

Report File standard “Imprese 2012-2021”

Lo scopo di questo documento è quello di presentare la costruzione del file standard di “Imprese 2012-2021”, un dataset di microdati a livello di impresa¹, che, coerentemente alle linee guida di istituto per il 2022, sarà messo a disposizione dei ricercatori esterni all’istituto.

Il documento si compone di diverse parti. Nella prima parte si mostra in modo dettagliato come il campionamento di impresa sia effettivamente casuale. Partendo dall’archivio delle anagrafiche di impresa, si effettua una randomizzazione che comprende il 15% delle matricole contributive, e si verifica che, lungo diverse dimensioni (territorio, settore, età dell’impresa, ecc.), l’incidenza osservata nel campione sia la stessa di quella osservata nella popolazione.

Successivamente, si descrivono le variabili che sono presenti nel file standard, che provengono sia dalla anagrafica di impresa che dagli archivi dei rapporti di lavoro annuali, dal 2012 al 2021. Sono state selezionate quelle variabili che si ritiene potrebbero essere di maggiore interesse da parte dei ricercatori accademici ed istituzionali.

Nell’ultima parte del documento si descrive come è stato affrontato il problema di variabili che cambiano definizione nel tempo, come la natura giuridica e la qualifica inerente alla professione, al fine di derivare variabili sintetiche e omogenee dal punto di vista temporale.

1. Procedura per il campionamento casuale e relativa verifica empirica

La base dei dati utilizzata come punto di partenza per la costruzione del file standard è la versione più recente del database anagrafico delle imprese e delle relative posizioni contributive, disponibile presso la DCSR (Anagrafica_aziende_v7). Questo archivio riporta una serie di informazioni relative alle aziende italiane censite per matricola contributiva (anonimizzata). Si è scelto di utilizzare questa variabile anonimizzata come base per la selezione del campione poiché, tra le varie opzioni disponibili, la matricola contributiva anonimizzata (opportunamente trattata) si presenta come una variabile costituita anche da elementi casuali e quindi esogeni rispetto alle caratteristiche dell’azienda.

Per la randomizzazione sono stati estratti due numeri dalla stringa anonimizzata del codice fiscale/partita iva della matricola contributiva, ed è stata generata una nuova variabile con cento possibili valori (cioè le possibili combinazioni dei due numeri estratti). Una volta verificata la distribuzione uniforme delle diverse modalità ne sono state selezionate casualmente 15. Pertanto, alle matricole aziendali caratterizzate dall’aver una delle 15 modalità selezionate è stato associato il valore 1 della variabile binaria che indica, per ogni singola osservazione, l’inclusione/esclusione dal sample dello specifico file standard.

Il dataset così creato derivante dall’anagrafica aziende è stato unito ai dati sui rapporti di lavoro Uniemens, dai quali i dati individuali sono stati ‘collassati’ a livello di matricola contributiva. Si è quindi proceduto a valutare la capacità del campione di descrivere l’universo di riferimento analizzando la distribuzione di alcune

¹ Si noti che nel dataset l’unità di osservazione è la matricola contributiva e non il codice fiscale di impresa e che per una quota limitata di imprese le due definizioni non coincidono, in quanto vi possono essere imprese con più matricole contributive.

variabili di particolare rilevanza. In particolare, si è calcolata l'incidenza delle osservazioni per modalità/classi di valori e si è proceduto ad un confronto tra i risultati nell'universo e nel campione.

Campionamento da Anagrafica_aziende_v7

Estrazione caratteri matricola
contributiva da "Anagrafica aziende"

controllo uniforme distribuzione delle
modalità

selezione del 15% delle modalità e
creazione variabile binaria (1 in 0 out)

Inclusione nel campione delle aziende
con le modalità scelte

Matching con Aziende Annuali 2012-2021

Numero di osservazioni nel Db "AZIENDE
ANNUALI": 17.085.240

Matching con risultati campionamento.
Numero di osservazioni: 2.564.961
(15.01% della popolazione complessiva)

controllo bilanciamento (confronto
sample - popolazione)

1.1 Caratteristiche del campione

Considerando il periodo 2012-2021, nei dataset rapporti di lavoro annuali sono contenute 17.085.240 osservazioni, di cui 2.564.961 (il 15%) corrispondono al campione estratto casualmente dal dataset sulle anagrafiche aziende e andranno a far parte del file standard. La Tabella 1 mostra il numero di osservazioni e il dettaglio di incidenza sul totale per singolo anno di riferimento. Dal confronto si evince che campione e popolazione sono sostanzialmente sovrapponibili in termini di incidenza annuale delle osservazioni sul totale.

Tabella 1: Incidenza annuale nel periodo (2012-2021) del campione e della popolazione.

ANNO	Popolazione		Campione	
	No.	%	No.	%
2012	1808578	10.59	272059	10.61
2013	1742401	10.2	261938	10.21
2014	1689315	9.89	253786	9.89
2015	1711503	10.02	256684	10.01
2016	1684671	9.86	252495	9.84
2017	1696907	9.93	254502	9.92
2018	1698764	9.94	254841	9.94
2019	1706009	9.99	256111	9.98
2020	1656647	9.7	248791	9.7
2021	1690445	9.89	253754	9.89
Totale	17085240	100	2564961	100

1.2 Classi dimensionali

La struttura del tessuto produttivo italiano è caratterizzata da una cospicua presenza di aziende di piccole dimensioni. Come si evince dalla Tabella 2, il sample estratto è in grado di rilevare accuratamente questa peculiarità: campione e popolazione sono del tutto simili in termini di incidenza delle aziende per classe dimensionale. La distribuzione risulta essere adeguata anche analizzando la caratteristica del campione per ogni singolo anno di riferimento (Tabella 3).

Tabella 2: Consistenza e incidenza sul totale, per classi dimensionali (periodo 2012-2021)

classidip	Campione		Popolazione	
	No.	%	No.	%
fino a 2	7501305	43.91	1126888	43.93
[3, 5[3153616	18.46	472593	18.42
[5, 10[3181680	18.62	477650	18.62
[10, 15[1247812	7.3	187079	7.29
[15, 50[1561240	9.14	235122	9.17
50 e oltre	439587	2.57	65629	2.56
Totale	17085240	100	2564961	100

Tabella 3: Incidenza per anno e classi dimensionali (periodo 2012-2021)

ANNO	Popolazione						Campione					
	fino a 2	[3, 5[[5, 10[[10, 15[[15, 50[50 e oltre	fino a 2	[3, 5[[5, 10[[10, 15[[15, 50[50 e oltre
	%	%	%	%	%	%	%	%	%	%	%	%
2012	43.81	18.83	18.97	7.29	8.8	2.3	43.67	18.92	18.98	7.29	8.84	2.31
2013	44.92	18.89	18.58	7.03	8.31	2.28	44.94	18.83	18.54	7.1	8.31	2.27
2014	45.52	18.75	18.24	6.89	8.27	2.32	45.42	18.78	18.3	6.91	8.29	2.3
2015	44.74	18.63	18.43	7.03	8.7	2.48	44.69	18.65	18.42	7.02	8.75	2.47
2016	45.39	18.51	18.2	6.94	8.52	2.44	45.44	18.47	18.13	6.91	8.62	2.43
2017	44.19	18.35	18.41	7.24	9.21	2.6	44.35	18.22	18.33	7.24	9.26	2.59
2018	43.29	18.14	18.58	7.46	9.74	2.78	43.41	18.06	18.58	7.41	9.77	2.76
2019	41.95	18.06	18.89	7.8	10.3	2.98	41.99	18.08	18.93	7.72	10.3	2.96
2020	42.97	18.28	18.91	7.6	9.53	2.71	43.11	18.13	18.94	7.58	9.56	2.68
2021	42.23	18.1	19.01	7.76	10.0	2.87	42.29	18.05	19.05	7.77	9.98	2.86
Totale	43.91	18.46	18.62	7.3	9.14	2.57	43.93	18.42	18.62	7.29	9.17	2.56

1.3 Incidenza di CF aziendali con più matricole contributive

È possibile che una azienda svolga diverse attività e abbia pertanto più posizioni contributive legate allo stesso codice fiscale. Si è ritenuto opportuno calcolare l'incidenza di questa casistica e valutarne l'esatta rappresentazione nel sample. La Tabella 4 mostra l'esito soddisfacente della verifica. Come si evince dalla Tabella 5 non ci sono sostanziali deviazioni dalla struttura della popolazione nemmeno analizzando dettagliatamente i singoli anni.

Tabella 4: Codici fiscali per numero di matricole contributive associate (periodo 2012-2021)

Periodo 2012-2021 Tipologia CF	Popolazione		Campione	
	No.	%	No.	%
Una matricola	15911542	93.13	2388402	93.12
Più di una matricola	1173698	6.87	176559	6.88
Totale	17085240	100	2564961	100

Tabella 5: Incidenza CF per numero di matricole associate e anno (periodo 2012-2021)

ANNO	Popolazione		Campione	
	Una matricola %	Più di una matricola %	Una matricola %	Più di una matricola %
2012	91.12	8.88	91.15	8.85
2013	91.79	8.21	91.79	8.21
2014	92.25	7.75	92.21	7.79
2015	93.09	6.91	93.07	6.93
2016	93.76	6.24	93.75	6.25
2017	93.82	6.18	93.79	6.21
2018	93.87	6.13	93.87	6.13
2019	93.84	6.16	93.81	6.19
2020	94.06	5.94	94.05	5.95
2021	93.90	6.10	93.86	6.14
Totale	93.13	6.87	93.12	6.88

1.4 Attività economica (ATECO 2007)²

Le Tabella 6 illustra il confronto tra popolazione e campione in base al settore economico di appartenenza della matricola contributiva, mostrando la buona qualità del campione in tal senso. Nella tabella si utilizza la classificazione Ateco a sottosezioni (21 modalità).

Tabella 6: Consistenza e incidenza per settore di attività economica (Ateco 2007; periodo 2012-2021)

Periodo 2012-2021	Popolazione		Campione	
	No.	%	No.	%
ATECO 2007				
Agricoltura	169438	0.99	24832	0.97
Estrazione	22180	0.13	3404	0.13
Manifattura	2750996	16.1	413898	16.14
Energia	19775	0.12	3050	0.12
Acqua	60514	0.35	8986	0.35
Costruzioni	2002958	11.72	300958	11.73
Commercio	3848698	22.53	577741	22.52
Trasporto	614962	3.6	93122	3.63
Alloggio/ristorazione	2295392	13.43	344808	13.44
ITC	353401	2.07	53151	2.07
Finanza/assic.	237823	1.39	36011	1.4
Immobiliare	220540	1.29	33276	1.3
Att. Professionali/Scientifiche	1123497	6.58	168328	6.56
Noleggio/ag.viaggio	797371	4.67	119560	4.66
PA	11170	0.07	1657	0.06
Istruzione	199199	1.17	29274	1.14
Sanità	780273	4.57	115815	4.52
Arte-Sport-Intratt.	163034	0.95	24104	0.94
Altri Servizi	1067873	6.25	160244	6.25
Servizi alla famiglia	327270	1.92	49918	1.95
ONG	5132	0.03	827	0.03
Totale	17085240	100	2564961	100

² Per l'elevato numero di celle si è preferito non includere in questo documento le tabelle di confronto tra popolazione e campione dettagliato per anno, tanto per le categorie di attività economica quanto per la distribuzione regionale nel paragrafo successivo. Tali informazioni sono comunque disponibili su richiesta.

1.5 Distribuzione geografica

Analizzando le caratteristiche del campione dal punto di vista geografico, si rileva che non ci sono significativi scostamenti tra la distribuzione delle osservazioni nel campione e quanto rilevato nella popolazione: l'incidenza per area misurata tanto a livello regionale (Tabella 7) quanto al livello provinciale (Tabella 8) indica ancora una volta l'affidabilità del campione estratto.

Tabella 7: Incidenza a livello regionale (periodo 2012-2021)

Periodo 2012-2021	Popolazione		Campione	
	Regione	No.	%	No.
ABRUZZO	403894	2.36	59894	2.34
BASILICATA	146340	0.86	21357	0.83
CALABRIA	453703	2.66	67566	2.63
CAMPANIA	1477617	8.65	220358	8.59
EMILIA-ROMAGNA	1324146	7.75	198971	7.76
FRIULI-VENEZIA GIULIA	318812	1.87	47976	1.87
LAZIO	1703409	9.97	257671	10.05
LIGURIA	480695	2.81	72670	2.83
LOMBARDIA	2934403	17.18	442371	17.25
MARCHE	495392	2.9	73527	2.87
MOLISE	84712	0.5	12681	0.49
PIEMONTE	1136960	6.65	170789	6.66
PUGLIA	1092124	6.39	163161	6.36
SARDEGNA	428079	2.51	63827	2.49
SICILIA	1206628	7.06	181094	7.06
TOSCANA	1237230	7.24	186231	7.26
TRENTINO-ALTO ADIGE	365200	2.14	55387	2.16
UMBRIA	260573	1.53	39687	1.55
VALLE D'AOSTA	46108	0.27	6645	0.26
VENETO	1489215	8.72	223098	8.7
Totale	17085240	100	2564961	100

Tabella 8: Incidenza a livello provinciale. Confronto popolazione e campione (periodo 2012-2021)

Periodo 2012-2021	Popolazione		Campione		Popolazione		Campione		
	Provincia sede	No.	%	No.	%	Provincia sede	No.	%	No.
AG	91393	0.53	13809	0.54	MO	212195	1.24	32179	1.25
AL	113822	0.67	17037	0.66	MS	64366	0.38	9239	0.36
AN	138864	0.81	20903	0.81	MT	52701	0.31	7434	0.29
AO	45997	0.27	6616	0.26	NA	754127	4.41	111811	4.36
AP	70804	0.41	10181	0.4	NO	92394	0.54	13607	0.53
AQ	88791	0.52	13226	0.52	NU	53112	0.31	7990	0.31
AR	103778	0.61	15797	0.62	OR	37384	0.22	5685	0.22
AT	52418	0.31	7594	0.3	PA	272711	1.6	41039	1.6
AV	109362	0.64	16403	0.64	PC	82188	0.48	12197	0.48
BA	361277	2.11	54256	2.12	PD	289008	1.69	43388	1.69
BG	307660	1.8	46356	1.81	PE	98365	0.58	14291	0.56

BI	42498	0.25	6673	0.26	PG	197486	1.16	30116	1.17
BL	57307	0.34	8631	0.34	PI	123012	0.72	18396	0.72
BN	76442	0.45	11688	0.46	PN	79201	0.46	11596	0.45
BO	295011	1.73	43567	1.7	PO	123996	0.73	18535	0.72
BR	102948	0.6	15488	0.6	PR	133836	0.78	19916	0.78
BS	392867	2.3	59849	2.33	PT	85445	0.5	12414	0.48
BT	111854	0.65	16692	0.65	PU	118190	0.69	17357	0.68
BZ	191455	1.12	29151	1.14	PV	124728	0.73	18593	0.72
CA	127346	0.75	18621	0.73	PZ	93058	0.54	13867	0.54
CB	59453	0.35	9034	0.35	RA	111053	0.65	16372	0.64
CE	220376	1.29	33482	1.31	RC	119904	0.7	17732	0.69
CH	110961	0.65	16541	0.64	RE	147696	0.86	22809	0.89
CL	58648	0.34	8869	0.35	RG	94195	0.55	14071	0.55
CN	157543	0.92	23428	0.91	RI	36863	0.22	5757	0.22
CO	161654	0.95	23847	0.93	RM	1276185	7.47	192192	7.49
CR	83467	0.49	12813	0.5	RN	136478	0.8	20649	0.81
CS	168221	0.98	24837	0.97	RO	68466	0.4	10365	0.4
CT	258576	1.51	38495	1.5	SA	316655	1.85	46781	1.82
CZ	86483	0.51	12973	0.51	SI	82769	0.48	12707	0.5
EE	169	0	51	0	SO	59037	0.35	9154	0.36
EN	36092	0.21	5320	0.21	SP	69957	0.41	10448	0.41
FC	118327	0.69	17898	0.7	SR	95532	0.56	14722	0.57
FE	88002	0.52	13431	0.52	SS	162774	0.95	24362	0.95
FG	153070	0.9	23093	0.9	SU	46669	0.27	7041	0.27
FI	336639	1.97	51321	2	SV	93274	0.55	14163	0.55
FM	63560	0.37	9372	0.37	TA	132152	0.77	19163	0.75
FR	136249	0.8	20723	0.81	TE	105536	0.62	15882	0.62
GE	248335	1.45	37429	1.46	TN	173333	1.01	26195	1.02
GO	35671	0.21	5197	0.2	TO	588433	3.44	89428	3.49
GR	73741	0.43	10780	0.42	TP	119432	0.7	17407	0.68
IM	68912	0.4	10649	0.42	TR	62784	0.37	9467	0.37
IS	25149	0.15	3677	0.14	TS	58318	0.34	8703	0.34
KR	38585	0.23	5695	0.22	TV	259320	1.52	38716	1.51
LC	88941	0.52	13300	0.52	UD	146013	0.85	22583	0.88
LE	230776	1.35	34532	1.35	VA	230003	1.35	34905	1.36
LI	108474	0.63	16788	0.65	VB	46389	0.27	6914	0.27
LO	47349	0.28	7023	0.27	VC	42508	0.25	6041	0.24
LT	172348	1.01	26486	1.03	VE	270191	1.58	40514	1.58
LU	134809	0.79	20125	0.78	VI	261592	1.53	39044	1.52
MB	228222	1.34	34153	1.33	VR	283611	1.66	42438	1.65
MC	103721	0.61	15649	0.61	VT	82652	0.48	12675	0.49
ME	179604	1.05	27244	1.06	VV	39658	0.23	6199	0.24
MI	1107225	6.48	167004	6.51	Totale	17085240	100	2564961	100
MN	107048	0.63	15915	0.62					

1.6 Ulteriore verifica di bilanciamento: la regressione multivariata

Come ulteriore strumento di controllo delle caratteristiche del campione estratto si è proceduto a effettuare una regressione di bilanciamento: si tratta di una regressione lineare in cui la variabile dipendente è la variabile binaria che “decide” per ogni osservazione la sua inclusione o esclusione dal campione; tra i regressori sono incluse invece le variabili di cui si vuole valutare l’impatto nel determinare, in media, una maggiore o minore probabilità di essere parte del campione per ogni singola osservazione. Un coefficiente non statisticamente significativo indica un impatto non diverso da zero della variabile di interesse nel guidare l’inclusione o meno nel campione.

Le tabelle seguenti riportano le differenze normalizzate per alcune covariate (Tabella 9) e i coefficienti di regressione (Tabella 10) per le varie modalità di ogni variabile, con il livello di significatività per i due periodi oggetto di analisi.

Come si noterà, le differenze normalizzate rimangono ben al di sotto del livello di guardia di 0.25 come suggerito in letteratura³; per quanto concerne la regressione, anche nei casi in cui il coefficiente risulti statisticamente significativo, l’impatto della variabile nel determinare l’appartenenza o meno al campione è sostanzialmente trascurabile (cioè, il coefficiente ha un valore prossimo allo zero), e la significatività statistica è essenzialmente dovuta all’elevata numerosità di osservazioni nel modello di regressione.

³ La differenza normalizzata riportata in tabella è, per ogni covariata, la differenza tra le medie nel campione e nella popolazione, rapportata alla radice quadrata della somma delle deviazioni standard nei due gruppi. Cfr. Imbens, G. W. and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of economic literature*, 47(1):5–86. La Tabella include solo una selezione delle covariate utilizzate; si segnala che il valore più elevato tra tutti i regressori è pari a 0.00243 per il periodo 1983-2004 e 0.0022 per il periodo 2005-2019.

Tabella 9: Statistiche descrittive e differenze normalizzate (periodo 2012-2021).

Ateco 2007	Normalized difference	Regione	Normalized difference
Agricoltura	-0.00075	ABRUZZO	-0.00135
Estrazione	-0.00170	BASILICATA	-0.00185
Manifattura	0.00056	CALABRIA	-0.00094
Energia	0.00067	CAMPANIA	-0.00145
Acqua	0.00065	EMILIA-ROMAGNA	0.00019
Costruzioni	-0.00046	FRIULI-VENEZIA GIULIA	0.00023
Commercio	0.00022	LAZIO	0.00178
Trasporto	-0.00005	LIGURIA	0.00084
Alloggio/ristorazione	0.00118	LOMBARDIA	0.00134
ITC	0.00016	MARCHE	-0.00139
Finanza/assic.	0.00018	MOLISE	-0.00014
Immobiliare	0.00072	PIEMONTE	0.00011
Att. Professionali/Scientifiche	0.00041	PUGLIA	-0.00090
Noleggio/ag. viaggio	-0.00038	SARDEGNA	-0.00078
PA	-0.00020	SICILIA	-0.00006
Istruzione	-0.00022	TOSCANA	0.00052
Sanità	-0.00163	TRENTINO-ALTO ADIGE	0.00107
Arte-Sport-Intratt.	-0.00176	UMBRIA	0.00127
Altri Servizi	-0.00106	VALLE D'AOSTA	-0.00149
Servizi alla famiglia	-0.00009	VENETO	-0.00046
ONG	0.00157		

Tabella 10: Regressione di bilanciamento. Variabile dipendente "File standard" (periodo 2012-2021).

Periodo 2012-2021. Variabile dipendente: File standard (0/1)			
Variabile	Coefficiente	Variabile	Coefficiente
2013.anno	-0.000097	BASILICATA	-0.00257**
2014.anno	-0.000186	CALABRIA	0.000676
2015.anno	-0.000429	CAMPANIA	0.000756
2016.anno	-0.000521	EMILIA-ROMAGNA	0.00196***
2017.anno	-0.000417	FRIULI-VENEZIA GIULIA	0.00223***
2018.anno	-0.000378	LAZIO	0.00293***
2019.anno	-0.000261	LIGURIA	0.00285***
2020.anno	-0.000199	LOMBARDIA	0.00235***
2021.anno	-0.000265	MARCHE	0.000122
1.multimatricola	0.000414	MOLISE	0.00136
Dipendenti [3, 5[-0.000389	PIEMONTE	0.00185***
Dipendenti [5, 10[-0.000185	PUGLIA	0.00105
Dipendenti [10, 15[-0.000430	SARDEGNA	0.000800
Dipendenti [15, 50[0.000205	SICILIA	0.00187***
Dipendenti 50 e oltre	-0.00118**	TOSCANA	0.00218***
Estrazione	0.00686***	TRENTINO-ALTO ADIGE	0.00331***
Manifattura	0.00387***	UMBRIA	0.00396***
Energia	0.00752***	VALLE D'AOSTA	-0.00428**
Acqua	0.00204	VENETO	0.00148**
Costruzioni	0.00369***	Costante	0.145***
Commercio	0.00352***		
Trasporto	0.00485***		
Alloggio/ristorazione	0.00360***		
ITC	0.00366***		
Finanza/assic.	0.00472***		
Immobiliare	0.00400***		
Att.Professionali/Scientifiche	0.00307***		
Noleggio/ag.viaggio	0.00324***		
PA	0.00169		
Istruzione	0.000318		
Sanità	0.00179*		
Arte-Sport-Intratt.	0.00129		
Altri Servizi	0.00338***		
Servizi alla famiglia	0.00546***		
ONG	0.0138***		

Osservazioni 17,071,496

R-squared 0.000

*** p<0.01, ** p<0.05, * p<0.1

2. Descrizione e derivazione delle variabili presenti nel file standard

Le variabili contenute nel file standard provengono dai due archivi principali: l'archivio delle anagrafiche delle matricole contributive (attualmente nella versione più recente, la v7, presente sul server della DCSR); l'archivio dei rapporti di lavori annuali per i lavoratori dipendenti privati extra agricoli, con particolare riferimento ai file "visitinps_rapporti_lavoro_annuali_xxx.dta" per il periodo 2012-2021, presenti nella banca dati della DCSR. Tali dati individuati vengono 'collassati' a livello di matricola contributiva, e una volta derivato un dataset a livello di matricola contributiva viene realizzata la metodologia di campionamento presentata nel paragrafo precedente, che considera circa il 15% delle matricole contributive derivante dal file delle anagrafiche imprese. Il file finale contiene in totale 27 variabili. La Tabella 11 riporta il numero di imprese presenti nel dataset per ogni anno.

Tabella 11: Numero di imprese nel file standard nel periodo considerato.

Anno	Osservazioni	Percentuale	Cumulata
2012	269141	10.63	10.63
2013	258663	10.22	20.85
2014	250774	9.91	30.76
2015	253934	10.03	40.80
2016	250097	9.88	50.68
2017	252447	9.97	60.65
2018	253006	10.00	70.65
2019	254223	10.04	80.69
2020	226827	9.36	90.05
2021	251806	9.95	100.00
Totale	2530918 ⁴	100.00	

Le chiavi per aggregare l'archivio delle anagrafiche con quello dei rapporti annuali sono le seguenti:

- **id_firm**: corrisponde alla matricola dell'azienda anonimizzata
- **year**: che specifica l'informazione inerente all'anno di interesse nell'archivio dei rapporti di lavoro annuali.

Di seguito viene presentata la lista delle variabili incluse nel file standard, specificando anche l'archivio di provenienza, e nel caso in cui la variabile è stata calcolata partendo da informazioni originarie dell'archivio viene brevemente descritta la modalità di derivazione.

Le seguenti variabili sono derivate dall'archivio delle anagrafiche delle matricole contributive (in particolare il file "visitinps_anagrafica_aziende_v7.dta"), e riguardano informazioni che generalmente non mostrano variabilità temporale.

- **legal_form**, che fornisce l'informazione rispetto alla tipologia di impresa (ditte individuali/familiari, società di persone, società di capitali ecc.);

⁴ Il numero di osservazioni differisce leggermente rispetto al paragrafo precedente perché sono state eliminate delle righe che presentavano anomalie.

- industry, che indica il settore di attività, secondo la classificazione Ateco, a 3 digit;
- province, che indica la provincia dove l'impresa è localizzata;
- date_establishment: data di costituzione della matricola contributiva. Tale variabile, (ove disponibile), come indicato nella Tabella 12, è stata riarrangiata in sei classi di dimensione simile.

Tabella 12: Classi della variabile date of establishment.

date_establishment	Frequenza	Percentuale	Cumulata
Before 1995	463688	18.32	18.32
From 1995 to 2004	491432	19.42	37.74
From 2005 to 2010	539925	21.33	59.07
From 2011 to 2014	516043	20.39	79.46
After 2014	519830	20.54	100.00
TOTALE	2530918	100.00	

Dall'archivio dei rapporti di lavoro annuali 2012-2021 sono state selezionate le seguenti variabili, che sono talvolta variabili originarie e altre volte da esse derivate.

Per quanto riguarda la composizione della forza lavoro si forniscono le seguenti variabili aggregate a livello di impresa:

- n_workers: numero di contratti di lavoro dichiarati dall'azienda nell'anno (poiché ogni lavoratore nell'anno può aver intrattenuto più di un rapporto di lavoro con una data impresa questo numero può essere superiore al numero di lavoratori che nell'anno hanno intrattenuto rapporti di lavoro con l'azienda);
- n_workers_fte: numero di rapporti di lavoro pesato per il numero di giorni lavorati e per la percentuale di part-time (full time equivalent);
- n_part_time_fte: numero di contratti di lavoro part time (full time equivalent);
- n_fixed_term_fte: numero di contratti di lavoro a tempo determinato (full time equivalent);
- n_men_fte: numero di contratti di lavoro intestati a uomini (full time equivalent);
- n_women_fte: numero di contratti di lavoro intestati a donne, pesato per il numero di giorni lavorati e per la percentuale di part-time (full time equivalent);
- n_blue_collar_fte: numero di contratti di lavoro con qualifica da operaio (full time equivalent);
- n_white_collar_fte: numero di contratti di lavoro con qualifica da impiegato (full time equivalent);
- n_manager_fte: numero di contratti di lavoro con qualifica da dirigente o quadro (full time equivalent);
- n_apprentice_fte: numero di contratti di lavoro con qualifica da apprendista (full time equivalent).

La variabile n_workers è stata rielaborata ed organizzata in classi come segue (per minimizzare rischi di identificazione):

- Valori da 1 a 15: immutati (cioè non si creano classi di dimensione, essendo la numerosità di impresa elevata fino ai 15 lavoratori);
- Valori da 16 a 20 sono aggregati in una unica categoria;
- Valori da 21 a 30 sono aggregati in una unica categoria;
- Valori da 31 a 40 sono aggregati in una unica categoria;
- Valori da 41 a 50 sono aggregati in una unica categoria;
- Valori da 51 a 60 sono aggregati in una unica categoria;

- Valori da 61 a 75 sono aggregati in una unica categoria;
- Valori da 76 a 100 sono aggregati in una unica categoria;
- Valori maggiori di 100 sono aggregati in una unica categoria.

Si viene pertanto a delineare una nuova variabile `n_workers` a 23 modalità.

Le altre variabili sul numero di dipendenti misurati in full time equivalent sono state arrotondate alla prima cifra decimale e tutti i valori superiori a 100 sono stati inclusi in un'unica categoria generica.

Seguono variabili riguardanti il salario, le quali per ragioni di privacy sono state rielaborate secondo procedure di top/bottom coding e arrotondamento. In primo luogo, si è applicata la winsorizzazione, ossia, i valori che sono inferiori al primo percentile e superiori al novantanovesimo percentile sono stati sostituiti rispettivamente con il valore corrispondente al primo percentile e il valore corrispondente al novantanovesimo percentile. Successivamente, i valori rimanenti sono stati arrotondati come segue: i valori compresi tra il primo e il quinto percentile sono stati arrotondati ai 500 euro più vicini, i valori compresi tra il quinto e il novantacinquesimo percentile sono stati arrotondati ai 100 euro più vicini e i valori compresi tra il novantacinquesimo e il novantanovesimo percentile sono stati arrotondati ai 500 euro più vicini. Tale arrotondamento variabile è giustificato dall'intento di riduzione di granularità nelle osservazioni vicine alle code, dove la frequenza delle imprese è minore con un aumento del rischio di identificabilità. Le variabili sul salario sono:

- `mean_wage`: salario medio pagato nell'anno per azienda;
- `mean_wage_part_time`: salario medio pagato dall'azienda per anno ai contratti part time;
- `mean_wage_fixed_term`: salario medio pagato dall'azienda per anno ai contratti a tempo determinato;
- `mean_wage_blue_collar`: salario medio pagato dall'azienda per anno per i contratti con qualifica da operaio;
- `mean_wage_white_collar`: salario medio pagato dall'azienda per anno per i contratti con qualifica da impiegato;
- `mean_wage_manager`: salario medio pagato dall'azienda per anno per i contratti con qualifica da dirigente o quadro;
- `mean_wage_apprentice`: salario medio pagato dall'azienda per anno per i contratti con qualifica da apprendista;
- `mean_wage_men`: salario medio pagato dall'azienda per contratti intestati a uomini;
- `mean_wage_women`: salario medio pagato dall'azienda per contratti intestati a donne.

Per quanto riguarda le variabili inerenti alla distribuzione dei salari all'interno dell'impresa, si è optato per l'indice di Gini calcolato per ciascuna impresa per ogni anno e arrotondato alla seconda cifra decimale. In particolare, sono stati calcolati:

- `gini`: indice di Gini calcolato prendendo in considerazione la variabile salari totali per i lavoratori dichiarati dall'azienda nell'anno di riferimento;
- `gini_fte`: indice di Gini: calcolato prendendo in considerazione la variabile salari totali per i lavoratori dichiarati dall'azienda nell'anno di riferimento, pesata per i giorni retribuiti e per la percentuale di part time (full time equivalent).

3 Ridefinizione di variabili con classificazione che varia nel tempo

L'analisi dei database di partenza ("Anagrafica_aziende_v7" e "Rapporti annuali 2012-2021") ha evidenziato la presenza di variabili le cui modalità sono cambiate nel tempo. Al fine di ottenere un sistema classificatorio sintetico e coerente è stata effettuata una valutazione delle loro principali caratteristiche, in base alle quali si è poi proceduto ad una riclassificazione.

3.1 Riclassificazione variabile "Legal form"

La base di dati più recente relativa all'anagrafica delle aziende in attività e con almeno un dipendente, e che pertanto appartengono all'archivio dei lavoratori dipendenti privati extra agricoli, ha mostrato la presenza di 113 diversi tipi di forme in cui le diverse unità economiche risultano organizzate.

In generale, la maggior parte delle classificazioni, effettuate per unificare forme giuridiche troppo eterogenee e talvolta generiche, fa riferimento alla legislazione europea ed italiana. L'evoluzione normativa, infatti, ha determinato negli anni la nascita di nuove forme giuridiche e la modifica o cessazione di altre. In questo contesto, pertanto, i presupposti di origine giuridica possono offrire la possibilità di associare, ad una determinata categoria di soggetti economici, caratteristiche generali il più possibile uniformi.

Da una riflessione sui soggetti economici individuati si è passati quindi ad un confronto con le classificazioni delle forme organizzative registrate nei principali archivi amministrativi italiani. Da tale comparazione è emerso che la classificazione Uniemens ha un'adesione quasi totale con la ripartizione prevista dal Registro delle imprese; tuttavia, ciò non ha impedito di verificare anche le classificazioni adottate dall'Anagrafe tributaria e dall'Istat.

Nel Registro delle imprese sono censite circa il 95% delle imprese che esercitano attività economica. A partire da questa classificazione, Inps ha aggiunto altre forme di attività che, pur non essendo strettamente economiche, possono assumere personale alle loro dipendenze (es.: parrocchia, condominio).

Ne risulta che la ripartizione delle forme giuridiche in Uniemens è piuttosto variegata e tiene conto della normativa che le regola, che si presta difficilmente ad una agevole generalizzazione.

Si deve considerare, inoltre, che in Uniemens confluiscono contribuzioni di vario tipo (IVS, Naspi, Maternità, TFR, ecc.), circostanza che rende possibili 3 scenari: può accadere che alcune imprese siano obbligate a versare tutti i contributi alla gestione privata, altre per le quali è prescritto il versamento esclusivo alla gestione pubblica, ed una forma "mista" di aziende per le quali la legislazione prevede che i contributi principali siano versati alla gestione privata e quelli minori alla gestione pubblica (o viceversa).

A questo proposito è esemplare il caso di quelle aziende pubbliche che, dai primi anni '90 fino al 2016, hanno subito processi di privatizzazione (aziende a partecipazione statale, banche e imprese pubbliche, aziende di servizio pubblico, telecomunicazioni, settore energetico, ecc.). In molti di questi casi la trasformazione in S.p.A. ha determinato il passaggio di alcune contribuzioni dalla gestione pubblica alla gestione privata, mentre in altri è stato possibile mantenere lo stesso regime di contribuzione. A questo deve aggiungersi che, frequentemente, è stata garantita, ai dipendenti, la possibilità di optare per un sistema pensionistico diverso da quello fino ad allora attribuito generando, anche in questo caso, uno spostamento di contribuzione da una gestione all'altra. L'inquadramento delle imprese in INPS avviene attraverso l'attività esercitata e dichiarata dall'impresa stessa, che viene preliminarmente codificata in base alla tabella ATECO 2007.

Successivamente l'INPS associa a ciascuna attività un codice statistico contributivo (C.S.C.), composto da cinque cifre, sulla base del quale vengono assegnate all'impresa le aliquote contributive relative all'attività svolta ed alle assicurazioni cui è soggetta.

Al C.S.C. viene sempre abbinato il codice ISTAT, che descrive nel particolare l'attività aziendale. Per alcune categorie di imprese, sono attribuiti anche i Codici Autorizzazione (C.A.), che hanno lo scopo di individuare, all'interno di imprese aventi il medesimo C.S.C., quelle soggette ad un particolare regime contributivo o che beneficiano di sgravi e riduzioni.

L'insieme di tutti i codici attribuiti definisce il regime contributivo dell'azienda e quindi l'aliquota che deve essere applicata per il versamento dei contributi.

Questo sistema articolato di codificazioni non consente, tuttavia, di identificare agevolmente le imprese che nel tempo sono state privatizzate o hanno trasformato la propria forma giuridica; pertanto, è estremamente complesso risalire alla loro evoluzione "storica".

Per completezza di informazioni è bene considerare che nel 2015 sono confluiti in Uniemens i dati relativi ai lavoratori ex Enpals mentre, nell'aprile del 2020, il flusso è stato arricchito dalle contribuzioni dei lavoratori del settore agricolo.

Tenuto conto dei dati contenuti nel dataset e a partire dal confronto con le classificazioni delle forme giuridiche operate dal Registro delle imprese, dall'Anagrafe tributaria e dall'Istat, si è pensato di optare per una prima proposta organica individuando sette modalità (compresa una modalità residuale).

La Tabella 13 fornisce a sinistra l'informazione sulle modalità originali incluse nella variabile natura_giuridica, e la associate frequenza, mentre nell'ultima colonna viene riportata la riclassificazione sintetica e omogenea temporalmente che è stata scelta. Il numero totale delle osservazioni è pari a 9,149,824, delle quali 186,836 sono associate a dati mancanti (missing), quindi nelle tabelle il numero totale di osservazioni è 9,336,660.

Tabella 23: Classificazione analitica delle forme giuridiche presenti nel database "Anagrafica_aziende_v7".

	Frequenza	Percentuale	Cumulata	Riclassificazione
Ditta	3,014,361	32.29	32.29	Impresa indiv. e fam.
Società a responsabilità limitata	1,781,518	19.08	51.37	Soc. di capitali
Impresa individuale	1,314,055	14.07	65.44	Impresa indiv. e fam
Società in nome collettivo	816,046	8.74	74.18	Soc. di persone
Società in accomandita semplice	599,231	6.42	80.60	Soc. di persone
Soggetto Estero	273,235	2.93	83.53	Altre forme
Società a responsabilità limitata semplificata	194,438	2.08	85.61	Soc. di capitali
Società per azioni	174,716	1.87	89.48	Soc. di capitali
Società di fatto	147,239	1.58	91.06	Soc. di persone
Società a responsabilità limitata unico socio	135,857	1.46	92.51	Soc. di capitali
Società cooperativa	106,160	1,14	93.65	Soc. cooperative
Condominio	97,807	1.05	94.70	Altre forme
Associazione	91,128	0.98	95.67	Altre forme
Ente	60,173	0.64	96.32	Soc. cooperative
Società cooperativa a responsabilità limitata	58,988	0.63	96.95	Altre forme

Scuola	43,543	0.47	97.42	Altre forme
Studio associato	24,793	0.27	97.68	Soc. di persone
Amministrazione statale	16,707	0.18	97.86	Altre forme
Cooperativa sociale	15,986	0.17	98.03	Soc. cooperative
Altre Forme	13,995	0.15	98.18	Altre forme
Società semplice	11,889	0.13	98.31	Soc. di persone
Società consortile a responsabilità limitata	10,132	0.11	98.42	Impresa indiv. e fam
Datore di lavoro	9,883	0.11	98.52	Soc. di capitali
Società per azioni con socio unico	9,754	0.10	98.63	Altre forme
Società costituita in base a legge di altro Stato	8,985	0.10	98.72	Altre forme
Parrocchia	8,372	0.09	98.81	Altre forme
Persona fisica Proprietario	8,340	0.09	98.90	Soc. cooperative
Piccola società coop a resp. limitata	7,706	0.08	98.98	Impresa indiv. e fam
Circolo	6,847	0.07	99.06	Altre forme
Consorzio di cui al DLGS 267/2000	6,427	0.07	99.13	Altre forme
Fondazione	6,395	0.07	99.20	Altre forme
Sindacato	5,584	0.06	99.26	Altre forme
Consorzio	5,202	0.06	99.31	Altre forme
Istituto	5,114	0.05	99.37	Altre forme
Ente di diritto pubblico economico non economico	5,002	0.05	99.42	Altre forme
Agenzia	4,065	0.04	99.46	Altre forme
Soggetto estero	3,210	0.03	99.50	Soc. cooperative
Società cooperativa a responsabilità illimitata	2,979	0.03	99.53	Altre forme
Ente ecclesiastico civilmente riconosciuto	2,613	0.03	99.56	Soc. di capitali
SRL a capitale ridotto	2,085	0.02	99.58	Altre forme
Azienda agricola	2,011	0.02	99.60	Altre forme
Ente ecclesiastico	1,986	0.02	99.62	Altre forme
Chiesa	1,902	0.02	99.64	Altre forme
Consorzio con attività esterna	1,880	0.02	99.66	Altre forme
Federazione	1,832	0.02	99.68	Altre forme
Impresa familiare	1,535	0.02	99.70	Impresa individ. e familiare
Istituto religioso	1,526	0.02	99.72	Altre forme
Ente morale	1,489	0.02	99.73	Altre forme
Azienda municipale	1,153	0.01	99.73	Altre forme
Società consortile in accomandita semplice	1,153	0.01	99.75	Altre forme
Società consortile in nome collettivo	1,153	0.01	99.76	Altre forme
Comitato	1,149	0.01	99.77	Altre forme
Organizzazione	1,103	0.01	99.78	Altre forme
Comunione ereditaria	1,086	0.01	99.80	Soc. di persone
Gruppo	1,062	0.01	99.81	Altre forme
Partito politico	1,044	0.01	99.82	Altre forme

Associazione temporanea d'impresa	1,004	0.01	99.83	Altre forme
Carovana	955	0.01	99.84	Altre forme
Società consortile cooperativa a responsabilità limitata	920	0.01	99.85	Altre forme
Unione	879	0.01	99.86	Altre forme
Ente pubblico economico	866	0.01	99.87	Altre forme
Compagnia	853	0.01	99.88	Altre forme
Società fra professionisti	833	0.01	99.89	Soc. di persone
Società consortile per azioni	736	0.01	99.90	Altre forme
Congregazione	733	0.01	99.91	Altre forme
Scuola elementare (direzione didattica)	611	0.01	99.92	Altre forme
Ministero	486	0.01	99.92	Altre forme
Società agricola	478	0.01	99.93	Altre forme
Società cooperativa Consortile	462	0.01	99.93	Soc. cooperative
Confederazione	415	0.00	99.94	Altre forme
Ambasciata	412	0.00	99.94	Altre forme
Azienda speciale	394	0.00	99.94	Altre forme
Arciconfraternita	334	0.00	99.95	Altre forme
Società in accomandita per azioni	329	0.00	99.95	Soc. di persone
Confraternita	285	0.00	99.96	Altre forme
Ente Impresa	257	0.00	99.96	Altre forme
Azienda Speciale di cui al DLGS 267/2000	250	0.00	99.96	Altre forme
Società di mutuo soccorso	248	0.00	99.96	Altre forme
Università	217	0.00	99.97	Altre forme
Comando	196	0.00	99.97	Altre forme
Conservatorio	196	0.00	99.97	Altre forme
Azienda provinciale	179	0.00	99.97	Altre forme
Ispettorato	178	0.00	99.97	Altre forme
Azienda regionale	169	0.00	99.98	Altre forme
Rappresentanza diplomatica	154	0.00	99.98	Altre forme
Piccola Società cooperativa	151	0.00	99.98	Soc. cooperative
Ente sociale	143	0.00	99.98	Altre forme
Contratto di rete dotato di soggettività giuridica	140	0.00	99.98	Altre forme
Gruppo europeo di interesse economico	135	0.00	99.98	Altre forme
Azienda autonoma statale	130	0.00	99.99	Altre forme
Consorzio con o senza attività esterna	128	0.00	99.99	Altre forme
Provveditorato	119	0.00	99.99	Altre forme
Società europea	100	0.00	99.99	Soc. di capitali
Azienda speciale di ente locale	96	0.00	99.99	Altre forme
Fondazione Impresa	94	0.00	99.99	Altre forme
Ente pubblico commerciale	89	0.00	99.99	Altre forme
Società anonima	89	0.00	99.99	Altre forme
Società irregolare	76	0.00	100.00	Soc. di persone
Consorzio intercomunale	60	0.00	100.00	Altre forme
Società consortile	60	0.00	100.00	Altre forme

Commissione	50	0.00	100.00	Altre forme
Istituto di credito	46	0.00	100.00	Altre forme
Mutua assicurazione	42	0.00	100.00	Altre forme
Società non prevista dalla legislazione italiana	40	0.00	100.00	Altre forme
Società mutua assicuratrice	33	0.00	100.00	Soc. cooperative
Legazione	18	0.00	100.00	Altre forme
Consorzio municipale	17	0.00	100.00	Altre forme
Associazione in partecipazione	15	0.00	100.00	Altre forme
Società cooperativa Europea	9	0.00	100.00	Soc. di capitali
Natura Giuridica Non Presente	8	0.00	100.00	Altro
Ente di cui alla l.r. 21-12-93, N.88	7	0.00	100.00	Altre forme
Consorzio Municipale	6	0.00	100.00	Altre forme
Consorzi Fidi	4	0.00	100.00	Altre forme
TOTALE	8,884,679	100.00		

Si può notare come la modalità “Impresa individuale e familiare”, da sola, assorba quasi la metà delle frequenze totali. Ciò è dovuto al raggruppamento dei soggetti economici identificati come “Ditta” con la tipologia indicata come “Impresa individuale”. La disciplina della ditta, prevista dal Codice civile, è stata pensata per l’imprenditore individuale. Con lo sviluppo dell’attività di impresa in forma societaria, tuttavia, la disciplina è stata applicata anche alla ragione sociale (soc. di persone) e alla denominazione sociale (soc. di capitali). Oggi si applica anche a realtà di grandi dimensioni (società che possono avere anche più di una ditta).

Mettendo a confronto la variabile “NATURA GIURIDICA” con la variabile “DATA di COSTITUZIONE” emerge che, nella stragrande maggioranza dei casi, la “Ditta” si riferisce ad imprese costituite prima del 2000. Successivamente diventa più frequente la sigla “DI” (Ditta Individuale). Ciò non toglie che esistano casi di “Ditta” con data di costituzione successiva al 2000 e “Ditta Individuale” precedente al 2000, ma si tratta di occorrenze trascurabili ai fini di questa trattazione.

Si è pensato, inoltre, di fondere le modalità “Impresa individuale” e “Impresa Familiare” in quanto concettualmente simili. Per avere una impresa familiare, infatti, si devono escludere rapporti di natura diversa, quali ad esempio la subordinazione, la società o il contratto di associazione in partecipazione. La tesi dominante della dottrina giuridica italiana, inoltre, ritiene che l’impresa familiare sia un istituto peculiare dell’impresa individuale, alla luce del ruolo predominante dell’imprenditore per la conduzione della stessa; pertanto, la loro unificazione è apparsa appropriata. La Tabella 14 riporta le frequenze per la riclassificazione sintetica e omogenea temporalmente.

Tabella 14: Riclassificazione delle forme giuridiche.

Legal form	Frequenza	Percentuale	Cumulata
Impresa indiv. e fam.	4249757	47.83	47.83
Soc. di capitali	2176615	24.50	72.33
Soc. di persone	1579853	17.78	90.11
Altre forme⁵	691081	7.78	97.89
Soc. cooperative	187375	2.11	100.00
TOTALE	8884681	100.00	

Nella versione finale del file, si è deciso di semplificare ulteriormente la classificazione unendo le due categorie impresa individuali e familiari e società di persone nella categoria “partnership”. Le società di capitali sono state rinominate “corporation”. Inoltre, si è deciso di unire le società cooperative nelle altre forme (other). Segue tabella riassuntiva della classificazione finale:

Tabella 15: Classificazione finale della variabile legal form.

Legal form	Frequenza	Percentuale	Cumulata
Partnership	5829610	65.61	65.61
Corporation	217615	24.50	90.11
Other	878456	9.89	100
TOTALE	8884681	100.00	

⁵ Tra le “Altre forme”, il 3% è rappresentato dalla forma giuridica “Soggetto estero” che è costituita, nella quasi totalità dei casi, da studi professionali: legali, tecnici, commerciali, studi medici, dentistici, veterinari, analisi cliniche, agenzie di pubblicità, disbrigo pratiche, marittime, lotterie, intermediari: ag. immobiliari, finanziarie, depositi, logistica. La seconda modalità più frequente riguarda, invece, la forma giuridica “Condominio” e copre l’1% del dataset.

3.2 Riclassificazione variabile “Qualifica”

Per la riclassificazione della variabile “Qualifica” sono stati analizzati i dati contenuti nei file “Rapporti annuali” del 2021. L’analisi ha messo in luce la presenza di 26 modalità diverse per questa variabile negli archivi.

La Tabella 16 mostra le modalità presenti all’anno considerato (2021), dove nell’ultima colonna è presente anche la riclassificazione sintetica scelta per ogni modalità originaria.

Tabella 16: Classificazione analitica variabile “Qualifica” – anno 2021

QUALIFICA	Frequenza	Percentuale	Cumulata	Riclassificazione
Operaio	11998608	58.55	58.55	Operaio
Impiegato	6439454	31.42	89.97	Impiegato
Apprendista	1069337	5.22	95.19	Apprendista
Lavoratore con qualifica di quadro	548707	2.68	97.87	Dirigente e quadro
Apprendista mantenuto in servizio come operaio	104685	0.51	98.38	Dirigente e quadro
Apprendista qualificato impiegato	94072	0.46	98.84	Apprendista
Dirigente	93909	0.46	99.30	Apprendista
Dirigenti di aziende industriali	52171	0.25	99.55	Dirigente e quadro
Viaggiatore o piazzista	23436	0.11	99.67	Altro
Lavoratore domestico dipendente da ag. interinale	15377	0.08	99.74	Altro
Assistente di volo	10023	0.05	99.79	Altro
Lavoratore in esodo	9479	0.05	99.84	Altro
Equiparato impiegato	9094	0.04	99.88	Impiegato
Lavoratore a domicilio	8215	0.04	99.92	Apprendista
Apprendista non soggetto all’assicurazione infortuni	715	0.03	99.96	Altro
Pilota (fondo volo)	5852	0.03	99.98	Altro
Apprendista non professionalizzante in servizio come operaio	1883	0.01	99.99	Apprendista
Apprendista non professionalizzante in servizio come impiegato	910	0.00	100.00	Altro
Lavoratore escluso da contribuzione Inps	779	0.00	100.00	Altro
Pilota in addestramento (primi 12 mesi)	161	0.00	100.00	Apprendista
Lavoratore contribuz figurativa correlata esodo	133	0.00	100.00	Altro
Tecnico di volo per i collaudi	71	0.00	100.00	Altro
Pilota collaudatore	32	0.00	100.00	Altro
Giornalista professionista, praticante	12	0.00	100.00	Altro
Atipica ex INPDAI	3	0.00	100.00	Altro
Tecnico di volo	2	0.00	100.00	Altro
TOTALE	20492776	100.00		

Anche in questo caso si ritiene che i cambiamenti normativi e contributivi abbiano apportato, negli anni, integrazioni e specificazioni più dettagliate dei diversi tipi di occupazione. Tuttavia, le qualifiche che vengono utilizzate maggiormente nel mondo del lavoro sono molto poche, rispetto alla totalità delle modalità

disponibili. Si è ritenuto opportuno, quindi, concentrare la riclassificazione su tali modalità. Le frequenze associate alla nuova classificazione sono riportate nella Tabella 17.

Tabella 17: Riclassificazione variabile "Qualifica" (periodo 2012-2021).

QUALIFICA	Frequenza	Percentuale	Cumulata
Operaio	9910385	58.05	58.05
Impiegato	5411864	31.70	89.75
Apprendista	148344	9.22	90.62
Dirigente e quadro	1573422	0.87	99.84
Altro	27495	0.16	100.00
TOTALE	17071510	100.00	