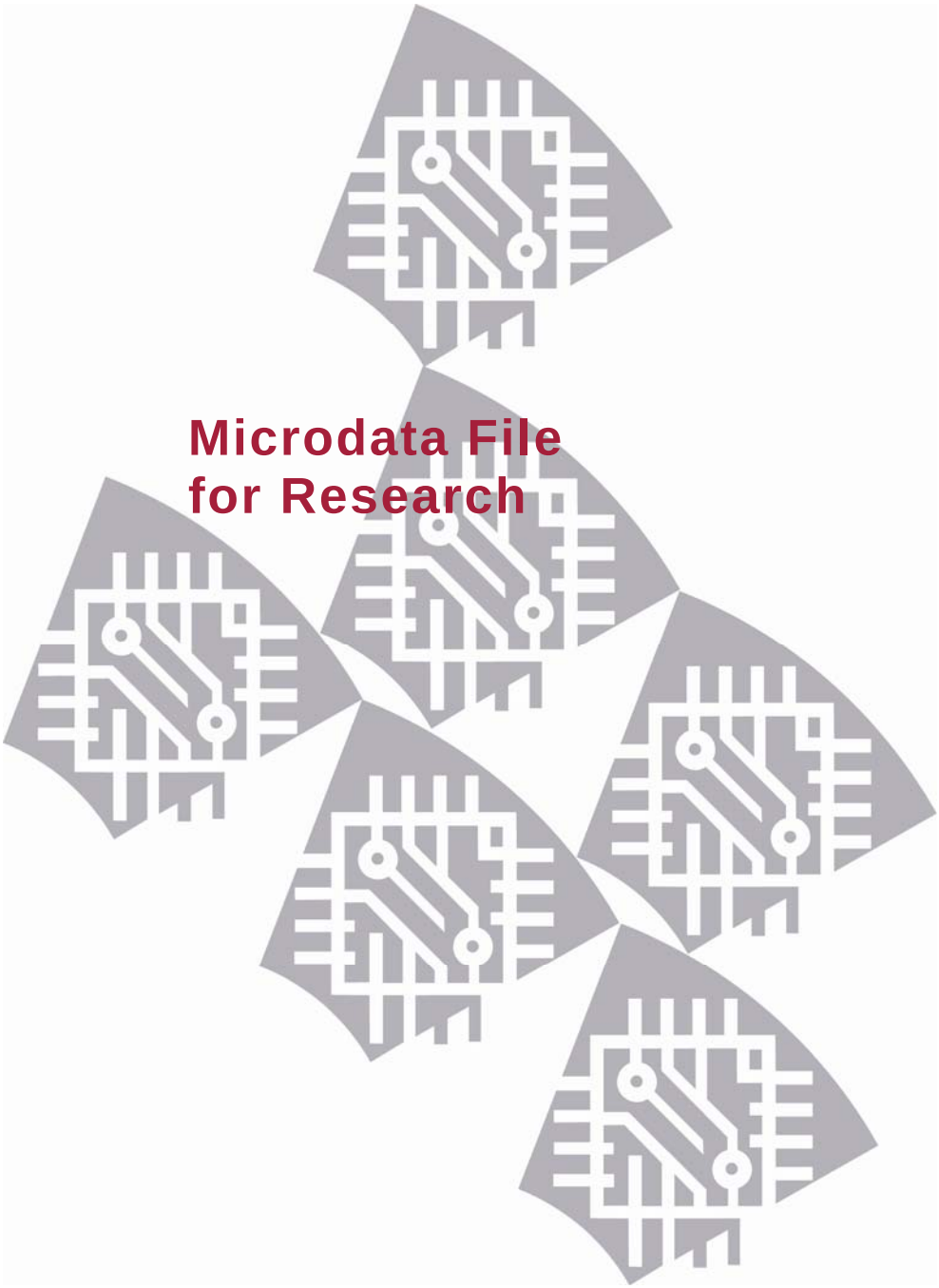
Istat
Istituto Nazionale di Statistica



Rilevazione comunitaria sulla struttura e produzione delle aziende agricole

Anno di riferimento: 2005

Descrizione del file

Luglio 2009

INDICE

1. Introduzione 3
2. Descrizione delle variabili 4
3. Metodologia statistica per la tutela della riservatezza 5
4. Analisi del contenuto informativo 6
5. Riferiementi bibliografici 10
6. Contatti 10

1. Introduzione

Nella raccolta dei dati l'Istat deve garantire che l'informazione statistica non sia rilasciata quando informazioni riservate relative ad un singolo rispondente possano essere dedotte in modo troppo accurato. Per il rilascio di un file di dati elementari per la ricerca (MFR), per la rilevazione sulla struttura e produzione delle aziende agricole 2005 (SPA 2005), è stata messa a punto una metodologia statistica per la riduzione del rischio di violazione della riservatezza dei rispondenti.

L'utente del file di microdati per la ricerca deve rispettare le condizioni stabilite nel contratto di rilascio.

Informazioni sullo status di ciascuna variabile nel MFR sono fornite nella sezione 2. La metodologia statistica per la tutela della riservatezza è brevemente illustrata nella sezione 3. Nella sezione 4, l'impatto del metodo di protezione è valutato attraverso alcuni confronti con i dati elementari originali.

2. Descrizione delle variabili

Questo paragrafo riassume i cambiamenti delle variabili relative al file di dati elementari originale. Tali cambiamenti sono necessari al fine di proteggere la riservatezza dei rispondenti.

Le etichette e i nomi delle variabili seguono il Regolamento della Commissione 2139/2004/EC.

Il separatore decimale è ".".
Nel file delimitato da tabulatore, la prima riga contiene i nomi delle variabili.
Il carattere missing è ":".
Il MFR contiene informazioni sulle aziende agricole.
Il MFR contiene 47780 records e 319 variabili. La somma dei coefficienti di riporto all'universo (variabile A09) è pari a 1728532.

Le variabili seguenti sono state soppresse: A04A, A04D, A05, A08, A09A.

Alcune variabili numeriche sono state soppresse; le seguenti relazioni fra le variabili sono state preservate:

Variabili soppresse	**Variabili rilasciate**	**Relazione**
D09E1, D09F, D09G	D09	D09 = D09E1 + D09F + D09G
D18A, D18B, D18B1, D18B23	D18	D18 = D18A + D18B + D18B1 + D18B23
F01, F02	F	F = F01 + F02
G01A, G01B, G01C	G01	G01 = G01A + G01B + G01C
G03A, G03B	G03	G03 = G03A + G03B
G04C, G04D	G04A, G04B, G04	G04 = G04A + G04B + G04C + G04D
I08AD22, I08B, I08C, I08D, I08E	I08	I08 = I08AD22 + I08B + I08C + I08D + I08E
J09A, J09B	J09	J09 = J09A + J09B
J10A, J10B	J10	J10 = J10A + J10B
J16A, J16B, J16C, J16D	J16	J16 = J16A + J16B + J16C+J16D
D01, D02, D03, D04, D05, D06, D08	**D01**	**D01** = D01 + D02 + D03+ D04 + D05 + D06 + D08
D10, D11, D12	**D10**	**D10** = D10 + D11 + D12
D23, D24, D25	**D23**	**D23** = D23 + D24 + D25
D26, D27, D28, D29, D30	**D26**	**D26** = D26 + D27 + D28 + D29 + D30
D31, D32, D33	**D31**	**D31** = D31 + D32 + D33

Tabella 1. Variabili numeriche soppresse e relazioni preservate. In grassetto le variabili costruite come aggregazioni di variabili rilevate.

Le variabili numeriche sono state perturbate, eccetto A06, A09, A12, A13, A14, A15, A16, A17, A18 e le variabili che iniziano per lettera "K" oppure "L". La variabile A06 è stata ricodificata.

3. Metodologia statistica per la tutela della riservatezza

3.1 Scenari di violazione della riservatezza

Considerando le informazioni registrate nella SPA 2005, sono stati adottati due scenari di violazione della riservatezza: uno sulla base delle informazioni contenute nelle variabili strutturali e uno sulla base delle informazioni contenute nelle variabili non-strutturali. Anche la tipologia del rilascio è stata presa in considerazione.

3.2. Modifiche del contenuto informativo delle variabili

3.2.1 Soppressione delle variabili

Le variabili soppresse dal MFR sono le seguenti:

1. Variabili identificative dirette (ragione sociale, indirizzo, ecc): ad esempio A08, A09A;
2. Variabili considerate troppo dettagliate: A04A, A04D;
3. Variabili che provocano frequenze (campionarie e nella popolazione) troppo basse (inferiori a 3): A05;
4. Alcune variabili numeriche che contribuiscono alla determinazione dell'orientamento tecnico-economico (OTE) ad un livello troppo dettagliato: prima colonna della Tabella 1.

La struttura del MFR è stata preservata sostituendo i valori delle variabili soppresse con il carattere ":".

3.2.2 Ricodifica globale

Per il MFR SPA 2005, la variabile A06 è stata ricodificata in modo da contenere informazioni solo sulle OTE *particolari* (50 categorie, OTE3).

L'informazione sulla collocazione geografica è rilasciata a livello regionale (NUTS2), variabile A07.

3.3. Modifica delle variabili numeriche

Il rischio di violazione della riservatezza delle aziende agricole è stato ridotto utilizzando la soppressione di alcune variabili, l'aggregazione di alcune variabili e l'individual ranking. Le variabili soppresse sono indicate nella prima colonna della Tabella 1. Le variabili che sono state derivate come aggregazioni di variabili rilevate sono indicate in grassetto nella seconda colonna della Tabella 1. La microaggregazione è stata applicata a tutte le variabili numeriche, ad eccezione di A06, A09, A12, A13, A14, A15, A16, A17, A18 e le variabili che iniziano per "K" oppure "L". La microaggregazione è stata applicata indipendentemente per ciascuna variabile numerica (individual ranking), vedi Defays *et. al.* (1998). Il parametro k dell'individual ranking è stato posto pari a 3. La variabile A07 è stata utilizzata come variabile di bloccaggio. Una versione pesata dell'individual ranking è stata applicata, usando la variabile A09 (coefficiente di riporto all'universo) come variabile "peso". Un'ulteriore controllo è stato implementato per tener conto della sparsità del fenomeno osservato.

4. Analisi del contenuto informativo

Nell'applicazione delle metodologie di riduzione del rischio di violazione della riservatezza, è inevitabile una perdita di informazione (se così non fosse sarebbe impossibile garantire la riservatezza dei rispondenti).

Ovviamente, il contenuto informativo delle variabili non modificate è rimasto invariato. Di conseguenza, circa il 64% del contenuto informativo delle variabili è rimasto invariato.

Variazioni delle medie pesate

Grazie all'utilizzo della variabile A07 come variabile di bloccaggio, le medie pesate delle variabili numeriche perturbate sono state preservate a livello regionale (NUTS2).

Sono stati calcolati i quartili delle variazioni delle medie pesate. Più precisamente, per ciascuna variabile numerica perturbata e per ciascuna combinazione di dettaglio geografico (NUTS2, variabile A07) e OTE *generali* (OTE1, variabile A06) , le medie pesate sono state calcolate sia per la variabile non-perturbata (WM_{orig}) che per la variabile perturbata (WM_{pert}). In seguito, la variazione percentuale è stata calcolata: $v = \frac{WM_{orig} - WM_{pert}}{WM_{orig}} * 100$. Nella Tabella 2, per ciascuna variabile numerica perturbata, sono indicati i quartili della distribuzione di v nelle 186 combinazioni di A07 e OTE *generali* (A06). Le distribuzioni asimmetriche sono una conseguenza del fenomeno osservato. Alcune variabili non sono state indicate nella Tabella 2 a causa della elevato numero di valori uguali a zero (fenomeno assente).

Variabile	Q1	Q2	Q3	Variabile	Q1	Q2	Q3
A11	-0.07	0.00	0.08	*G02*	-1.48	0.04	0.84
A13	0.00	0.00	0.00	*G03*	-0.20	-0.02	0.18
A14	0.00	0.00	0.00	*G04*	-0.26	0.00	0.21
A15	0.00	0.00	0.00	*G04A*	-0.65	0.03	0.87
A16	0.00	0.00	0.00	*G04B*	-0.45	0.00	0.35
A17	0.00	0.00	0.00	*G05*	-5.59	0.01	4.42
A18	0.00	0.00	0.00	*G06*	-18.26	1.19	13.89
CC01	-0.12	-0.01	0.07	*G07*	-6.62	-0.01	0.15
CC02	-0.28	0.00	0.25	*H01*	-0.79	0.00	0.66
CC03	-1.87	0.01	1.55	*H02*	-0.41	0.00	0.32
CC05A	-1.79	0.07	1.28	*H03*	-0.17	0.00	0.15
CC05D	-14.91	-0.27	15.31	*I01*	-1.47	0.00	1.31
D01	-0.10	-0.01	0.08	*I03A*	-0.24	-0.01	0.16
D07	-0.92	-0.01	1.02	*I03B*	-0.27	0.00	0.20
D09	-2.35	-0.17	1.88	*I08*	-0.63	0.01	0.55
D10	-1.40	0.00	1.36	*J01*	-0.71	0.00	0.57
D14	-0.85	0.00	0.90	*J02*	-0.19	0.00	0.10
D14A	-1.04	-0.12	0.35	*J03*	-0.76	0.00	1.19
D14B	-6.95	-0.03	7.04	*J04*	-0.45	0.00	0.31
D15	-3.35	-0.06	2.09	*J05*	-0.48	0.00	1.10
D16	-1.28	0.00	7.96	*J06*	-0.75	0.00	0.71
D17	-13.50	-0.04	3.87	*J07*	-0.85	0.00	0.26
D18	-0.30	-0.01	0.14	*J08*	-0.49	0.01	1.06

D19	-10.19	-0.23	11.50	*J09*	-0.50	-0.01	0.49
D21	-1.02	-0.01	1.24	*J10*	-0.63	0.00	0.51
D23	-7.31	-0.12	0.32	*J11*	-8.94	-0.06	5.56
D26	-1.09	-0.02	0.84	*J12*	-1.85	0.00	3.24
D31	-17.70	0.00	53.40	*J13*	-1.41	0.00	0.39
D34	-78.06	0.08	10.80	*J14*	-6.35	0.00	1.17
D35	-52.00	0.00	2.28	*J15*	-4.84	0.00	1.55
E	-0.09	0.00	0.04	*J16*	-24.69	-1.58	1.69
F	-0.32	0.00	0.20	*J17*	-11.30	0.00	3.25
G01	-0.75	-0.04	0.43	*J18*	-11.94	-0.36	4.32

Tabella 2. I quartili delle variazione percentuale delle medie pesate nelle 186 combinazioni di A07 and OTE *generali* (A06). Q1 = primo quartile, Q2 = secondo quartile (mediana) e Q3 = terzo quartile.

Variazioni delle varianze

Sono stati calcolati i quartili delle variazioni delle varianze. Più precisamente, per ciascuna variabile numerica perturbata e per ciascuna combinazione del dettaglio geografico (NUTS2, variabile A07) e OTE *generali* (OTE1, variabile A06), le varianze sono state calcolate sia per la variabile non-perturbata (V_{orig}) che per la variabile perturbata (V_{pert}). In seguito, la variazione percentuale è stata calcolata: $v = \frac{V_{orig} - V_{pert}}{V_{orig}} * 100$. Nella Tabella 3, per ciascuna variabile numerica perturbata, sono indicati i quartili della distribuzione di v nelle 186 combinazioni di A07 e OTE *generali* (A06). Le distribuzioni asimmetriche sono una conseguenza del fenomeno osservato. Alcune variabili non sono state indicate nella Tabella 3 a causa della elevato numero di valori uguali a zero (fenomeno assente).

Variabile	**Q1**	**Q2**	**Q3**	**Variabile**	**Q1**	**Q2**	**Q3**
A11	-1.11	0.00	1.67	*G02*	-5.63	0.00	3.26
A13	0.00	0.00	0.00	*G03*	-4.84	0.00	1.99
A14	0.00	0.00	0.00	*G04*	-1.98	0.00	2.78
A15	0.00	0.00	0.00	*G04A*	-2.53	0.10	6.78
A16	0.00	0.00	0.00	*G04B*	-5.55	0.00	3.47
A17	0.00	0.00	0.00	*G05*	-12.95	1.80	19.87
A18	0.00	0.00	0.00	*G06*	-57.24	5.18	35.90
CC01	-1.06	0.07	1.92	*G07*	-77.76	-7.10	46.46
CC02	-1.69	0.13	2.75	*H01*	-4.03	0.16	7.39
CC03	-9.11	0.18	8.91	*H02*	-3.04	0.25	5.12
CC05A	-6.40	0.38	6.12	*H03*	-4.39	0.00	2.93
CC05D	-45.09	2.95	52.85	*I01*	-7.30	0.32	9.54
D01	-1.78	-0.15	1.01	*I03A*	-1.35	0.02	2.41
D07	-1.54	0.51	10.04	*I03B*	-1.79	0.01	2.30
D09	-7.86	-0.36	12.23	*I08*	-4.42	0.27	4.06
D10	-3.90	0.70	6.83	*J01*	-9.04	0.00	8.95
D14	-4.37	0.09	6.89	*J02*	-0.81	0.00	2.28
D14A	-4.52	0.00	5.38	*J03*	-4.64	0.00	9.97
D14B	-10.34	0.49	23.38	*J04*	-2.43	0.00	3.25
D15	-9.86	0.00	13.25	*J05*	-3.99	0.00	13.05
D16	-11.11	1.12	36.45	*J06*	-2.43	0.00	5.76

D17	-28.82	2.29	18.17	*J07*	-1.35	0.00	4.07
D18	-4.18	-0.02	1.96	*J08*	-2.59	0.47	8.89
D19	-36.15	-0.24	33.86	*J09*	-2.67	0.00	3.64
D21	-5.96	0.25	8.11	*J10*	-2.98	0.00	7.56
D23	-18.03	-1.50	43.98	*J11*	-29.96	0.10	23.69
D26	-2.26	0.00	3.12	*J12*	-9.62	0.00	18.49
D31	-171.16	70.61	90.59	*J13*	-14.95	0.00	3.26
D34	-98.81	7.66	57.41	*J14*	-25.72	0.00	12.21
D35	-152.35	-109.14	84.67	*J15*	-12.48	-0.13	8.91
E	-0.38	0.00	0.39	*J16*	-95.04	-3.57	8.76
F	-1.73	0.32	2.88	*J17*	-42.57	0.00	22.31
G01	-5.53	-0.02	3.40	*J18*	-45.92	-4.78	25.45

Tabella 3. I quartili delle variazione percentuale delle varianze nelle 186 combinazioni di A07 e OTE generali (A06). Q1 = primo quartile, Q2 = secondo quartile (mediana) e Q3 = terzo quartile.

Numero di valori nulli modificati

Per la rilevazione SPA 2005, i valori nulli rappresentano un fenomeno assente. Applicando l'individual ranking, alcuni valori nulli potrebbero essere trasformati in valori non nulli positivi. In questo modo la perturbazione crea un fenomeno artificiale. Nella Tabella 4 è indicato, per ciascuna variabile numerica, il numero N di valori nulli che è stato trasformato in valori non nulli. I valori più elevati di N corrispondono alle variazioni più significative delle medie pesate/varianze.

Variabile	N	Variabile	N
A11	0	*G02*	2
A13	0	*G03*	3
A14	0	*G04*	0
A15	0	*G04A*	0
A16	0	*G04B*	0
A17	0	*G05*	3
A18	0	*G06*	0
CC01	0	*G07*	13
CC02	0	*H01*	0
CC03	0	*H02*	0
CC05A	0	*H03*	0
CC05D	7	*I01*	0
D01	0	*I03A*	0
D07	2	*I03B*	0
D09	2	*I08*	4
D10	0	*J01*	0
D14	0	*J02*	0
D14A	0	*J03*	0
D14B	0	*J04*	0
D15	1	*J05*	0
D16	7	*J06*	0
D17	5	*J07*	0
D18	0	*J08*	0
D19	6	*J09*	0
D21	1	*J10*	0

D23	7	*J11*	2
D26	4	*J12*	2
D31	6	*J13*	0
D34	11	*J14*	3
D35	16	*J15*	0
E	0	*J16*	2
F	0	*J17*	0
G01	0	*J18*	6

Tabella 4. Numbero N di valori nulli modificati.

Perturbazioni puntuali

L'impatto del metodo di perturbazione sulla perdita d'informazione potrebbe essere valutato utilizzando le perturbazioni puntuali. Nella tabella 5, è presentata per ciascuna variabile numerica, la mediana delle perturbazioni puntuali.

Variabile	Mediana	Variabile	Mediana
A11	-0.001	*G04A*	-0.002
CC01	-0.001	*G04B*	-0.001
CC02	-0.001	*G05*	0.002
CC03	-0.001	*G06*	0.002
CC05A	0.001	*G07*	-0.028
CC05D	-0.030	*H01*	-0.001
D01	-0.001	*H02*	-0.001
D07	0.017	*H03*	-0.001
D09	0.001	*I01*	0.001
D10	0.001	*I03A*	-0.001
D14	-0.001	*I03B*	-0.001
D14A	-0.001	*I08*	0.001
D14B	-0.001	*J01*	-1.000
D15	-0.001	*J02*	-1.000
D16	-0.002	*J03*	-1.000
D17	0.001	*J04*	-1.000
D18	0.001	*J05*	-1.000
D19	-0.002	*J06*	-1.000
D21	-0.001	*J07*	-1.000
D23	-0.001	*J08*	1.000
D26	-0.001	*J09*	1.000
D31	-0.075	*J10*	-1.000
D34	-0.013	*J11*	1.000
D35	-0.303	*J12*	1.000
E	-0.001	*J13*	-1.000
F	0.001	*J14*	-1.000
G01	-0.001	*J15*	-1.000
G02	-0.001	*J16*	-1.000
G03	-0.001	*J17*	-1.000
G04	-0.001	*J18*	-1.000

Tabella 5. Mediana delle perturbazioni puntuali.

5. Riferimenti bibliografici

D. Ichim (2009), "*Italian Farm Structure Survey: SDC Methodology for an MFR Dissemination*", Documenti Istat 2009, in corso di pubblicazione.

D. Defays and M.N. Anwar (1998) , "*Masking Microdata Using Micro-Aggregation*", Journal of Official Statistics, Vol.14, No.4, 1998. pp. 449-461.

6. Contatti

Per ulteriori dettagli sulla tutela della riservatezza, contattare rilascio.microdati@istat.it.