Unione europea
Fondo sociale europeo

REGIONE DEL VENETO

POR Fondo Sociale Europeo 2014-2020

ASSE 1 OCCUPABILITA'
ASSE 2 INCLUSIONE SOCIALE

DGR n. 1095 del 13 luglio 2017

Direttiva per la sperimentazione dell'Assegno per il lavoro per la ricollocazione di lavoratori disoccupati

Assegno per il Lavoro – Anno 2017

MODELLO PER LA PROFILAZIONE DEI DESTINATARI DELL'ASSEGNO PER IL LAVORO

VENETO LAVORO

Versione del 30 settembre 2017

Sommario

1. MISSION 3
2. IDENTIFICAZIONE DELL'UNIVERSO DA TESTARE E PREPARAZIONE DEI DATI 3
3. IDENTIFICAZIONE DELLA VARIABILE DIPENDENTE 3
4. SELEZIONE DELLE VARIABILI INDIPENDENTI 4
Fig. 1 Matrice di associazione 5
Tab. 1 Distribuzione dei soggetti osservati secondo le loro principali caratteristiche 6
Fig. 2 Distribuzione della distanza in mesi 7
5. MODELLO E COEFFICIENTI STIMATI 7
Tab. 2 Tabella riassuntiva del modello logit 9
Tab. 3 Tabella errore del modello logit 10
6. DETERMINAZIONE DELLE FASCE DI AIUTO 10
Tab. 4 Tabella fasce intensità di aiuto 10
Fig. 3 Decili della variabile dipendente stimata e suddivisione in fasce 11
Fig. 4 Distanza did ricollocamento in mesi per fascia di aiuto 12
Fig. 5 Distribuzione dell'età per fascia di aiuto 13
Fig. 6 Distribuzione del titolo di studio per fascia di aiuto 14
7. INOCCUPATI 15
8. CONCLUSIONI 15

1. MISSION

All'Osservatorio Regionale del Lavoro è stato chiesto di profilare i soggetti destinatari dell'Assegno per il Lavoro (AxL) in funzione del grado di occupabilità, definito come l'inverso del rischio di permanenza nella condizione di non lavoro. La dgr nr. 1095 del 13/7/2017 prevede la suddivisione dei beneficiari dell' AxL in tre fasce a seconda dell'intensità di aiuto necessario per la loro ricollocazione ("bassa intensità di aiuto", "media intensità di aiuto", "alta intensità di aiuto").

2. IDENTIFICAZIONE DELL'UNIVERSO DA TESTARE E PREPARAZIONE DEI DATI

La prima fase di lavoro ha riguardato l'identificazione della popolazione da testare nell'ambito dell'attività di profilazione.[1] Sono stati così selezionati 135.942 lavoratori disoccupati di età superiore ai 35 anni, che hanno rilasciato la dichiarazione di immediata disponibilità (did) in Veneto nel biennio 2013-2014[2]. Nel caso di più did rilasciate nel periodo è stata selezionata la prima.

Sono state effettuate alcune operazioni di editing dei dati per migliorare la qualità e l'interpretabilità del modello tra cui l'assegnazione del titolo di studio "licenza elementare" a tutte le unità per le quali il titolo di studio non era disponibile.

Sono stati esclusi dall'analisi i soggetti che al momento del rilascio della did risultavano alla ricerca della prima occupazione, in quanto numericamente insufficienti per la stima di un ulteriore modello e privi delle informazioni sulla storia lavorativa precedente. Comunque è stato possibile assegnare anche questi individui a una fascia di svantaggio utilizzando esclusivamente le informazioni anagrafiche.

3. IDENTIFICAZIONE DELLA VARIABILE DIPENDENTE

La seconda fase ha riguardato la definizione della variabile dipendente. In linea con l'approccio utilizzato nella profilazione attuata dall'ANPAL e con il profiling del programma Garanzia Giovani, è stata selezionata come variabile risposta la probabilità di iniziare un rapporto di lavoro a tempo indeterminato (incluso l'apprendistato) o a tempo determinato (inclusi i rapporti di lavoro somministrati) con durata prevista superiore ai 3 mesi entro 24 mesi dal rilascio della did. La

1 Le pre-scelte in materia di profilazione sono state illustrate nei documenti interni n. 120 (maggio 2017) e n. 122 (luglio 2017).

2 La scelta del biennio 2013-2014 è funzionale all'osservabilità di ciascun soggetto per almeno 24 mesi dal rilascio della did.

variabile osservata è quindi dicotomica, in quanto nella popolazione osservata assume due valori, 1 se il lavoratore inizia un rapporto di lavoro nel periodo considerato, 0 altrimenti. Nella popolazione in esame 62.848 soggetti hanno iniziato un nuovo rapporto di lavoro, mentre 73.094 sono rimasti disoccupati (con riferimento ai rapporti di lavoro qui considerati: ciò non esclude quindi che essi possano aver avuto accesso a tirocini o siano stati assunti con rapporti di lavoro intermittente etc.).

4. SELEZIONE DELLE VARIABILI INDIPENDENTI

Alla popolazione osservata sono state applicate adeguate metodologie statistiche al fine di individuare in che misura i diversi caratteri concorrono a determinare la probabilità di rioccupazione, intesa sempre nel senso già indicato, nei 24 mesi successivi alla did.
Inizialmente sono stati presi in considerazione i seguenti gruppi di variabili:

a) variabili riferite a caratteristiche socio anagrafiche individuali

1) genere;
2) età al momento del rilascio della did;
3) cittadinanza (italiano o straniero);
4) titolo di studio posseduto al momento dell'entrata in disoccupazione (da did);
5) CPI di appartenenza;

b) informazioni sull'evento di disoccupazione

1) percezione o meno dell'ammortizzatore sociale;

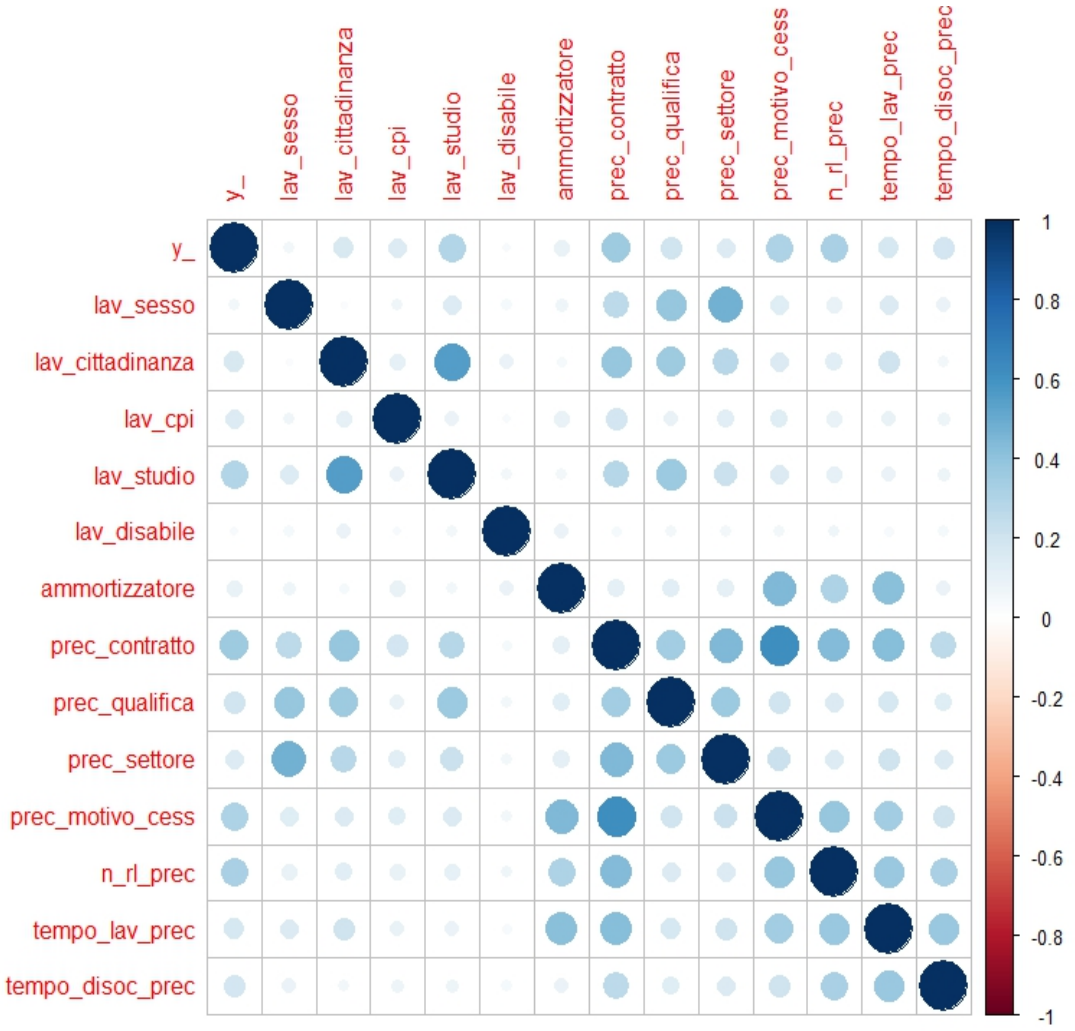
c) informazioni sulla storia lavorativa antecedente all'entrata in disoccupazione:

1) caratteristiche dell'ultimo rapporto di lavoro (settore, contratto, qualifica, motivo cessazione);
2) numero di rapporti lavorativi nei cinque anni precedenti l'evento di disoccupazione;
3) giorni lavorati nei cinque anni precedenti il rilascio della did;
4) giorni di effettiva disoccupazione amministrativa nei cinque anni precedenti il rilascio della did.

Le analisi successive hanno evidenziato la presenza di variabili eccessivamente legate tra loro, come mostrato nella **fig. 1**, che riporta il livello di associazione tra le variabili (qualitative) considerate. In particolare, le informazioni relative alla storia lavorativa del soggetto risultano essere collegate tra di loro e quindi ripetitive. È il caso del motivo di cessazione del rapporto

precedente alla did, fortemente associato sia alla tipologia contrattuale sia all'eventuale percezione di un ammortizzatore sociale, o delle variabili relative al tempo lavorato e passato in disoccupazione nei cinque anni precedenti che risultano essere collegate tra di loro e con altre informazioni sulla storia lavorativa del soggetto.

Fig. 1 – Matrice di associazione



Si è quindi deciso di omettere le variabili ridondanti, il cui effetto sulla probabilità di ricollocamento viene catturato già da altre caratteristiche. Le informazioni escluse dal modello sono pertanto quelle relative alla tipologia del contratto precedente al rilascio della did, all'ammortizzatore eventualmente percepito e alla storia lavorativa nei 5 anni precedenti.

La **tab. 1** riporta le caratteristiche dei lavoratori oggetto della nostra analisi.

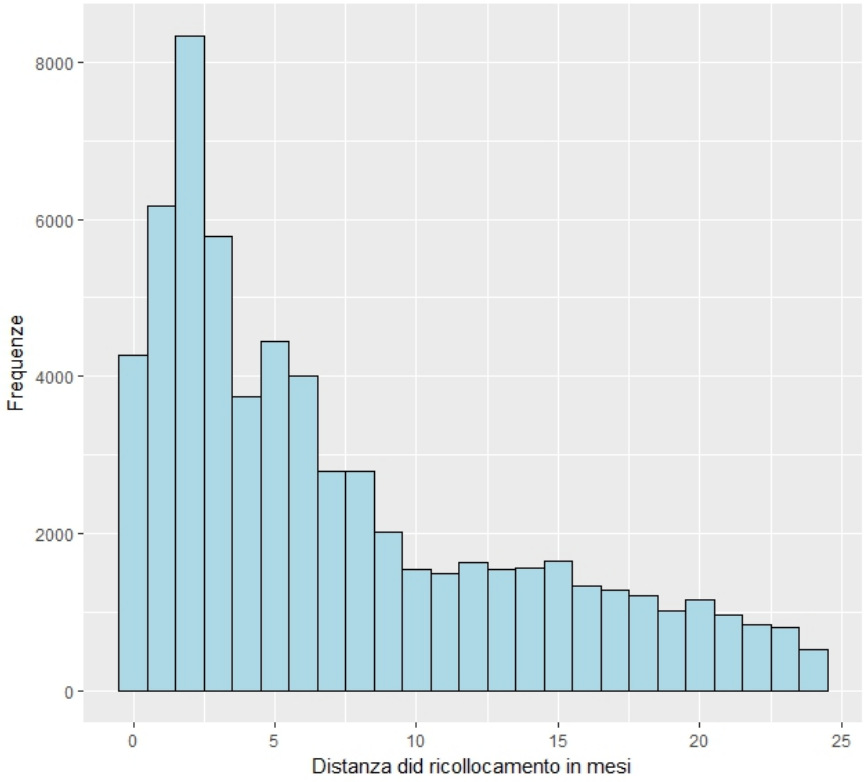
Allegato A al Decreto n. 771 del 29/11/2017

Tab. 1 – Distribuzione dei soggetti osservati secondo le loro principali caratteristiche

Variabile		Non hanno trovato lavoro	Hanno trovato lavoro	Totale
Genere				
	maschi	34.287	32.795	67.082
	femmine	38.807	30.053	68.860
Età (v. continua)				
	mediana	47	43	45
	media	47,61	44,44	46,14
Cittadinanza				
	italiana	50.339	52.446	102.785
	straniera	22.755	10.402	33.157
Livello di studio				
	Lic. Elementare	17.791	3.601	21.392
	Lic. Media	30.617	24.513	55.130
	Qualifica prof.	4.144	3.891	8.035
	Diploma	16.601	21.920	38.521
	Laurea	3.941	8.923	12.864
Lavoratore disabile				
	sì	3.988	2.516	6.504
	no	69.106	60.332	129.438
Motivo cess. rapp. precedente				
	Perdita involontaria	43.872	25.645	69.517
	Perdita volontaria	9.173	5.003	14.176
	Fine termine	15.545	31.102	46.647
	Altro	4.504	1.098	5.602
Qualifica rapp. precedente				
	Professioni intellettuali	2.435	8.365	10.800
	Professioni tecniche	4.892	4.868	9.760
	Impiegati	8.480	8.178	16.658
	Professioni qualif. dei servizi	18.088	13.075	31.163
	Operai specializzati	15.196	11.547	26.743
	Conduttori e operai semi-spec.	9.317	6.876	16.193
	Professioni non qualificate	14.686	9.939	24.625
Ateco rapp. precedente				
	Made in italy	12.308	8.281	20.589
	Metalmeccanico	9.199	6.487	15.686
	Altre industrie	4.011	2.806	6.817
	Costruzioni	6.491	5.961	12.452
	Comm.-tempo libero	9.960	14.747	24.707
	Ingrosso e logistica	6.673	6.637	13.310
	Terziario avanzato	2.872	2.350	5.222
	Servizi alla persona	17.648	12.524	30.172
	Altri servizi	3.932	3.055	6.987

Oltre alle variabili inserite come covariate nel modello, è stata analizzata anche la distanza intercorsa tra il rilascio della did e l'inizio di un rapporto di lavoro nei 24 mesi successivi. Tale informazione è stata utilizzata come controllo dei risultati ottenuti, per verificare se l'assegnazione ad una fascia di svantaggio superiore corrisponde ad un'effettiva maggior distanza del ricollocamento. La **fig. 2** sotto riportata mostra la distribuzione delle frequenze del tempo di ricollocamento entro i 24 mesi.

Fig. 2 – Distribuzione della distanza in mesi



5. MODELLO E COEFFICIENTI STIMATI

Come detto in precedenza, la variabile risposta è data dalla probabilità che il lavoratore ha di ricollocarsi con uno dei contratti considerati entro 24 mesi dalla did.

Data la natura binaria della variabile dipendente utilizzata, è stato stimato un modello logit, il quale permette di determinare quali caratteristiche incidono sulla probabilità del fenomeno oggetto di studio.

Il soggetto preso come baseline nella stima del modello presenta le seguenti caratteristiche:

- genere: maschile

- livello di studio: diploma
- lavoratore disabile: no
- qualifica del rapporto precedente: operaio specializzato
- settore del rapporto precedente: metalmeccanico
- motivo cessazione rapporto precedente: perdita involontaria.

Nel caso di variabili qualitative che presentano due o più modalità, è infatti necessario determinare una modalità di riferimento rispetto alla quale verranno interpretati i coefficienti delle altre. Tale scelta non ha alcun impatto sui risultati ottenuti con il modello.

Prima di procedere con la stima del modello si è provveduto a dividere il campione in due parti:

- la prima contiene il 75% delle osservazioni che verranno utilizzate per la stima del modello
- la seconda contiene il rimanente 25% che sarà utilizzato per le operazioni di verifica.

I risultati ottenuti con la stima del modello sono riportati nella **tab. 2**. Di particolare interesse, oltre al livello di significatività di ciascuna variabile, sono il segno dei coefficienti stimati e il valore degli effetti marginali. Infatti, nei modelli per variabili binarie i coefficienti stimati indicano solamente il tipo di effetto di ciascuna variabile sulla risposta (aumento o diminuzione), ma non l'entità della variazione. Quest'ultima viene determinata attraverso il calcolo dei cosiddetti "effetti marginali".

Le variabili inserite nel modello risultano essere tutte significative per spiegare la probabilità di rioccupazione entro i 24 mesi. Gli effetti marginali mostrano che, a parità di tutte le altre condizioni, le caratteristiche che determinano una maggiore probabilità di re-impiego sono:

- sesso: maschile
- età: intorno ai 35 anni
- cittadinanza: italiana
- livello di studio: laurea
- motivo cessazione: fine termine
- qualifica: professioni intellettuali
- settore ateco: commercio e tempo libero.

Tab. 2 – Tabella riassuntiva del modello logit

Variabile	Coefficiente	Effetto marginale	Std. Error	z value	p-value	
Intercetta	2,129427		0,053717	39,642	<2e-16	***
Genere						
femmine	-0,421462	-10,39%	0,016316	-25,830	<2e-16	***
Età (v, continua)	-0,046922	-1,16%	0,001001	-46,854	<2e-16	***
Cittadinanza						
straniera	-0,285922	-7,00%	0,020654	-13,844	<2e-16	***
Livello di studio						
Lic. Elementare	-1,376398	-29,94%	0,029915	-46,011	<2e-16	***
Lic. Media	-0,357635	-8,79%	0,018666	-19,160	<2e-16	***
Qualifica prof.	-0,210787	-5,15%	0,030541	-6,902	5,14e-12	***
Laurea	0,065171	1,62%	0,029824	2,185	0,02887	*
Lavoratore disabile						
sì	-0,319310	-7,73%	0,032713	-9,761	<2e-16	***
Motivo cess, rapp, precedente						
Perdita volontaria	-0,206667	-5,06%	0,023982	-8,618	<2e-16	***
Fine termine	1,003204	24,55%	0,016720	59,999	<2e-16	***
Altro	-0,459083	-10,95%	0,043632	-10,522	<2e-16	***
Qualifica rapp, precedente						
Professioni intellettuali	0,660026	16,33%	0,039812	16,579	<2e-16	***
Professioni tecniche	-0,096885	-2,39%	0,032091	-3,019	0,00254	**
Impiegati	-0,039195	-0,97%	0,028308	-1,385	0,16618	
Professioni qualif. dei servizi	-0,299581	-7,33%	0,028114	-10,656	<2e-16	***
Conduttori e operai semi-spec.	-0,045697	-1,13%	0,026171	-1,746	0,08080	,
Professioni non qualificate	-0,153845	-3,78%	0,024708	-6,226	4,77e-10	***
Ateco rapp, precedente						
Made in italy	0,069833	1,73%	0,026873	2,599	0,00936	**
Altre industrie	-0,033032	-0,82%	0,036419	-0,907	0,36441	
Costruzioni	0,407820	10,16%	0,030259	13,478	<2e-16	***
Comm.-tempo libero	0,735648	18,18%	0,031063	23,682	<2e-16	***
Ingrosso e logistica	0,359002	8,94%	0,029855	12,025	<2e-16	***
Terziario avanzato	-0,081136	-2,00%	0,041382	-1,961	0,04992	*
Servizi alla persona	0,156217	3,88%	0,031224	5,003	5,64e-07	***
Altri servizi	0,158952	3,95%	0,037530	4,235	2,28e-05	***

Per determinare la bontà del modello stimato nel prevedere la variabile dipendente, oltre a una valutazione della significatività dei coefficienti, esso è stato nuovamente stimato sull'insieme di test e stati verificati gli scostamenti rispetto al valore reale della variabile risposta.

Tali differenze sono riportate nella **tab. 3**.

Tab. 3 – Tabella errore del modello logit

	Valori osservati	
Valori previsti	0	1
0	13.396	6.123
1	4.947	9.520

errore totale	0,33
falsi positivi	0,34
falsi negativi	0,31

Dalle analisi condotte risulta che il modello stimato possiede una buona capacità predittiva come evidenziato anche dalle curve Lift e ROC che non presentano anomalie[3].

6. DETERMINAZIONE DELLE FASCE DI AIUTO

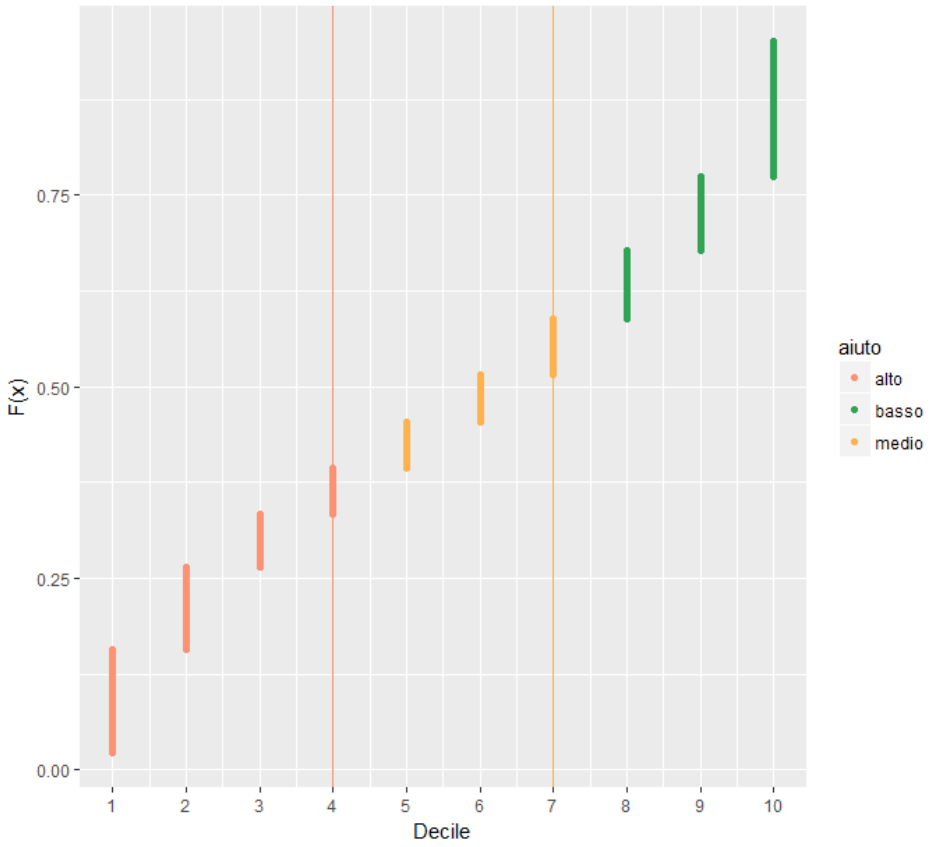
Le tre fasce di aiuto sono state calcolate a partire dalla funzione di ripartizione empirica della variabile dipendente stimata (**fig. 3**). Riportiamo di seguito gli estremi e le numerosità per le tre classi:

Tab. 4 – Tabella fasce intensità di aiuto

fascia ad alta intensità di aiuto	1 – 4	0 – 0,395	13.593
fascia a media intensità di aiuto	5 - 7	0,396 – 0,589	10.196
fascia a bassa intensità di aiuto	8 - 10	0,59 – 1	10.197
totale			33.986

3 La curva ROC rappresenta il rapporto tra vero positivo e falso positivo mentre la curva LIFT mostra le performance del modello rispetto al classificatore casuale.

Fig. 3 – Decili della variabile dipendente stimata e suddivisione in fasce



Per verificare se i risultati ottenuti dividendo la popolazione in fasce sono coerenti con le aspettative, sono stati indagati i valori assunti in ciascuna classe dalle variabili considerate.

La **fig. 3** rappresenta la distribuzione della distanza in mesi tra la did e il ricollocamento per le tre fasce individuate. Nella fascia di aiuto alto meno del 25% degli individui ha iniziato un nuovo rapporto di lavoro entro il periodo preso in esame, nella media il 52% e nella bassa quasi il 70% .

La differenza più evidente che si nota dal grafico riguarda la diversa distribuzione del tempo necessario al re-impiego: la mediana della fascia bassa (7 mesi) è di circa due terzi inferiore a quella della fascia media (20 mesi).

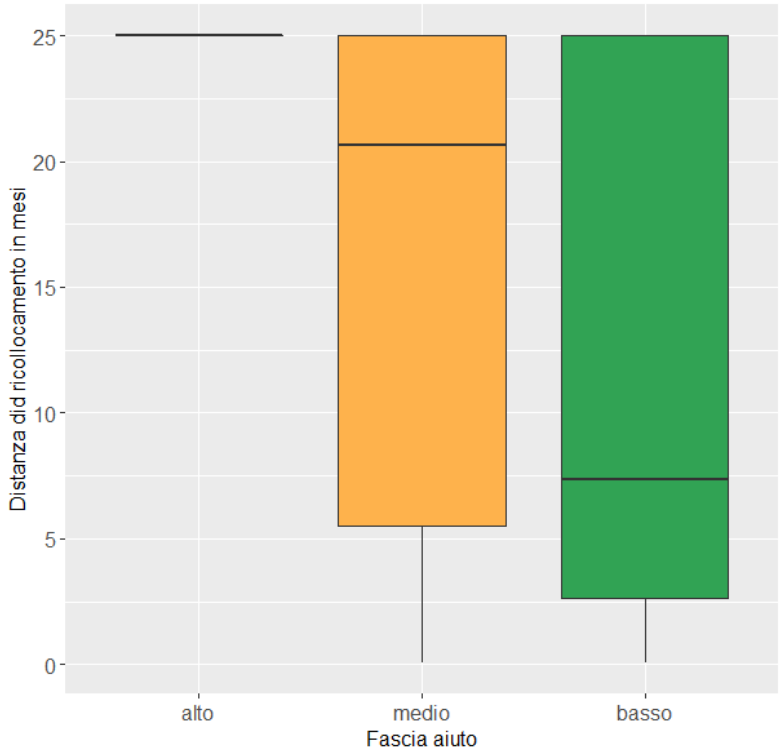


Fig. 4 – Distanza did ricollocamento in mesi per fascia di aiuto

Anche la distribuzione dell'età mostra una certa variabilità tra le tre classi ed è sensibilmente più elevata per la fascia di aiuto alta (mediana sopra ai 50 anni); tale indicatore scende a 44 anni per la fascia media e a 41 per la bassa.

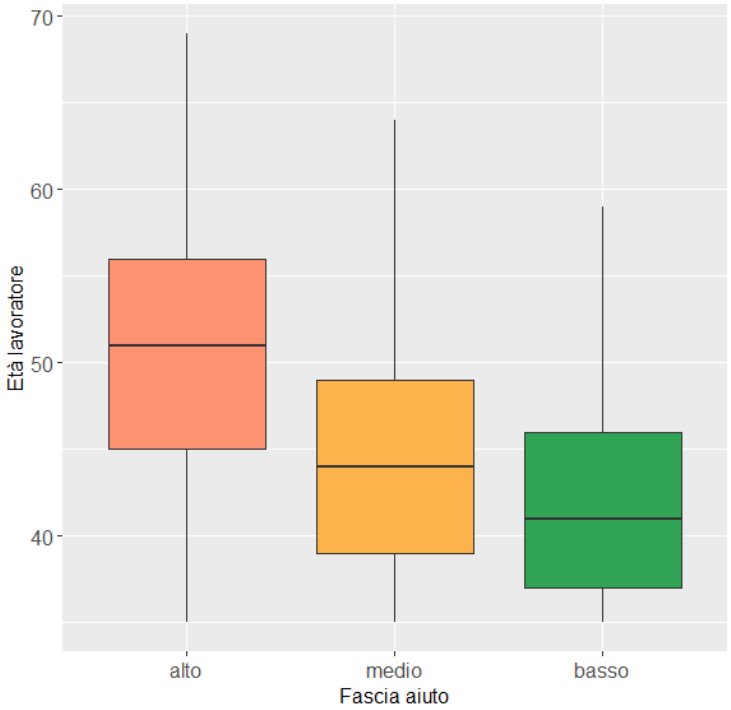


Fig. 5– Distribuzione dell'età per fascia di aiuto

Se analizziamo invece il genere del lavoratore, si può notare una prevalenza femminile nella fascia alta, al contrario della fascia media dove la maggioranza è maschile; la differente distribuzione tra le due modalità è invece meno evidente nella fascia bassa.

Per quanto riguarda la cittadinanza la maggior parte degli stranieri presenta un livello di svantaggio elevato.

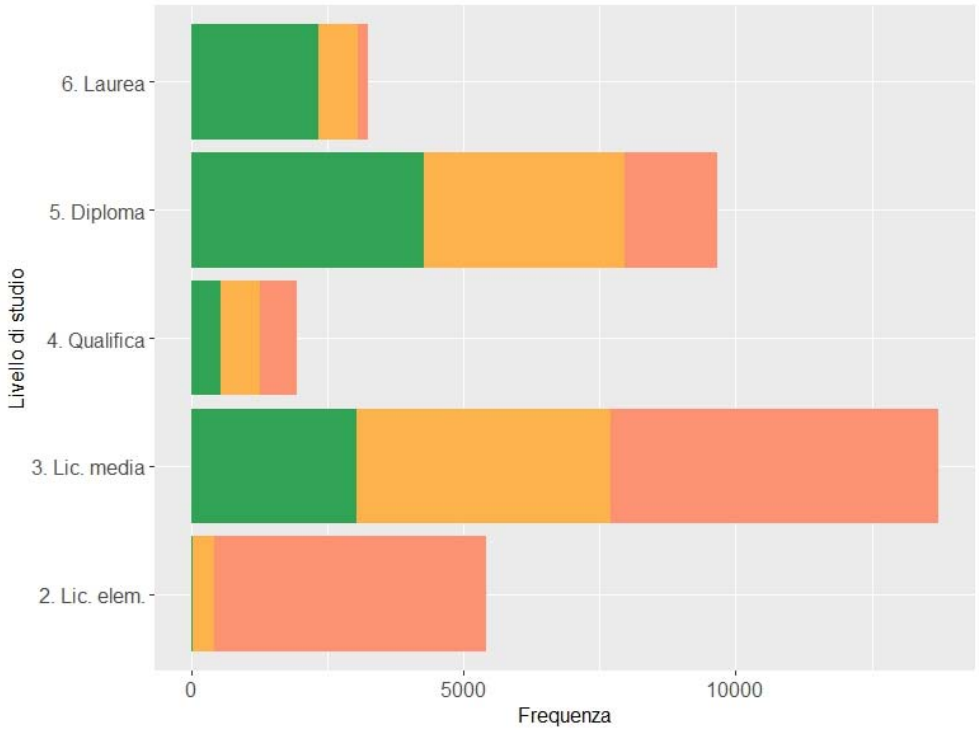


Fig. 6 – Distribuzione del titolo di studio per fascia di aiuto

Un'altra caratteristica discriminante è il livello di studio posseduto al momento della did: nelle fasce di aiuto più alte predominano i titoli di studio più bassi. I laureati e i diplomati si concentrano invece nella classe con minore intensità di aiuto.

Considerando i lavoratori disabili emerge che il 58% si concentra nella fascia di aiuto più elevata, mentre la classe media e quella bassa raccolgono rispettivamente il 27% e il 15%. Il numero di lavoratori disabili rappresenta il 5% dei soggetti analizzati.

Per quanto riguarda le caratteristiche dell'ultimo rapporto di lavoro precedente alla did si nota come il motivo di cessazione più frequente nella classe alta e media è il licenziamento, mentre nella fascia bassa prevale il fine termine. Inoltre, le professioni meno qualificate sono più comuni nella classe di aiuto più elevata, al contrario della fascia inferiore che si caratterizza per una maggior presenza di professioni più qualificate. Infine, i settori dominanti nella fascia di maggior svantaggio sono i servizi alla persona e il made in Italy; nella fascia bassa prevale il settore del commercio e tempo libero.

7. INOCCUPATI

Alla luce di quanto emerso dall'analisi condotta sui disoccupati (condizionalità della storia lavorativa e di disoccupazione pregressa) si è provveduto ad assegnare anche agli inoccupati una fascia di svantaggio utilizzando le informazioni anagrafiche ed assegnando alle variabili "storicizzate" i parametri usati come baseline nella stima del modello per disoccupati.

Infatti, data la relativamente bassa numerosità degli inoccupati (7.544 soggetti) e l'alto sbilanciamento delle frequenze della variabile risposta (solamente 1.285 trovano un impiego entro i 24 mesi dalla did) non è possibile ottenere risultati sufficientemente affidabili.

Mantenendo la suddivisione in fasce definita nel paragrafo precedente otteniamo che 7.086 soggetti vengono classificati nella fascia alta e 458 nella fascia media. Questo è dovuto a un'età più elevata (mediana 47 anni), alla prevalenza del genere femminile, all'alto numero di cittadini stranieri e al titolo di studio basso.

8. CONCLUSIONI

L'analisi effettuata ha permesso di profilare i soggetti destinatari dell'Assegno per il Lavoro con un modello logistico che stima la probabilità di rioccupazione entro 24 mesi dal rilascio della did e la successiva determinazione delle tre fasce di intensità di aiuto. Le tre classi così individuate raccolgono al loro interno soggetti con caratteristiche differenti:

- fascia alta: principalmente donne, età avanzata, basso livello di studio, italiani (ma con alta incidenza anche di stranieri), rapporto di lavoro precedente nel settore dei servizi alla persona e del Made in Italy, bassa qualifica rapporto di lavoro antecedente concluso con un licenziamento.
- fascia media: maggioranza di uomini, età intorno ai 44 anni, possessori di un diploma o licenza media, rapporto di lavoro precedente concluso con un licenziamento.
- fascia bassa: soggetti con età intorno ai 41 anni, possessori di diploma o laurea, rapporto di lavoro precedente nel settore del commercio e tempo libero o dei servizi alla persona, qualifica relativamente elevata e rapporto di lavoro concluso alla scadenza prevista.