

Nota metodologica INTANGIBLE ASSETS SURVEY-1

Denominazione: Rilevazione statistica sugli Investimenti Intangibili
(INTANGIBLE ASSETS SURVEY-1)

Codice PSN SISTAN: IAP-00012 (ex ISF-00059)

Rapporto di ricerca: Angotti Roberto (a cura di) (2017), *INTANGIBLE ASSETS SURVEY. I risultati della Rilevazione statistica sugli investimenti intangibili delle imprese*, INAPP, Roma, in <<https://goo.gl/wY27sy>> (23.11.2018)

1. La popolazione di riferimento dell'indagine

Il campo di osservazione dell'Indagine sugli Investimenti Intangibili è costituito dalle imprese attive con 10 o più addetti che operano nelle divisioni della Classificazione Ateco 2007 riportate in Tabella 1. Da essa si desume come siano state escluse le imprese appartenenti alle seguenti divisioni: 45, 55, 56, 68, 69, 70 e dal 74 al 99. Questa scelta è stata effettuata per focalizzare il campione su settori con maggiore densità di attività intangibili, sulla base di quanto emerso a seguito della indagine pilota.

Tabella 1 - Settori economici (Sezioni e Divisioni) inclusi nel campo di osservazione dell'indagine (Ateco 2007)

Sezione	Divisione
B. Estrazione di minerali da cave e miniere	05 - 06 - 07 - 08 - 09
C. Attività manifatturiere	10 - 11 - 12 - 13 - 14 - 15 - 16 - 17 - 18 - 19
	20 - 21 - 22 - 23 - 24 - 25 - 26 - 27 - 28 - 29
	30 - 31 - 32 - 33
D. Fornitura di energia elettrica, gas, vapore e aria condizionata	35
E. Fornitura di acqua; Reti fognarie, attività di gestione dei rifiuti e risanamento	36 - 37 - 38 - 39
G. Commercio all'ingrosso e al dettaglio; riparazione di autoveicoli e motocicli	46 - 47
H. Trasporto e magazzinaggio	49 - 50 - 51 - 52 - 53
J. Servizi di informazione e comunicazione	58 - 59 - 60 - 61 - 62 - 63
K. Attività finanziarie e assicurative	64 - 65 - 66
M. Attività professionali, scientifiche e tecniche	71 - 72 - 73

Nel complesso, l'universo di riferimento è composto da più di 120 mila imprese, presso le quali (sulla base dell'archivio ASIA 2011) risultavano occupati 5.945 milioni di addetti. Il 58,2% fa parte dell'industria, mentre il restante 41,8% appartiene ai servizi (Tab. 2). Nel dettaglio, i settori più importanti sono il commercio, con quasi un quarto delle imprese (29.570 unità), cui seguono con percentuali pressoché dimezzate la produzione di metalli (11,4%) e la meccanica (12,6%); sopra le 10 mila unità si collocano le attività di trasporto e il tessile. Questi cinque comparti comprendono il 56,8% del totale delle imprese, un valore che arriva a pesare per il 64,3% se si includono anche le imprese farmaceutiche e chimiche che contano per il 7,5% (9.096 unità).

Se si considera l'ampiezza aziendale, si tratta di un tessuto produttivo caratterizzato da un'elevata presenza (86,1%) di imprese di piccole dimensioni, vale a dire con un numero di addetti compreso fra 10 e 49 unità; gli addetti in esse occupati sono poco meno di due milioni e rappresentano il 33,3% del totale. Le imprese con almeno 250 ed oltre addetti costituiscono una quota marginale (2,3%); dal punto di vista occupazionale, tuttavia, esse rivestono un'indubbia importanza dal momento che presso di esse sono occupati oltre 2,5 milioni di addetti, pari al 34,6% del totale.

Per tutti i settori economici, tanto del secondario quanto del terziario, la prima classe dimensionale rappresenta il gruppo più numeroso, con punte che sfiorano o che oltrepassano la soglia del 90% nel caso del tessile, del comparto del mobile e del commercio. Le imprese sopra i 250 addetti sono il 2,3% dell'universo complessivo, con una quota che aumenta per la fabbricazione dei mezzi di trasporto e per i servizi finanziari (rispettivamente, 8,7% e 8,9%), così come, sebbene in forma più attenuata, per la fornitura di elettricità, gas, acqua e per le telecomunicazioni.

Tabella 2 - Imprese, per settore economico e classe dimensionale. Anno 2011 (v.a. e %)

Settore economico	Classe dimensionale					Totale
	10-49	50-99	100-249	250-499	500 ed oltre	
INDUSTRIA	60.224	5.877	2.990	873	565	70.529
Industrie estrattive	466	30	20	0	1	517
Industrie alimentari e del tabacco	5.970	489	237	86	36	6.818
Tessile e abbigliamento	9.117	683	284	73	48	10.206
Industria della carta, cartone, stampa	2.825	222	123	40	14	3.224
Industria chimica, farmaceutica, gomma, plastica	7.344	866	575	194	117	9.096
Produzione di metalli e prodotti in metallo	12.083	1.076	448	117	41	13.764
Apparecchi meccanici, elettrici, elettronici	12.722	1.494	724	178	157	15.274
Mezzi di trasporto	1.178	191	143	77	67	1.655
Legno, mobili e altre industrie manifatturiere	6.476	507	226	25	18	7.252
Fornitura di elettricità, gas, acqua e gestione rifiuti	2.043	319	211	84	66	2.723
SERVIZI	44.116	3.458	1.754	693	621	50.642
Commercio	26.895	1.594	651	208	222	29.570
Attività di trasporto, magazzinaggio e posta	9.488	1.037	536	217	153	11.431
Telecomunicazioni, editoria, informatica	4.492	361	263	142	121	5.379
Servizi finanziari, assicurazioni e fondi pensione	1.519	292	211	84	113	2.219
Attività professionali, scientifiche, tecniche	1.722	174	94	42	12	2.043
TOTALE	104.340	9.335	4.745	1.566	1.186	121.171

Fonte: ASIA (ISTAT)

Dalla Tabella 3 si evince come quasi i due terzi delle imprese hanno una sede legale situata nelle regioni settentrionali (35,5% nel Nord-Ovest e 31,4% nel Nord-Est); segue il Centro con il 19,1% e il Sud con il 14,1%. In tutte le ripartizioni le piccole imprese fra 10 e 49 addetti rappresentano la larga maggioranza del tessuto produttivo, sebbene questa frammentazione sia maggiore nelle regioni meridionali dove arrivano a pesare per l'89,1%. (Tab. 4)

Tabella 3 - Imprese, per classe dimensionale e ripartizione geografica. Anno 2011 (v.a.)

Classe dimensionale	Ripartizione geografica				TOTALE
	Nord-Ovest	Nord-Est	Centro	Sud ed Isole	
10-49	36.250	32.616	20.287	15.188	104.340
50-99	3.649	3.102	1.495	1.090	9.335
100-250	1.903	1.563	768	512	4.745
250-499	698	452	270	145	1.566
500 ed oltre	468	340	258	120	1.186
TOTALE	42.967	38.073	23.077	17.054	121.171

Fonte: ASIA (ISTAT)

Le aree geografiche del Paese sono accomunate dalla prevalenza di realtà imprenditoriali e produttive di piccole dimensioni; un elemento che le contraddistingue è dato, per certi versi, dalla differente composizione settoriale (Tab. 4). I servizi - che contano a livello generale per il 41,8% - nelle regioni meridionali arrivano a costituire il 48,4% del totale delle imprese, un risultato che si deve soprattutto all'aumento dell'incidenza del commercio (30,2%) e, in seconda battuta, delle attività di trasporto (12,5%). Un'altra specificità del Sud è rappresentata dalla maggiore consistenza dell'agroalimentare: 11 imprese su 100 appartengono a questo comparto a fronte di una media italiana del 5,6%. Le ripartizioni settentrionali, invece, si caratterizzano per una più spiccata vocazione industriale dal momento che le imprese afferenti questo macrosettore incidono per circa il 60%. In confronto al dato nazionale, poi, il Nord-Ovest si distingue per una maggiore rilevanza della produzione di metalli e della meccanica (rispettivamente, 15,0% e 14,5%), mentre nel Nord-Est si ravvisa una quota più importante, oltre che della meccanica, del comparto del mobile (8,1%).

Tabella 4 - Imprese, per settore economico e ripartizione geografica. Anno 2011 (v.a.)

Settore economico	Ripartizione geografica				TOTALE
	Nord-Ovest	Nord-Est	Centro	Sud ed Isole	
INDUSTRIA	26.030	22.849	12.852	8.798	70.529
Industrie estrattive	125	156	124	112	517
Industrie alimentari e del tabacco	2.345	1.878	755	1.840	6.818
Tessile e abbigliamento	2.628	2.838	3.655	1.086	10.206
Industria della carta, cartone, stampa	1.202	1.041	644	336	3.224
Industria chimica, farmaceutica, gomma, plastica	3.740	2.827	1.429	1.099	9.096
Produzione di metalli e prodotti in metallo	6.438	4.058	1.801	1.466	13.764
Apparecchi meccanici, elettrici, elettronici	6.238	5.865	1.769	1.402	15.274
Mezzi di trasporto	597	524	354	180	1.655
Legno, mobili e altre industrie manifatturiere	1.861	3.100	1.651	641	7.252
Fornitura di elettricità, gas, acqua e gestione rifiuti	855	562	670	636	2.723
SERVIZI	16.937	15.224	10.225	8.256	50.642
Commercio	9.918	8.917	5.581	5.155	29.570
Attività di trasporto, magazzinaggio e posta	3.103	3.755	2.447	2.127	11.431
Telecomunicazioni, editoria, informatica	2.228	1.448	1.212	490	5.379
Servizi finanziari, assicurazioni e fondi pensione	873	635	417	294	2.219
Attività professionali, scientifiche, tecniche	816	468	569	190	2.043
TOTALE	42.967	38.073	23.077	17.054	121.171

Fonte: ASIA (ISTAT)

2. Il piano di campionamento

Sulla base degli obiettivi della rilevazione (osservazione della distribuzione degli investimenti intangibili nelle imprese italiane con almeno 10 addetti nel corso del 2012) e della numerosità campionaria totale, è stata definita una strategia campionaria che consentisse di massimizzare la copertura delle imprese di maggiore dimensione (quelle con un livello atteso più elevato di investimenti intangibili) pur assicurando una copertura adeguata del fenomeno tra le piccole imprese e negli strati meno numerosi.

Il disegno di campionamento adottato, utilizzando le informazioni contenute nell'archivio ASIA 2011, è di tipo probabilistico a uno stadio stratificato, comunemente utilizzato per le indagini sulle imprese. In generale, l'obiettivo della stratificazione è quello di migliorare l'efficienza delle stime rispetto a un campione non stratificato della stessa dimensione. Tale obiettivo è stato reso possibile dalla disponibilità di una lista contenente, per tutte le unità della popolazione d'interesse, oltre alle informazioni identificative, ulteriori variabili caratterizzanti l'impresa e correlate con le variabili oggetto di studio: la dimensione dell'impresa, il settore di attività economica ed, infine, il territorio in cui è localizzata la sede locale.

La popolazione di riferimento è stata ripartita tra le imprese con almeno 50 addetti (soggette a censimento) e le imprese con 10-49 addetti (soggette a campionamento). Successivamente si è proceduto a stratificare le unità per settore di attività economica con riferimento alle Divisioni NACE rev.2 riportate in Tabella 1. Per la parte censuaria si è quindi proceduto all'estrazione dall'archivio ASIA dell'intera popolazione di riferimento. Per la parte campionaria è stato scelto un campionamento casuale semplice, ed è stata utilizzata la procedura SAS SURVEYSELECT per l'estrazione casuale delle unità oggetto di rilevazione dopo aver definito la numerosità campionaria.

Tenendo conto sia dei vincoli di disponibilità economica, sia degli obiettivi di indagine, ai fini del calcolo della numerosità campionaria e della relativa allocazione tra i suddetti domini, non disponendo di informazioni sufficienti sulla variabilità dei fenomeni di interesse, la procedura di allocazione è stata impostata considerando come parametro di stima una frequenza relativa del 10% in tutti i domini di interesse; in tal modo il disegno campionario è tale da garantire una sufficiente precisione per stime corrispondenti almeno al 10% a livello nazionale, per classe di addetti e per settore di attività economica. Tuttavia sono state considerate alcune eccezioni: gli strati con meno di 50 osservazioni sono stati censiti anch'essi, mentre negli strati con meno di 400 osservazioni il campione è stato fissato al 50% della popolazione, come illustrato in Tabella 5.

Tabella 5 - Stratificazioni delle unità, probabilità di selezione e numerosità campionaria

Numero Strati =5 con meno di 50 osservazioni.	Numero Strati =13 con 50-400 osservazioni.	Numero Strati =33 con oltre 400 osservazioni.
Probabilità di selezione=100% n=87	Probabilità di selezione=50% n=1.232	Probabilità di selezione=10% n=10.166

Il metodo adottato consente di definire le numerosità campionarie negli strati in modo tale da rispettare i vincoli sugli errori attesi di campionamento (espressi in termini di coefficienti di variazione) delle stime "tipiche", prefissati a livello dei differenti domini pianificati. I criteri adottati hanno portato alla definizione di una dimensione campionaria di 11.485 unità che, assieme alle imprese censite (16.836), determinano una numerosità totale del campione pari a 28.321 unità.

3. La rilevazione

La rilevazione, realizzata fra luglio e novembre 2013 da una società esterna incaricata dall'ISFOL, ha interessato in maniera censuaria le imprese con almeno 50 addetti e un campione di imprese dai 10 ai 49 addetti, per un totale di 28.321 unità.

Il primo contatto con le imprese del campione è avvenuto tramite PEC - Posta Elettronica Certificata - in coerenza con le recenti disposizioni di Agenda digitale. Inizialmente era previsto un sistema di contatto delle imprese tramite invio postale (lettera ufficiale e password per accedere al sito predisposto per la compilazione on line). Mentre era in corso la fase preparatoria della rilevazione è però entrata in vigore la normativa prevista dal Decreto del Presidente del Consiglio dei Ministri 22 luglio 2011 e successive modificazioni, sull'utilizzo della Posta Elettronica Certificata (PEC). È stato, quindi, utilizzato l'invio delle e-mail tramite PEC a circa 25.000 mila imprese estratte dall'archivio ASIA-ISTAT (Archivio Statistico delle Imprese Attive) e presenti nell'anagrafe PEC delle Camere di Commercio. Invece, per le imprese che ancora non si erano adeguate alla nuova normativa, si è fatto ricorso al contatto telefonico.

Per la raccolta dati è stata adottata una metodologia mista, con tecniche CAWI (Computer Assisted Web Interview) e CATI (Computer Assisted Telephone Interview), attraverso un questionario strutturato in otto sezioni. Il questionario è stato informatizzato attraverso lo sviluppo di questi due sistemi di rilevazione per consentire, se richiesto dall'impresa, la raccolta dati mediante intervista telefonica o la registrazione di questionari cartacei inviati per e-mail o fax.

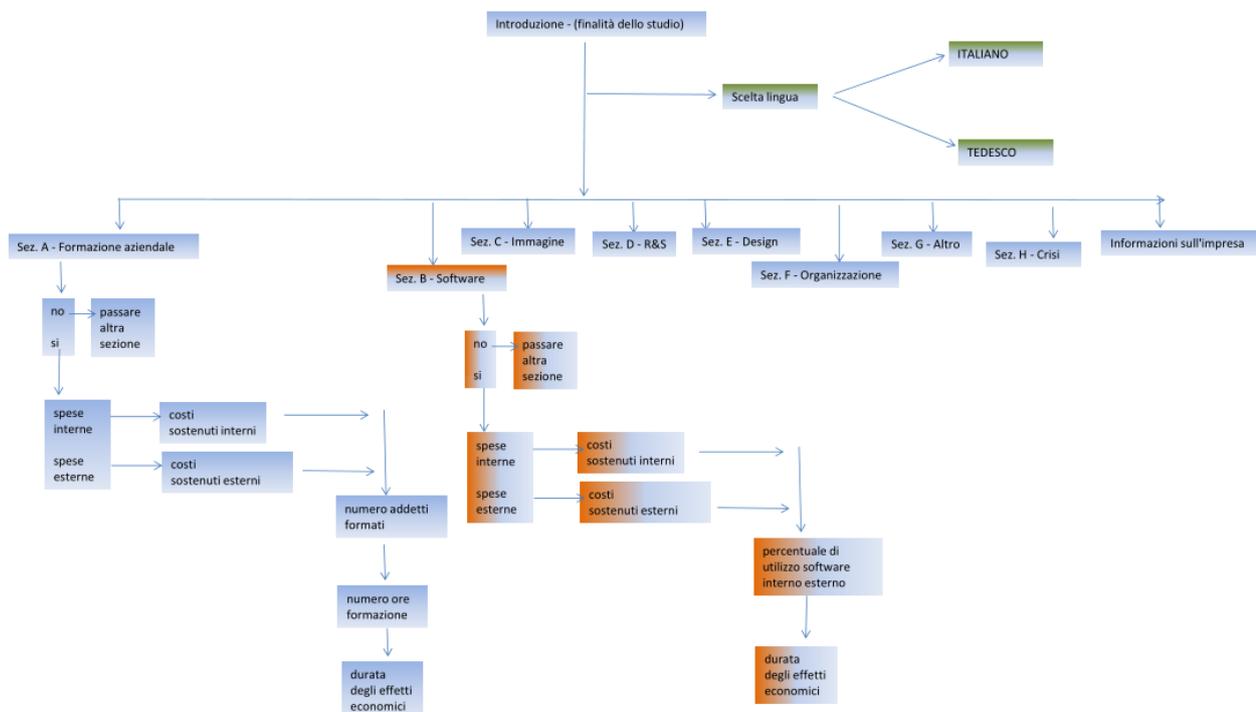
Il questionario di rilevazione è stato strutturato in modo modulare e non sequenziale (vedi schema 1) per consentire una compilazione più flessibile data la complessità e la natura prevalentemente quantitativa delle informazioni richieste. Tale scelta è stata determinata principalmente dalla consapevolezza che i dati richiesti non erano immediatamente disponibili soprattutto per le imprese con più investimenti intangibili. Per rispondere a tali esigenze è stato, quindi, predisposto un sistema di rilevazione che dava la possibilità di compilare parzialmente il questionario e completare la compilazione in più accessi successivi. Infatti, specialmente in imprese di medie e grandi dimensioni, solitamente organizzate in settori o unità specifiche, i rispondenti hanno dovuto cooperare con i responsabili delle diverse unità per compilare correttamente il questionario con i dati richiesti; mentre nelle piccole imprese la maggior parte delle informazioni sono state fornite dall'imprenditore, spesso con il supporto del commercialista.

Il sistema CAWI-CATI è stato sviluppato tenendo conto dei filtri relativi al questionario ed implementando dei controlli sia "hard checks" (con blocco della compilazione), che "soft checks", con la comparsa di specifici messaggi di errore o di messaggi di "warning". Il questionario è stato testato con una rilevazione pilota realizzata nel 2011, tuttavia data la natura essenzialmente innovativa della rilevazione si è deciso di introdurre, al termine di ogni sezione, un campo aperto per inserire osservazioni o ulteriori informazioni su quell'investimento intangibile. Invece, per le domande della sezione H, sulle misure messe in atto dall'impresa per contrastare gli effetti della crisi economica, il sistema CAWI e CATI è stato sviluppato in modo da far ruotare gli items per non influenzare l'esito delle risposte.

L'utilizzo di una tecnica mista ha consentito di massimizzare l'efficacia della rilevazione, sfruttando al massimo le potenzialità delle due metodologie e garantendo così un alto tasso di risposta del target oggetto di studio. Le imprese che non hanno compilato il questionario, entro la data indicata nella mail iniziale, sono state sollecitate tramite diversi recall inviati via mail e successivamente sollecitate telefonicamente. Al termine della raccolta dati sono pervenuti 10.631 questionari: di questi la quasi totalità

è stata compilata con il sistema CAWI (85% circa), solo una piccola percentuale attraverso CATI (meno del 5%) mentre il 10% circa ha preferito compilare un questionario cartaceo.

Schema 1 – La struttura del questionario di rilevazione on line



Durante la fase di rilevazione sul campo la società incaricata ha attivato un numero verde per fornire un supporto ai partecipanti; gli intervistati potevano chiamare questo numero di telefono per avere chiarimenti e richiedere assistenza sulla rilevazione. Invece, l'ISFOL ha attivato una mail dedicata, per tutto il periodo della rilevazione, per fornire assistenza a tutte le imprese che chiedevano informazioni sia su aspetti formali dell'indagine (obbligo di risposta, titolarità dell'indagine, segreto statistico e tutela della riservatezza, ecc.) sia per specifiche problematiche relative ai diversi assets; in particolare, sono state ricontattate telefonicamente tutte quelle imprese che ponevano problemi di interpretazione delle definizioni o quesiti molto peculiari.

È stato, inoltre, predisposto un sistema per monitorare l'andamento della rilevazione, con una periodicità giornaliera (numero di accessi al sito, numero di registrazioni, numero di questionari compilati parzialmente e numero di questionari completi inviati) e settimanale (analisi delle variabili di interesse) per tenere sotto controllo la tempistica e il rispetto del disegno campionario. Sulla base delle informazioni ricevute sono stati calcolati alcuni indicatori che hanno consentito di mettere in atto adeguati interventi (recall inviati via mail alle imprese registrate, solleciti telefonici dopo un mese dall'invio ufficiale, ecc.).

4. Descrizione del campione intervistato e tasso di risposta

Nell'ambito della rilevazione sugli investimenti intangibili sono state realizzate, attraverso la duplice modalità CAWI-CATI, interviste per un numero pari a 10.631 unità; pertanto, a fronte di un campione di 28.321 unità, il tasso di risposta si è attestato al 37,5%. Le imprese hanno avuto nei confronti della rilevazione un diverso atteggiamento a seconda dell'area geografica in cui esse hanno la sede legale. Infatti,

la partecipazione è stata molto più ampia nel Nord-Ovest e ancor di più nel Nord-Est (42,4%), meno nel Centro e soprattutto nel Sud (28,8%). (Tab. 6)

Tabella 6 - Statistiche riassuntive dell'andamento della rilevazione, per ripartizione geografica. Anno 2012 (v.a. e %)

Ripartizione geografica	Campione		Interviste		
	v.a.	%	v.a.	%	% di rispondenti
Nord-Ovest	10.682	37,7	4.115	38,7	38,5
Nord-Est	7.960	28,1	3.374	31,7	42,4
Centro	5.201	18,4	1.851	17,4	35,6
Sud ed Isole	4.478	15,8	1.291	12,1	28,8
TOTALE	28.321	100,0	10.631	100,0	37,5

Fonte: ISFOL "Rilevazione statistica sugli investimenti intangibili delle imprese (Intangible Assets Survey)"

L'analisi dettagliata per regione (Tab. 7) evidenzia come il più elevato tasso di risposta si sia registrato in Trentino-Alto Adige dove oltre la metà delle imprese contattate ha compilato il questionario; pure la Valle d'Aosta e il Friuli Venezia Giulia fanno osservare tassi di copertura sensibilmente più alti della media nazionale. Al contrario, come anticipato, nell'area meridionale la partecipazione è stata più modesta rispetto a quanto avvenuto a livello complessivo; se poi il confronto viene effettuato con le regioni settentrionali, specie quelle Nord-orientali, il gap diventa davvero rilevante dal momento che ammonta a 13,6 punti percentuali in meno. Particolarmente basso il tasso di risposta in Sicilia e in Calabria.

Tabella 7 - Statistiche riassuntive dell'andamento della rilevazione, per regione geografica. Anno 2012 (v.a. e %)

Regione geografica	Campione		Interviste		
	v.a.	%	v.a.	%	% di rispondenti
Piemonte	2.367	8,4	818	7,7	34,6
Valle d'Aosta	49	0,2	24	0,2	49,0
Lombardia	7.682	27,1	3.040	28,6	39,6
Trentino -Alto Adige	666	2,4	345	3,2	51,8
Veneto	3.613	12,8	1.465	13,8	40,5
Friuli-Venezia Giulia	681	2,4	301	2,8	44,2
Liguria	584	2,1	233	2,2	39,9
Emilia-Romagna	3.000	10,6	1.263	11,9	42,1
Toscana	1.821	6,4	714	6,7	39,2
Umbria	414	1,5	163	1,5	39,4
Marche	1.004	3,5	329	3,1	32,8
Lazio	1.962	6,9	645	6,1	32,9
Abruzzo	517	1,8	177	1,7	34,2
Molise	67	0,2	23	0,2	34,3
Campania	1.339	4,7	387	3,6	28,9
Puglia	931	3,3	273	2,6	29,3
Basilicata	112	0,4	35	0,3	31,3
Calabria	291	1,0	59	0,6	20,0
Sicilia	858	3,0	221	2,1	25,8
Sardegna	363	1,3	116	1,1	32,0
TOTALE	28.321	100,0	10.631	100,0	37,5

Fonte: ISFOL "Rilevazione statistica sugli investimenti intangibili delle imprese (Intangible Assets Survey)"

A livello di classe dimensionale (Tab. 8), si rileva una tendenza all'aumento del tasso di risposta man mano che cresce l'ampiezza aziendale, passando dal 32,4% per le imprese più piccole al 40,2% per quelle della fascia 100-250 addetti e al 49,9% per quelle con 500 ed oltre addetti. Abbastanza contenuta invece la variabilità a livello settoriale giacché industria e servizi fanno registrare il medesimo tasso di risposta (Tab. 9). Analizzando nel dettaglio i singoli comparti, notiamo come lo scostamento dalla media risulti nella maggior parte dei casi piuttosto circoscritto: le uniche eccezioni sono rappresentate, in negativo, dal tessile, dalle attività di trasporto e dalle industrie alimentari (rispettivamente, -3,8, -4,5 e -6 punti percentuali rispetto al dato complessivo); in positivo, dalle imprese finanziarie che hanno compilato il questionario nel 56,3% dei casi.

Tabella 8 - Statistiche riassuntive dell'andamento della rilevazione, per classe dimensionale. Anno 2012 (v.a. e %)

Classe dimensionale	Campione		Interviste		
	v.a.	%	v.a.	%	% di rispondenti
10-49	11.487	40,6	3.717	35,0	32,4
50-99	9.277	32,8	3.697	34,8	39,9
100 -250	5.017	17,7	2.016	19,0	40,2
250-499	1.480	5,2	672	6,3	45,4
500 ed oltre	1.060	3,7	529	5,0	49,9
TOTALE	28.321	100,0	10.631	100,0	37,5

Fonte: ISFOL "Rilevazione statistica sugli investimenti intangibili delle imprese (Intangible Assets Survey)"

Tabella 9 - Statistiche riassuntive dell'andamento della rilevazione, per settore economico. Anno 2012 (v.a. e %)

Settore economico	Campione		Interviste		
	v.a.	%	v.a.	%	% di rispondenti
INDUSTRIA	16.722	59,0	6.278	59,1	37,5
Industrie estrattive	109	0,4	41	0,4	37,6
Industrie alimentari e del tabacco	1.447	5,1	488	4,6	33,7
Tessile e abbigliamento	2.000	7,1	630	5,9	31,5
Industria della carta, cartone, stampa	681	2,4	263	2,5	38,6
Industria chimica, farmaceutica, gomma, plastica	2.571	9,1	1.006	9,5	39,1
Produzione di metalli e prodotti in metallo	2.888	10,2	1.107	10,4	38,3
Apparecchi meccanici, elettrici, elettronici	3.823	13,5	1.524	14,3	39,9
Mezzi di trasporto	594	2,1	223	2,1	37,5
Legno, mobili e altre industrie manifatturiere	1.423	5,0	529	5,0	37,2
Fornitura di elettricità, gas, acqua e gestione rifiuti	1.186	4,2	467	4,4	39,4
SERVIZI	11.599	41,0	4.353	40,9	37,5
Commercio	5.364	18,9	1.964	18,5	36,6
Attività di trasporto, magazzinaggio e posta	3.021	10,7	999	9,4	33,1
Telecomunicazioni, editoria, informatica	1.754	6,2	657	6,2	37,5
Servizi finanziari, assicurazioni e fondi pensione	883	3,1	497	4,7	56,3
Attività professionali, scientifiche, tecniche	577	2,0	236	2,2	40,9
TOTALE	28.321	100,0	10.631	100,0	37,5

Fonte: ISFOL "Rilevazione statistica sugli investimenti intangibili delle imprese (Intangible Assets Survey)"

5. Il calcolo e la calibrazione dei pesi

Il principio su cui è basato ogni metodo di stima campionaria consiste nella condizione che le unità incluse nel campione rappresentino anche le restanti unità della popolazione oggetto di studio. Questo principio viene realizzato attribuendo a ogni unità campionaria un peso, che indica il numero di unità della popolazione rappresentate dall'unità medesima.

Pertanto, all'insieme delle unità rispondenti è stato assegnato un sistema di pesi determinato con la tecnica nota in letteratura con il nome di calibration (Deville, Särndal, 1992; Singh, Mohl, 1996), che costituisce, peraltro, il metodo di stima standard per la maggior parte delle indagini campionarie.

Al fine di rendere più chiara la successiva esposizione, introduciamo la seguente simbologia. Sia

- k indice di impresa
- d indice di dominio di stima ($d = 1, \dots, D$)
- h indice di strato ($h = 1, \dots, H_d$)
- H_d numero di strati inerenti al dominio d
- N_h numero di imprese dello strato h
- U popolazione oggetto di studio di dimensione $N \left(N = \sum_{d=1}^D \sum_{h=1}^{H_d} N_h \right)$
- n_h numero di imprese campione dello strato h
- s campione estratto da U di dimensione $n \left(n = \sum_{d=1}^D \sum_{h=1}^{H_d} n_h \right)$
- y generica variabile di interesse
- y_{hk} valore della variabile y rilevato sull'unità k dello strato h .

Ipotizziamo di voler stimare, con riferimento ad un generico dominio d , il totale della variabile y oggetto di indagine, espresso dalla seguente relazione:

$$y_d = \sum_{h=1}^{H_d} \sum_{k=1}^{N_h} y_{hk} \quad (1)$$

La stima del totale (1) è data da:

$$\hat{y}_d = \sum_{h=1}^{H_d} \hat{y}_h = \sum_{h=1}^{H_d} \sum_{k=1}^{n_h} w_{hk} y_{hk} \quad (2)$$

in cui w_{hk} è il peso finale da attribuire alle n_h aziende dello strato h .

La procedura di calibrazione, in generale, viene utilizzata al fine di: i) correggere l'effetto distorsivo della mancata risposta totale; nonché ii) tenere conto della conoscenza di totali noti di rilevanti variabili ausiliarie (disponibili da fonti esterne all'indagine).

In particolare, la procedura per l'ottenimento dei pesi finali in oggetto è articolata nelle seguenti fasi:

1. calcolo dei pesi base come reciproco della probabilità di inclusione nel campione delle unità costituenti la popolazione d'interesse;

2. costruzione dei fattori correttivi che consentono di soddisfare le condizioni di uguaglianza tra i totali noti delle variabili ausiliarie e le corrispondenti stime campionarie;
3. determinazione dei pesi finali come prodotto dei pesi base, di cui al passo 1, per i fattori correttivi ottenuti al passo 2.

I fattori correttivi, sopra citati, sono ottenuti dalla risoluzione di un problema di minimo vincolato, in cui la funzione da minimizzare è una funzione di distanza (opportunosamente prescelta) tra i pesi base e i pesi finali, mentre i vincoli sono definiti dalla condizione di uguaglianza tra stime campionarie dei totali noti di popolazione e valori noti degli stessi. La funzione di distanza generalmente utilizzata è la funzione logaritmica troncata; l'adozione di tale funzione garantisce che i pesi finali siano positivi e contenuti in un predeterminato intervallo di valori possibili, eliminando in tal modo i pesi positivi estremi (troppo grandi o troppo piccoli).

Un importante stimatore appartenente alla classe dei *calibration estimators*, che si ottiene utilizzando la funzione di distanza euclidea, è lo stimatore di regressione generalizzata. Tale stimatore riveste un ruolo centrale perché è possibile dimostrare che tutti i *calibration estimators*¹ convergono asintoticamente, all'aumentare della numerosità campionaria, allo stimatore di regressione generalizzata (Deville, Särndal, 1992).

Nell'indagine in oggetto, si è deciso di utilizzare come totale noto esterno ai fini della calibrazione il numero di imprese e di addetti a livello settoriale e di dimensione dell'impresa allo scopo di garantire la coerenza tra la popolazione dell'universo ASIA e quella stimata dall'indagine.

Indicando, quindi, con ${}_j X$ ($j=1, \dots, J$) il totale noto della variabile ausiliaria J per la generica regione geografica e con ${}_j X_{hk}$ il valore assunto dalla variabile ausiliaria J sull'impresa rispondente k dello strato h , la condizione sopra descritta è espressa dalla seguente uguaglianza:

$${}_j X = {}_j \hat{X} = \sum_{h=1}^H \sum_{k=1}^{n_h} w_{hk} {}_j X_{hk} \quad (j=1, \dots, J) \quad (3)$$

in cui H indica il numero complessivo di strati definiti nella regione.

¹ È bene sottolineare che i *calibration estimators* sono approssimativamente corretti per n sufficientemente grande.

6. Valutazione del livello di precisione delle stime

Una valutazione dell'attendibilità delle stime prodotte da un'indagine è fornita dalla stima della varianza campionaria. Per il calcolo della varianza di campionamento è stato utilizzato il package R ReGenesees (Zardetto, 2012), sviluppato in Istat, che consente di ottenere le stime della varianza di campionamento.

Indicando con $\hat{V}_{dm}(\hat{RB}_d)$ la stima della varianza di \hat{RB} per il generico dominio di stima d ($d = 1, \dots, D$), la stima dell'errore di campionamento assoluto di \hat{RB} si ottiene mediante la seguente espressione:

$$\hat{\sigma}_{dm}(\hat{RB}_d) = \sqrt{\hat{V}_{dm}(\hat{RB}_d)} \quad (4)$$

A partire dall'errore assoluto è possibile costruire un intervallo di confidenza per ciascuna stima di interesse. Tale intervallo contiene il valore vero del parametro oggetto di stima con un livello di fiducia pari a P ed è ottenuto come:

$$\{\hat{RB}_d - k_P \hat{\sigma}_{dm}(\hat{RB}_d) \leq RB_d \leq \hat{RB}_d + k_P \hat{\sigma}_{dm}(\hat{RB}_d)\} \quad (5)$$

Nella (36) il valore di k_P dipende dal valore fissato per il livello di fiducia P , ad esempio, per $P = 0.95$ si ha $k_{0.95} = 1.96$.

Data la natura del parametro di interesse, definito come differenza relativa, gli errori di campionamento delle stime dell'errore di misura sono presentati in termini di intervalli di confidenza ed è stato scelto come livello di fiducia $P = 95\%$.

Tabella 10 - Coefficienti di variazione (CV %) per i domini di stima

	CV					
	Sez. A	Sez. B	Sez. C	Sez. D	Sez. E	Sez. F
	a1	b1	c1	d1	e1	f1
Ripartizione Geografica						
Nord Ovest	2,90	3,68	3,82	4,77	7,89	4,72
Nord Est	3,17	4,03	4,05	5,00	7,01	5,00
Centro	4,79	5,71	6,01	7,93	10,65	7,23
Sud	7,45	9,40	8,78	13,53	19,30	10,67
Isole	11,66	16,01	14,19	31,15	37,17	22,40
Settore di attività economica						
C=ATTIVITÀ MANIFATTURIERE	2,00	2,80	2,79	3,14	4,91	3,53
D=FORNITURA DI ENERGIA ELET., GAS, VAPORE E ARIA CONDIZ.	4,60	8,21	13,63	22,22	34,04	11,95
E=FORNITURA DI ACQUA; RETI FOGNARIE, ATTIVITÀ DI GESTIONE DEI RIFIUTI E RISANAMENTO	6,93	12,86	14,15	29,30	54,19	16,67
G=COMMERCIO ALL'INGROSSO E AL DETTAGLIO; RIPARAZIONE DI AUTOVEICOLI E MOTOCICLI	3,59	4,70	3,98	10,87	12,00	7,14
H=TRASPORTO E MAGAZZINAGGIO	5,27	9,21	10,41	30,12	50,52	11,94
J=SERVIZI DI INFORMAZIONE E COMUNICAZIONE	4,34	5,17	7,38	8,64	19,44	9,46
K=ATTIVITÀ FINANZIARIE E ASSICURATIVE	2,84	7,66	8,13	23,19	25,39	9,07
M=ATTIVITÀ PROFESSIONALI, SCIENTIFICHE E TECNICHE	6,62	8,76	11,58	15,50	37,09	13,89
Classe Addetti						
10 - 49 addetti	1,86	2,62	2,49	3,92	5,51	3,58
50 - 99 addetti	1,45	2,07	2,23	2,47	4,23	2,63
100 - 249 addetti	2,15	2,68	2,98	3,26	5,42	3,38
250 ed oltre addetti	2,19	2,74	3,13	3,45	5,56	3,36
Italia	1,47	2,05	2,05	2,79	4,33	2,73

Fonte: ISFOL "Rilevazione statistica sugli investimenti intangibili delle imprese (Intangible Assets Survey)"

7. Le procedure di controllo e correzione dei dati

La fase dell'identificazione e trattamento degli errori di misura e delle mancate risposte è stata progettata in modo da ottenere un insieme di dati completo (senza mancate risposte) e privo di incoerenze, per assicurare una elevata qualità della base dati. A tal fine, si è tenuto conto di alcuni standard sviluppati a livello nazionale e internazionale, in particolare sono state utilizzate un insieme di tecniche e metodologie statistiche consolidate in esperienze su dati simili (CVTS - *Continuing Vocational Training Survey*, CIS - *Community Innovation Survey* e R&S - *Ricerca e Sviluppo*).

Le attività di controllo delle informazioni fornite dalle aziende rispondenti si sono concentrate principalmente sulle mancate risposte totali e parziali, sugli errori di misura, oltre che sulla coerenza interna delle informazioni.

Un'operazione preliminare alla fase di controllo è stata quella di codificare le domande aperte, inserite a chiusura di ogni sezione per consentire ai rispondenti di "aggiungere qualche osservazione o ulteriori informazioni". Sostanzialmente, sono state fornite due tipologie di risposte: un piccolo gruppo di imprese ha indicato il dettaglio di alcune voci di spesa prese in considerazione per calcolare la "spesa totale" sostenuta per quel determinato investimento mentre la maggior parte dei rispondenti ha dato spiegazioni sulla mancata quantificazione della spesa per quell'investimento. Ciò ha consentito innanzitutto di ridurre le mancate risposte parziali; di individuare determinati parametri per correggere e/o imputare i dati in modo coerente e uniforme. Inoltre, i rispondenti hanno fornito preziose informazioni da utilizzare per future rilevazioni ma anche per l'analisi dei dati.

Mancate risposte totali

Una mancata risposta totale (MRT) ha luogo quando, pur avendo accertato che un'impresa del campione fa parte del campo di osservazione, non è stato possibile rilevare alcuna informazione sull'impresa a causa, ad esempio, del rifiuto opposto da parte dell'impresa contattata. Questo tipo di mancata risposta può determinare tuttavia, se non opportunamente trattata, una distorsione nella stime. In fase di costruzione dei pesi di estensione del campione all'universo, per correggere la mancata risposta totale si ricorre a metodi di aggiustamento, basati sulla riponderazione che, utilizzando le informazioni ausiliarie disponibili, minimizzano gli effetti distorsivi. A conclusione della fase di rilevazione, la prima operazione effettuata sulla base dati è stata quella di determinare il numero esatto dei rispondenti. Sono state adottate tutte le misure atte a ridurre le mancate risposte parziali o totali cercando, nei limiti del possibile, di utilizzare altre fonti di dati, comunque, quando un record conteneva meno del 50% delle variabili chiave presenti, è stato considerato come una "mancata risposta totale"²; per questo motivo sono stati eliminati 74 record, pari allo 0,69% del totale. Quindi, i questionari validi ricevuti sono stati 10.631.

Mancate risposte parziali

La mancata risposta parziale (MRP) - che ha luogo quando l'informazione relativa al rispondente è incompleta - è stata trattata mediante imputazione. In questo caso gli aspetti più rilevanti da considerare sono:

1. individuazione delle variabili osservate maggiormente correlate alle variabili non osservate (missing);
2. scelta di un approccio parametrico, non parametrico, o misto, per l'utilizzo delle informazioni disponibili ai fini della ricostruzione dell'informazione mancante;
3. *tuning* delle tecniche di imputazione.

Ai fini del primo aspetto, è opportuno svolgere preliminarmente un'analisi esplorativa delle relazioni di interdipendenza tra le variabili di interesse. I risultati di tale analisi verranno utilizzati per la determinazione di strati all'interno dei quali applicare separatamente le procedure di imputazione (celle d'imputazione), o

² È stato applicato uno dei criteri utilizzati in CVTS. Si veda il Regolamento (CE) n.1552/2005 del Parlamento europeo e Consiglio relativo alle statistiche sulla formazione professionale nelle imprese.

per l'individuazione delle variabili da utilizzare come variabili esplicative nei metodi di imputazione sia parametrici che non parametrici. La determinazione delle celle di imputazione e delle covariate deve inoltre cercare di rendere plausibile l'ipotesi di "missing at random" su cui si basano tutte le metodologie impiegate.

Di particolare importanza è la scelta tra approccio parametrico o non parametrico nella adozione delle tecniche di imputazione. I valori mancanti vengono sostituiti con valori "plausibili" e ciò consente di applicare le analisi statistiche all'intero dataset. I vantaggi naturalmente sono la praticità e semplicità con cui, in una fase successiva, possono essere condotte le analisi. Gli svantaggi riguardano l'introduzione di una ulteriore fonte di variabilità da tenere in considerazione nell'analisi dei dati. E' evidente che la scelta della tecnica di imputazione deve essere effettuata sulla base della natura delle variabili da trattare. In particolare, l'utilizzo di metodi basati su regressione lineare classica, o tramite medie, potrebbe esser appropriato per variabili "strutturali" a carattere continuo come i costi sostenuti dell'impresa per le attività di formazione, ricerca e sviluppo.

La scelta dell'approccio e del particolare metodo di imputazione è stata effettuata sulla base di studi di simulazione effettuati sui dati completi. In altre parole, sono state introdotte artificialmente delle mancate risposte parziali su alcuni record, poi imputate con i metodi da testare. La valutazione è stata effettuata mediante confronto tra stime ottenute su dati completi e stime ottenute su dati imputati.

Da un'analisi preliminare, l'approccio più efficiente è risultato essere l'imputazione delle medie per cella³. Le celle sono state costruite stratificando per settore di attività economica (30 settori), classe di addetti, ripartizione geografica, e si è tenuto conto anche dell'attività svolta dall'azienda. Il criterio di costruzione degli strati ha seguito l'obiettivo di ottenere degli strati massimamente omogenei al loro interno e massimamente diversi fra di loro. Supposto che la probabilità che un valore sia mancante è indipendente dai valori osservati e dagli altri valori mancanti, chiamiamo la mancata risposta MCAR. Introduciamo J strati, chiamiamo con \bar{y}_{jR} la media dei rispondenti per la variabile Y nello strato j. La stima della media di Y diventa:

$$\frac{1}{n} \sum_{j=1}^J \left(\sum_{i=1}^{r_j} j_{ij} + \sum_{i=r_j+1}^{n_j} \bar{y}_{jR} \right) = \frac{1}{n} \sum_{j=1}^J n_j \bar{y}_{jR} = \bar{y}_{wc}$$

Analogamente allo stimatore nel campionamento stratificato, che tale imputazione richiama, lo stimatore è corretto nel caso in cui all'interno degli strati ci sia una mancata risposta MCAR.

Errori di misura

Gli errori di misura si classificano tra gli errori non campionari. L'errore non campionario è dovuto ad errori o lacune che si verificano durante il disegno e la realizzazione delle diverse fasi del processo d'indagine (definizioni, liste, rispondenti, intervistatori, ecc.).

Sono di natura sistematica gli errori di misura che si manifestano in modo consistente nei dati a causa generalmente di lacune concettuali o procedurali in una o più fasi del processo di raccolta, registrazione, elaborazione dei dati; sono invece di natura casuale gli errori che si manifestano nei dati in modo casuale, dovuti generalmente a cause accidentali nel corso del processo di indagine cioè che non hanno la tendenza a modificare le stime in modo consistente in nessuna precisa direzione.

³ L'imputazione per media nelle celle porta a risultati simili a quelli che si otterrebbero riponderando per mancata risposta all'interno della cella (ammesso che i pesi siano costanti all'interno della cella stessa). Adottando l'imputazione tramite regressione, si attribuisce il valore di regressione. L'imputazione per media è un caso particolare della regressione, utilizzando come variabili esplicative le variabili dummy che identificano gli strati di imputazione. Al contrario, per variabili a carattere semi-continuo, che presentano forti concentrazioni su alcuni valori specifici (tipicamente zero), una strategia parametrica è più problematica, mentre l'adozione di tecniche hot-deck (e.g. donatore di minima distanza) sembra essere più appropriata. Nell'imputazione hot-deck si sostituisce un valore preso da un rispondente "simile". Questo è il metodo più utilizzato, ma le proprietà statistiche rimangono ancora in gran parte da investigare. Per alcune variabili è possibile utilizzare la tecnica di imputazione cold-deck, con la quale si sostituisce un valore preso da una osservazione di un'altra indagine.

L'individuazione e trattamento degli errori di misura si è articolato in tre sotto-fasi.

1. Individuazione e correzione degli errori sistematici;

Gli errori sistematici sono errori riconducibili ad un meccanismo deterministico interpretabile che interviene su una porzione consistente di unità. E' evidente che tali errori possono avere un forte effetto distorsivo sulle stime di interesse. Tipici errori sistematici nel caso di variabili quantitative sono gli errori di unità di misura.

In molti casi la correzione degli errori sistematici è banale, in quanto si riconduce ad un'azione deterministica che ripristina il valore vero sulla base di quello errato (ad es. nel caso dell'errore di unità di misura). Il problema tuttavia consiste nell'individuare le unità affette da errore nel caso si sia appurata l'esistenza di una sorgente sistematica di errore.

In questo caso sono state impiegate tecniche di analisi esplorativa dei dati e/o analisi delle frequenze di violazione di opportune regole di compatibilità tra le variabili: l'errore che si è verificato più frequentemente è quello relativo ai costi indicati nelle diverse sezioni del questionario, in particolare sono stati aggiunti sistematicamente due zeri in più al valore 'vero' che si voleva indicare.

2. individuazione e trattamento dei valori anomali ed errori influenti;

I valori anomali sono valori/osservazioni che differiscono significativamente dal modello implicito o esplicito, univariato o multivariato assunto per i dati, e che sono influenti sulla stima obiettivo (la cui inclusione o esclusione nel calcolo della stima di interesse ha effetti statisticamente rilevanti sulla stima stessa).

L'individuazione degli errori influenti e degli outlier si avvale, da una parte della modellizzazione delle distribuzioni congiunte delle variabili di interesse, dall'altro dell'impiego di tecniche di editing selettivo, che tengano conto anche del peso campionario dell'unità affetta da errore. Ove possibile, le unità classificate come potenzialmente affette da errori influenti sono state trattate interattivamente, ove necessario è stato effettuato un controllo manuale del record e in alcuni casi, relativamente alle grandi aziende è stato effettuato un recall telefonico.

E' importante verificare l'origine del valore anomalo che può essere attribuita alla variabilità naturale del fenomeno, cioè il valore è anomalo perché il fenomeno ha una variabilità maggiore di quella attesa o, più in generale, ha una distribuzione diversa da quella ipotizzata; oppure può essere un errore di misura o di processo, cioè errori di interpretazione, errori di misurazione commessi in fase di rilevazione e/o di registrazione.

Grazie alle altre fonti disponibili di dati (CIS, CVTS e Ricerca & Sviluppo), sui fenomeni oggetto di studio della presente indagine, si sono costruite delle soglie di valori accettabili al di sopra delle quali sono stati controllati tutti i record. Se non era verificata la coerenza all'interno del record allora si è confermata l'ipotesi della presenza di valore anomalo e si è proceduto alla correzione del valore. Inoltre, sono stati realizzati altri controlli – attraverso la costruzione di alcuni indicatori chiave (es. ore di formazione per formato oppure costo per partecipante, ecc.) – per far emergere le incoerenze tra le informazioni riguardanti la stessa unità rispondente.

3. localizzazione e correzione di errori casuali non influenti.

Gli errori residui rispetto alle due tipologie precedentemente menzionate sono trattate con tecniche di localizzazione (e imputazione) automatica tipo Fellegi-Holt⁴.

⁴ Tre sono i criteri fondamentali per l'imputazione delle variabili qualitative alla base della metodologia di Fellegi e Holt (1976):

- in ogni record i dati devono soddisfare tutte le regole di validità e incompatibilità, cambiando il meno possibile il valore dei campi;
- le regole di imputazione devono essere derivate dalle regole di controllo, senza esplicita specificazione;
- le distribuzioni di frequenza marginali e congiunte devono essere mantenute il più possibile.