Istat
Istituto Nazionale di Statistica

File di microdati per la ricerca

Rilevazione statistica sull'innovazione nelle imprese

Periodo di riferimento: anni 2010 - 2012

Descrizione del file

Anno 2015

INDICE

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

1. Introduzione

I file di microdati sono collezioni di dati elementari relative ad alcune indagini svolte dall'Istat.

I file di microdati per scopi di ricerca (MFR) possono essere rilasciati su richiesta motivata e previa autorizzazione del Presidente dell'Istituto, purché resi anonimi e privi riferimenti che ne permettano il collegamento con singole persone fisiche o giuridiche.

I paragrafi 2 e 3 forniscono il dettaglio delle misure di protezione adottate e, rispettivamente, alcuni cenni sulla metodologia statistica per la tutela della riservatezza.

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

2. Descrizione delle variabili

In riferimento all'insieme di informazioni ottenuto secondo il questionario di indagine, la riduzione del rischio di violazione della riservatezza è stata perseguita adottando le misure di protezione indicate nel paragrafo 3.2. Per significati e articolazioni delle variabili rimaste immutate, si rinvia al questionario di indagine e al relativo tracciato record.

3. Metodologia statistica per la tutela della riservatezza

La produzione di file anonimizzati prevede una preventiva analisi del rischio di violazione della riservatezza del rispondente e una successiva procedura di protezione dei dati. La valutazione del rischio di intrusione viene svolta in riferimento alle variabili di scenario ossia, in termini intuitivi, quelle che si ipotizzano suscettibili di match con le informazioni d'indagine.

3.1. Violazione della riservatezza

Sono variabili identificative indirette (o chiave) quelle che da sole, o in combinazione con altre, possono portare alla re-identificazione di uno o più record. In particolare, sono considerate l'identificazione da archivio esterno e l'identificazione spontanea sulla base di conoscenze a priori ("circolo delle conoscenze"). L'esistenza di un archivio esterno pubblico e nominativo, contenente informazioni associabili a quelle contenute nel file rilasciato, permette all'intruso di identificare l'unità statistica ed apprendere informazioni su di essa. Diversa è la fattispecie dell'identificazione riconducibile al "circolo delle conoscenze" ossia ascrivibile ad informazioni sufficienti alla re-identificazione e note all'utente senza che venga posto in essere alcun tentativo di intrusione.

3.2. Protezione dei dati

Le misure di protezione adottate per tutelare statisticamente la confidenzialità (illustrate in Hundepool et al., 2012) sono riepilogate nella seguente tabella.

Tabella 1. Trattamento delle variabili a fini di tutela della riservatezza

Variabile	Descrizione	Misura di Protezione
codice_i	Codice Impresa	soppressione
nuts12	Localizzazione geografica	soppressione
div	Attività economica (divisione NACErev.2)	ricodifica: accorpamento classi 5 - 9 e 37 - 39
HO	Sede dell'impresa capogruppo	ricodifica: ITA, EU+EFTA, Resto del mondo, Non definito, Nessuna risposta/non applicabile
TURNMAR	Quota di fatturato 2012 derivante da prodotti o servizi nuovi (o significativamente migliorati) per il mercato di riferimento	perturbazione: arrotondamento casuale non distorto a multipli di 0.05
TURNIN	Quota di fatturato 2012 derivante da prodotti o servizi nuovi (o significativamente migliorati) solo per l'impresa	perturbazione: arrotondamento casuale non distorto a multipli di 0.05
TURNUNG	Quota di fatturato 2012 derivante da prodotti o servizi non modificati, o modificati solo marginalmente (compreso la vendita di nuovi prodotti acquistati da altre imprese)	perturbazione: arrotondamento casuale non distorto a multipli di 0.05
rrdinx	Spesa sostenuta per attività di R&S svolta all'interno dell'impresa nel 2010	rapporto con turn10, coerente con il top coding su rallx
rrdexx	Spesa sostenuta per l'acquisizione di servizi di R&S nel 2010	rapporto con turn10, coerente con il top coding su rallx
rmacx	Spesa sostenuta per l'acquisizione di macchinari e attrezzature nel 2010	rapporto con turn10, coerente con il top coding su rallx
roekx	Spesa sostenuta per l'acquisizione di altre tecnologie dall'esterno nel 2010	rapporto con turn10, coerente con il top coding su rallx
rdsgx	Spesa sostenuta per attività di progettazione (design) nel 2010	rapporto con turn10, coerente con il top coding su rallx
rotrx	Spesa sostenuta per altre attività innovative nel 2010	rapporto con turn10, coerente con il top coding su rallx
rallx	Spesa totale per l'innovazione sostenuta nel 2010	rapporto con turn10, top coding a partire da 1
turn10	Fatturato 2010	sostituzione con la variabile Gturn12 = turn12/turn10, top coding al 99° percentile
turn12	Fatturato 2012	perturbazione: microaggregazione

emp10	Addetti 2010	sostituzione con la variabile Gemp12 = emp12/emp10, top coding al 99° percentile
emp12	Addetti 2012	ricodifica: classi di addetti 10-49, 50-249, 250+
dom1	Dominio attività economica	soppressione

4. Analisi del contenuto informativo

Ad eccezione delle modifiche dovute alle misure di protezione riepilogate nella precedente tabella, i dati del MFR coincidono con i dati d'indagine. Per confrontare variabili originali e perturbate sono stati effettuati, in ciascuno strato definito da classe di attività economica e classe di addetti, i test univariati di Kolmogorov-Smirnov adattati alla presenza di intensità ripetute (Abadie, 2002) ed il test multivariato di identità distributiva proposto da Rizzo e Szekely (2010). I livelli di significatività osservati per i test univariati sono esposti nella tabella seguente:

Tabella 2. Livelli di significatività osservati per i test univariati

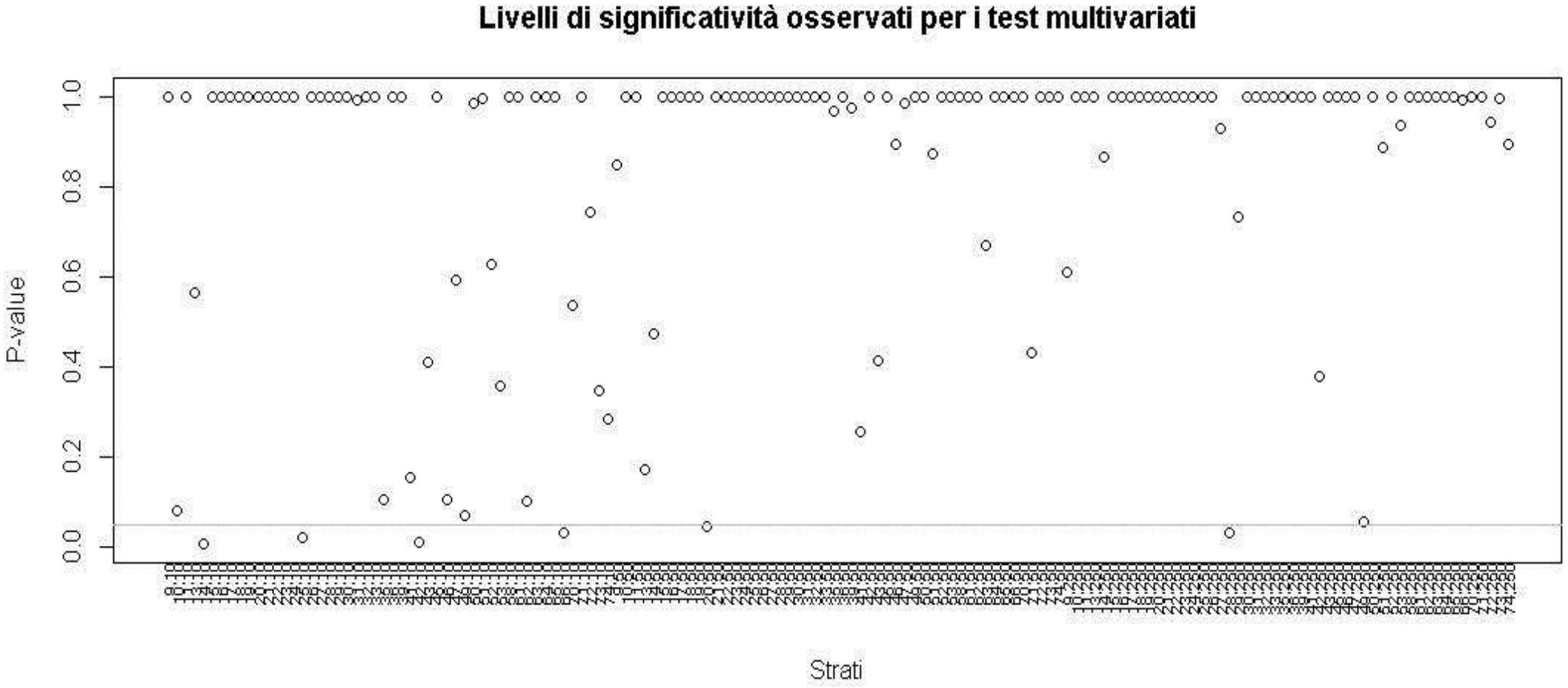
Strato	P-value Gemp12	P-value Gturn12	P-value turn12	P-value rallx	P-value TURNUNG	P-value TURNIN	P-value TURNMAR
9,10	1	1	1	1	0.913	1	0.542
10,10	1	1	1	0.942	0.263	0.885	0.837
11,10	1	1	1	1	0.853	0.992	0.834
13,10	1	1	1	1	0.987	0.997	1
14,10	1	1	1	1	0.989	0.977	0.831
15,10	1	1	1	1	1	1	1
16,10	1	1	1	1	0.997	0.998	1
17,10	1	1	1	1	0.877	1	0.979
18,10	1	1	1	1	0.917	0.997	1
19,10	1	1	1	1	0.696	1	1
20,10	1	1	1	1	0.814	0.877	0.935
21,10	1	1	1	1	0.911	0.994	1
22,10	1	1	1	1	0.948	0.999	0.968
23,10	1	1	1	1	0.96	1	0.974
24,10	1	1	1	1	0.993	0.982	0.998
25,10	1	1	1	1	0.992	0.869	0.91
26,10	1	1	1	1	1	1	1
27,10	1	1	1	1	1	0.998	0.996
28,10	1	1	1	1	0.934	0.892	0.957
29,10	1	1	1	1	0.922	1	1
30,10	1	1	1	1	0.987	0.992	0.943
31,10	1	1	1	1	0.803	0.751	1
32,10	1	1	1	1	1	0.998	1
33,10	1	1	1	1	0.875	0.985	1

35,10	1	1	1	0.976	0.792	0.954	0.617
36,10	1	1	1	1	0.711	0.341	0.602
39,10	1	1	1	1	0.772	1	1
41,10	1	1	1	0.792	0.841	0.914	0.966
42,10	1	1	1	1	1	1	1
43,10	1	1	1	0.774	0.485	0.463	0.915
45,10	1	1	1	1	0.356	0.347	0.84
46,10	1	1	1	0.955	0.032	0.496	0.377
47,10	1	1	1	0.583	0.272	0.553	0.708
49,10	1	1	1	0.719	0.899	0.902	0.714
50,10	1	1	1	1	0.94	0.928	1
51,10	1	1	1	1	0.935	0.937	1
52,10	0.97	0.959	1	0.452	0.995	0.976	0.986
53,10	0.998	0.996	1	0.694	0.86	0.913	1
58,10	1	1	1	1	0.998	1	1
61,10	1	1	1	1	1	1	1
62,10	1	1	1	1	0.87	0.999	0.998
63,10	1	1	1	1	0.944	0.999	1
64,10	1	1	1	1	0.662	0.99	0.992
65,10	1	1	1	1	0.987	1	0.84
66,10	1	1	1	0.999	0.935	0.996	0.993
70,10	1	1	1	1	0.998	1	1
71,10	1	1	1	1	0.764	0.981	0.995
72,10	1	1	1	0.046	0.941	0.944	0.964
73,10	1	1	1	1	1	1	0.985
74,10	1	1	1	1	0.591	1	1
9,50	1	1	1	1	0.999	0.97	0.677
10,50	1	1	1	1	0.797	0.895	1
11,50	1	1	1	1	0.135	0.505	0.791
13,50	1	1	1	1	0.972	1	1
14,50	1	1	1	1	0.92	0.976	0.877
15,50	1	1	1	1	0.999	0.994	0.986
16,50	1	1	1	1	0.913	0.91	0.973
17,50	1	1	1	1	0.993	0.706	0.748
18,50	1	1	1	1	0.76	1	0.996
19,50	1	1	1	1	0.992	0.979	1
20,50	1	1	1	1	0.822	0.557	0.689
21,50	1	0.999	1	1	0.987	1	1
22,50	1	1	1	1	0.247	0.897	0.994
23,50	1	1	1	1	0.998	1	0.981
24,50	1	1	1	1	1	0.977	1
25,50	1	1	1	1	0.594	0.965	0.822
26,50	1	1	1	1	0.996	0.973	0.999
27,50	1	1	1	1	1	0.997	1
28,50	1	1	1	1	0.671	0.915	0.877
29,50	1	1	1	1	1	1	1
30,50	1	1	1	1	0.994	0.972	1
31,50	1	1	1	1	0.968	1	1

32,50	1	1	1	1	0.958	0.986	0.98
33,50	1	1	1	1	0.945	0.875	0.97
35,50	1	1	1	1	0.795	0.994	1
36,50	1	1	1	1	0.997	0.997	1
39,50	1	1	1	1	0.952	0.975	0.73
41,50	1	1	1	1	0.936	0.98	0.984
42,50	1	1	1	1	0.999	0.992	1
43,50	0.989	0.998	1	0.809	0.753	0.911	0.955
45,50	1	1	1	1	1	0.993	0.997
46,50	1	1	1	1	0.22	0.396	0.627
47,50	1	1	1	0.957	0.766	0.935	0.596
49,50	1	1	1	0.957	0.981	1	0.864
50,50	1	1	1	1	0.981	0.95	1
51,50	1	1	1	1	0.82	0.803	1
52,50	0.33	0.427	1	0.057	0.942	1	0.917
53,50	1	1	1	1	1	1	1
58,50	1	1	1	1	0.754	0.982	0.977
61,50	1	1	1	1	0.933	0.86	0.988
62,50	1	1	1	1	0.705	0.956	0.997
63,50	1	1	1	1	0.935	0.979	0.918
64,50	1	1	1	1	0.052	0.263	0.384
65,50	1	1	1	1	0.953	1	0.854
66,50	1	1	1	1	1	0.827	0.91
70,50	1	1	1	1	0.695	0.799	0.914
71,50	1	1	1	1	1	0.981	0.981
72,50	1	1	1	0.407	0.894	1	0.839
73,50	1	1	1	1	0.606	0.751	0.95
74,50	1	1	1	1	1	1	1
9,250	1	1	1	1	1	1	1
10,250	1	1	1	1	0.288	0.954	0.439
11,250	1	1	1	1	0.896	0.729	0.994
13,250	1	1	1	1	0.926	0.991	1
14,250	1	1	1	1	0.892	0.994	0.987
15,250	1	1	1	1	0.977	1	1
16,250	1	1	1	1	0.979	0.985	0.957
17,250	1	1	1	1	0.855	0.961	1
18,250	1	1	1	1	0.615	1	0.91
19,250	1	1	1	1	0.849	0.534	0.86
20,250	1	1	1	1	0.931	0.247	0.699
21,250	1	1	1	1	0.243	0.704	0.79
22,250	1	1	1	1	0.557	0.665	0.996
23,250	1	1	1	1	0.876	0.734	0.993
24,250	1	1	1	1	0.656	0.568	0.334
25,250	1	1	1	1	0.819	0.813	0.885
26,250	1	1	1	1	0.979	1	0.864
27,250	1	1	1	1	0.871	0.987	0.146
28,250	1	1	1	1	0.262	0.943	0.801
29,250	1	1	1	1	0.905	1	0.982

30,250	1	1	1	1	0.783	0.646	0.279
31,250	1	1	1	1	0.879	1	1
32,250	1	1	1	1	1	1	1
33,250	1	1	1	1	0.997	1	0.993
35,250	1	1	1	1	0.929	0.689	1
36,250	1	1	1	1	0.912	0.88	0.992
39,250	1	1	1	1	0.73	0.814	0.873
41,250	1	1	1	1	0.992	0.993	0.549
42,250	1	1	1	0.986	0.993	0.937	0.981
43,250	1	1	1	1	0.897	0.714	0.81
45,250	1	1	1	1	1	1	1
46,250	1	1	1	1	0.297	0.58	0.328
47,250	1	1	1	0.989	0.506	0.33	1
49,250	1	1	1	1	0.613	0.907	0.261
50,250	1	1	1	1	0.997	0.981	0.994
51,250	1	1	1	1	0.8	0.756	1
52,250	1	1	1	1	0.845	0.682	0.8
53,250	1	1	1	1	0.659	0.656	0.546
58,250	1	1	1	1	0.824	0.719	0.69
61,250	1	1	1	1	1	0.999	1
62,250	1	1	1	1	0.904	0.748	0.999
63,250	1	1	1	1	0.422	0.997	0.852
64,250	1	1	1	1	0	0.026	0.064
65,250	1	1	1	1	0.498	0.377	0.985
66,250	1	1	1	1	0.956	0.963	1
70,250	1	1	1	1	0.965	0.794	0.974
71,250	1	1	1	1	0.947	0.996	1
72,250	1	1	1	0.94	1	1	1
73,250	1	1	1	1	0.658	0.266	1
74,250	1	1	1	1	1	1	1

La sostanziale identità delle distribuzioni univariate perturbate e originali non implica per sé che i record del MFR e d'indagine possano essere ritenuti realizzazioni campionarie della medesima popolazione generatrice. Tuttavia, le misure di protezione tramite perturbazione hanno riguardato una frazione limitata dei record, ed è ragionevole ritenere che anche i legami multivariati siano stati sostanzialmente mantenuti. Per verificare l'ipotesi di identica distribuzione (multivariata) dei record, restringendo l'attenzione alle sole variabili oggetto di perturbazione, i test delle ipotesi sono stati condotti previa standardizzazione delle intensità; i livelli di significatività osservati per ciascuno strato sono rappresentati nel diagramma della pagina seguente e si può constatare che eccedono la soglia del 5% in 144 strati su 150.


Livelli di significatività osservati per i test multivariati
P-value
0.0
0.2
0.4
0.6
0.8
1.0
Strati

5. Riferimenti bibliografici

Abadie, A. (2002). *Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models*. Journal of the American Statistical Association, 97:457 (March), 284-292.

Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E. S., Spicer, K. e de Wolf, P.-P. (2012). Statistical Disclosure Control. Wiley.

Istat (2004). Metodologie e tecniche di tutela della riservatezza nel rilascio di informazione statistica (Metodi e norme, n. 20, 2004), http://www.istat.it/dati/catalogo/20040706_00/.

Rizzo, M. L. e Szekely, G. J. (2010). *DISCO Analysis: A Nonparametric Extension of Analysis of Variance*. Annals of Applied Statistics, 4, 2, 1034-1055.

Willenborg, L. e de Waal, T. (2001). Elements of statistical disclosure control. Lecture Notes in Statistics, 115, New York: Springer-Verlag.

Curatore dei paragrafi

Il documento è stato redatto da Flavio Foschi.