

COLLANA
FOCUS

ANPAL
N° 146

ANPAL
Agenzia Nazionale Politiche Attive del Lavoro



METODOLOGIE E APPROFONDIMENTI

STRUMENTI PER L'ATTUAZIONE DELL'ASSESSMENT

La profilazione quantitativa degli utenti dei Servizi per l'impiego

Gennaio 2023



UNIONE EUROPEA

Fondo Sociale Europeo
Investiamo nel tuo futuro



L'ANPAL – Agenzia nazionale per le politiche attive del lavoro – è stata istituita dal D.lgs. 150/2015 con lo scopo di coordinare la rete dei servizi per le politiche del lavoro e la gestione delle politiche attive del lavoro e promuovere l'effettività dei diritti al lavoro, alla formazione e all'elevazione professionale, mediante interventi e servizi che migliorino l'efficienza del mercato. Tramite le proprie strutture di ricerca l'Agenzia svolge anche analisi, monitoraggio e valutazione delle politiche attive e dei servizi per il lavoro.

Commissario Straordinario: Raffaele Tangorra

ANPAL

Via Fornovo, 8

00192 Roma

www.anpal.gov.it

Il lavoro è coordinato dalla Struttura 3 - Ufficio di statistica e supporto metodologico (responsabile Giovanna Linfante).

Autori: Giovanna Linfante, Debora Radicchia, Enrico Toti

Il testo è stato chiuso a dicembre 2022.

Le opinioni espresse in questo lavoro impegnano la responsabilità degli autori e non necessariamente riflettono la posizione dell'Agenzia.

Alcuni diritti riservati [2022] [Anpal].

Quest'opera è rilasciata sotto i termini della licenza Creative Commons Attribuzione - Non commerciale.

Condividi allo stesso modo 4.0. Italia License.

<http://creativecommons.org/licenses/by-nc-sa/4.0>



ISSN 2724-5551

Collana Focus ANPAL

Valorizza gli avanzamenti periodici di monitoraggi e indagini e gli approfondimenti tecnici su iniziative e misure di politica attiva dell'Agenzia.

Prevede sotto-collane dedicate a temi specifici, quali: Garanzia Giovani in Italia, Incentivi per l'occupazione, Contratti di somministrazione, Metodologie e approfondimenti, Reddito di cittadinanza, Approfondimenti Covid-19.

Coordinamento editoriale: Orsola Fornara

INDICE

1. INTRODUZIONE	4
2. IL MODELLO DI ASSESSMENT DEI BENEFICIARI DEL PROGRAMMA GOL.....	4
3. LO STRUMENTO DI PROFILAZIONE QUANTITATIVO PER L'ASSESSMENT DEGLI UTENTI DEI CPI	7
3.1. <i>Modelli con variabile dipendente dicotomica</i>	7
3.2. <i>La scelta della variabile dipendente/outcome</i>	7
3.3. <i>Descrittive outcome (y) e variabili esplicative (X): eterogeneità regionali</i>	8
3.4. <i>Stime dei modelli</i>	11
4. UTILIZZO COMBINATO DEI MODELLI Y_0 E Y_1 : IDENTIFICAZIONE DELLE PLATEE PER FASCIA DI RISCHIO	12
4.1. <i>Falsi Negativi e Falsi Positivi</i>	13
4.2. <i>Definizione delle fasce di rischio</i>	16
5. CONCLUSIONI	17
5.1. <i>Simulazione – per alcuni strati della popolazione</i>	18
APPENDICE	24
BIBLIOGRAFIA	34

1. INTRODUZIONE

Il Programma per la Garanzia di occupabilità dei lavoratori (GOL), adottato con Decreto interministeriale nel novembre 2021¹, nel rispetto della relativa milestone stabilita dal PNRR, rappresenta il perno della riforma del sistema delle politiche attive del lavoro, in una logica organica e integrata. Il Programma si propone infatti di riqualificare l'offerta di politiche attive dei servizi pubblici per l'impiego, disegnando percorsi personalizzati per l'inserimento e il reinserimento occupazionale che garantiscano livelli essenziali di prestazioni uniformi su tutto il territorio nazionale.

Il Programma GOL disegna un nuovo paradigma dell'offerta di servizi da parte dei centri per l'impiego, basata su un approccio personalizzato che tenga conto dei differenti bisogni degli utenti, non solo in termini di competenze professionalizzanti da acquisire o valorizzare, ma anche delle condizioni socio-economiche degli stessi.

Il primo step del percorso occupazionale dell'utente è costituito da un orientamento di base, erogato attraverso un processo di profilazione quantitativa e qualitativa (assessment) che il Programma ha inteso ridefinire con l'adozione di una nuova metodologia.

ANPAL ai sensi dell'articolo 9, comma 1 lettera e) del d.lgs. 150/2015 è tenuta a definire le metodologie di profilazione degli utenti, allo scopo di determinarne il profilo personale di occupabilità.

L'articolo 19, comma 5, del citato decreto legislativo prevede che la valutazione del livello di occupabilità avvenga attraverso l'assegnazione ad una classe di profilazione sulla base delle informazioni fornite al momento della registrazione degli utenti ai servizi per l'impiego secondo una procedura automatizzata di elaborazione dei dati in linea con i migliori standard internazionali.

Alla luce di questa previsione normativa con la delibera n. 6/2016 del CDA dell'ANPAL è stata adottata una prima metodologia del sistema di profilazione dei beneficiari dei servizi per il lavoro. Tale metodologia è descritta nel dettaglio nella nota tecnica "Il sistema di profilazione dei beneficiari dei servizi per il lavoro" allegata alla delibera n. 6/2016.

Successivamente, con il DM n.4/2018 sono stati specificati i livelli essenziali delle prestazioni (LEP) in materia di politiche attive del lavoro da erogare su tutto il territorio nazionale, che prevedono espressamente nel Lep B, la profilazione quantitativa e nel Lep C, la profilazione qualitativa degli utenti dei Servizi per l'impiego.

Mentre, con la deliberazione del Consiglio di Amministrazione dell'ANPAL n. 19/2018, è stato approvato il documento "Linee guida per gli operatori dei centri per l'impiego (profilazione qualitativa)" e adottato un format per la profilazione qualitativa messo a disposizione degli operatori del Centro per l'impiego sul portale MyAnpal.

Al momento dell'adozione del Programma GOL, è stato previsto una revisione generale di questi strumenti per renderli il più possibile aderenti all'obiettivo di "personalizzazione dei percorsi, distinti sulla base di una valutazione multidimensionale dell'utenza".

Questo documento vuole introdurre in linea generale il nuovo Assessment di GOL, e più nello specifico illustrare in termini tecnici il nuovo strumento di profilazione quantitativa adottato, soffermandosi in particolare nella descrizione della metodologia, degli algoritmi applicati e delle metriche utilizzate per confrontare la capacità predittiva del modello, presentando le stime generali e specifiche su particolari strati della popolazione.

2. IL MODELLO DI ASSESSMENT DEI BENEFICIARI DEL PROGRAMMA GOL

Come riportato nel Programma GOL, "gli strumenti e le metodologie devono essere adeguati e resi comuni per individuare in maniera uniforme sul territorio nazionale l'appropriatezza dei percorsi nei servizi".

E inoltre viene precisato che "l'assessment rappresenta una valutazione più complessa e approfondita, in un contesto partecipato in cui non vi sono automatismi, che parte dalle competenze del lavoratore, dai suoi bisogni determinanti per l'occupabilità – anche oltre la sfera lavorativa e delle competenze –, dalle sue

¹ Decreto del Ministro del lavoro e delle politiche sociali, di concerto con il Ministro dell'economia e delle finanze, in data 11 novembre 2021 (GU n. 306 del 27 dicembre 2021).

aspirazioni, ma anche dalle concrete opportunità occupazionali che il mercato del lavoro e il sistema delle imprese locali offre.”

Il nuovo modello di Assessment previsto in Gol attribuisce agli operatori dei Centri per l’Impiego un ruolo centrale, per svolgere il quale avranno a disposizione due strumenti, uno di profilazione quantitativa e uno di profilazione qualitativa completamente rinnovati. Tali strumenti sono stati elaborati all’interno del Comitato Direttivo di GOL con il contributo di tutte le Regioni e Province Autonome e di ANPAL, e sono stati adottati con una specifica Delibera del Commissario Straordinario di ANPAL².

Gli strumenti adottati sono in linea con i migliori standard utilizzati a livello internazionale e presuppongono un modello di valutazione del profilo di occupabilità dell’utente dei servizi per l’impiego non automatizzato, che attribuisce all’operatore un ruolo chiave.

Il processo di Assessment parte dalla profilazione quantitativa, misurata in fase di rilascio delle DID o in una fase successiva quando si rende necessario l’aggiornamento del patto di servizio ai nuovi standard di GOL.

La logica dell’algoritmo della profilazione quantitativa è il risultato di un complesso modello statistico che utilizza congiuntamente i risultati della stima della probabilità dell’utente di diventare disoccupato di lunga durata, vale a dire permanere nella condizione di disoccupazione per 12 mesi o più, e la stima della probabilità che nell’arco dei 12 mesi l’individuo non risulti occupato per più di 90 giorni complessivi. L’algoritmo, per ciascun individuo, combina le due probabilità stimate e attribuisce un determinato livello di rischio, con riferimento alla più o meno accentuata difficoltà di reinserimento lavorativo, definito in tre classi: basso, medio e alto. La profilazione quantitativa fornirà dunque agli operatori una prima indicazione sul livello di occupabilità dell’utente, più precisamente, una misura del rischio basso/medio/alto di mancato (re)inserimento lavorativo nell’arco di un anno.

Il modello è descritto nel dettaglio nei paragrafi 3 e 4, ma in termini generali si costruisce partendo dalle caratteristiche anagrafiche dell’utente, quali genere, età, cittadinanza e titolo di studio che vengono integrate con le peculiarità delle esperienze lavorative negli ultimi 2 anni, in termini di durata, qualifica, settore di attività economica, e con la presenza di carichi familiari.

Nella stima si tiene conto anche della dimensione territoriale rappresentata dalla provincia del Cpi titolare della Scheda Anagrafica Professionale (SAP), in questo modo si considera anche la dinamicità del mercato del lavoro locale. Tutte queste informazioni, come è noto, incidono notevolmente sulla maggiore o minore difficoltà di inserimento lavorativo.

Dal punto di vista pratico, l’utente, in fase di rilascio della DID o in fase di calcolo del nuovo profiling per aggiornare il patto di servizio, fornirà solo alcune delle informazioni necessarie al calcolo del profiling quantitativo, in quanto la maggior parte dei dati che verranno utilizzati sono già presenti negli archivi amministrativi (in particolare nelle Schede Anagrafiche Professionali - SAP di competenza delle Regioni e Province Autonome e nell’archivio delle Comunicazioni Obbligatorie di competenza di MLPS).

Questa è una delle principali innovazioni del nuovo sistema di profilazione quantitativa; infatti, il modello statistico è stato stimato sulla popolazione di disoccupati che hanno rilasciato una DID nel 2018 e nel 2019, ricavando le informazioni necessarie dalle diverse componenti del Sistema Informativo Unitario. In questo modo i risultati sono maggiormente aderenti alle caratteristiche degli utenti che si rivolgono ai Centri per l’Impiego e si crea un minore disturbo statistico per l’utente.

Tutte le informazioni che l’algoritmo utilizza per calcolare la classe di rischio saranno messe a disposizione dell’operatore come informazioni preliminari allo svolgimento della profilazione qualitativa, senza creare alcun automatismo nella definizione del profilo dell’utente.

Infatti, una volta compreso il significato della classe di rischio starà agli operatori valorizzare tali indicazioni quando, in sede di colloquio di orientamento, acquisiranno, attraverso un’interazione dinamica con gli utenti,

² Delibera del Commissario Straordinario n. 5 del 9 maggio 2022 - Approvazione strumenti per l’attuazione dell’assessment: profilazione quantitativa, profilazione qualitativa, standard dei servizi di Gol e relative unità di costo standard <https://www.anpal.gov.it/documents/552016/1241390/Delibera+del+Commissario+straordinario+n.+5+del+9+maggio+2022.pdf/8f36dbf-17bb-b81a-15d8-20c5e28f25b2?t=1652191988355>

tutte le ulteriori informazioni necessarie a completare e qualificare lo specifico profilo dell'utente per far emergere i differenziati bisogni in termini di supporto, assistenza e accompagnamento nella ricerca di lavoro, oppure di formazione/riqualificazione delle competenze, o ancora di supporto nel dar risposta a bisogni complessi, attivando la rete dei servizi socio-sanitari territoriali.

Gli operatori in questa fase, hanno a disposizione lo strumento di profilazione qualitativa articolato in domande-guida che hanno la funzione di traccia di riferimento per la conduzione del colloquio, e che permette all'operatore di ricondurre le informazioni acquisite in un format strutturato³.

Lo strumento di profilazione qualitativa si presenta come un questionario articolato in 4 sezioni, ognuna organizzata in 5 ambiti tematici o aree indagate (tra parentesi nell'elenco che segue), ciascuna delle quali rileva uno specifico aspetto in termini di occupabilità.

Le prime 3 sezioni (A-B-C) sono rivolte a tutti gli utenti e riguardano:

- SEZIONE A - CONDIZIONE LAVORATIVA (grado di scolarizzazione, tempo trascorso dall'ultima esperienza di lavoro, percorso professionale, mobilità e spostamenti e lavoro ricercato);
- SEZIONE B – COMPETENZE (competenze comunicative in lingua italiana, competenze linguistiche, competenze digitali, formazione extrascolastica e competenze tecnico-professionali);
- SEZIONE C - CONDIZIONI PERSONALI (condizioni ostacolanti il lavoro/ricerca di lavoro, rete familiare, familiari a carico, condizione abitativa, cura della persona)

La quarta sezione (D- Approfondimento Condizioni Personali) è dedicata solo agli utenti per i quali, in fase di colloquio si rilevino elementi multiproblematici che necessitano di un approfondimento per valutare l'indirizzamento al percorso 4 (Lavoro e inclusione) con l'attivazione di una presa in carico integrata con altri servizi socio sanitari del territorio.

A queste 4 sezioni si aggiunge una Check list per la valutazione professionale, attraverso cui l'operatore approfondirà ulteriori aspetti importanti per l'indirizzamento finale ai percorsi (quali la compatibilità della professione ricercata dall'utente con le richieste del mercato del lavoro locale, la disponibilità a intraprendere percorsi formativi e alla mobilità geografica)

Nella fase conclusiva dell'assessment, quindi, l'operatore, mettendo insieme le indicazioni provenienti dalla profilazione quantitativa e qualitativa, dall'osservazione partecipata dell'utente in fase di colloquio di orientamento, dalla conoscenza del Mercato del Lavoro locale, attraverso la valutazione professionale sarà in grado di individuare il bisogno di cui l'utente è portatore e, di conseguenza, la tipologia di percorso da prevedere all'interno del Patto di Servizio Personalizzato tra quelli previsti dal programma GOL.

- **Percorso 1:** il reinserimento occupazionale: dedicato a coloro che risultano essere più vicini al mercato del lavoro e più facilmente occupabili (work ready).
- **Percorso 2:** aggiornamento (Upskilling): rivolto a coloro che hanno bisogno di adeguare le proprie competenze per inserirsi o re-inserirsi nel mercato del lavoro
- **Percorso 3:** riqualificazione (Reskilling): interessa quegli utenti che risultano nell'insieme più distanti dal mercato del lavoro e per i quali è necessaria una più robusta attività di formazione per avvicinarsi ai profili richiesti dal mercato stesso.
- **Percorso 4:** lavoro e inclusione: riguarda quelle persone per cui la distanza e la difficoltà di inserimento nel mercato del lavoro dipendono da una serie di condizioni e bisogni complessi, che vanno oltre la dimensione lavorativa.

Per mettere gli operatori dei Centri per l'impiego nelle condizioni di utilizzare al meglio gli strumenti di profilazione quantitativa e qualitativa per le finalità del Programma GOL sono stati programmati interventi formativi specifici.

I paragrafi che seguono descrivono nel dettaglio il modello, le stime e la capacità predittiva del nuovo strumento di profilazione quantitativa di GOL, approvato con Delibera del Commissario Straordinario di ANPAL n. 5 del 9 maggio 2022.

³ Per una analisi più dettagliata sul profiling qualitativo si rimanda all'allegato B "Strumenti per l'attuazione dell'assessment – Profilazione qualitativa" della Delibera n. 5 del Commissario Straordinario del 9 maggio 2022, già approvato dal Comitato direttivo del 4 Programma GOL nella riunione del 28 aprile 2022.

3. LO STRUMENTO DI PROFILAZIONE QUANTITATIVO PER L'ASSESSMENT DEGLI UTENTI DEI CPI

L'introduzione della DID (Dichiarazione di Immediata Disponibilità) on-line a partire dal 4 dicembre 2017 ha consentito un netto miglioramento qualitativo e quantitativo del patrimonio informativo dei Centri per l'impiego e più in generale del mercato del lavoro. In particolare, il miglioramento degli archivi amministrativi investe tre aspetti:

- completezza e tempestività delle informazioni sulla disoccupazione amministrativa (sistema DID);
- integrazione tra banche dati amministrative, in particolare i sistemi informativi ANPAL e MLPS-Comunicazioni Obbligatorie (d'ora in avanti MLPS-CO);
- allargamento del patrimonio informativo del disoccupato con le informazioni provenienti dalla profilazione quantitativa così come contemplata dal D.lgs. 150.

Questi tre aspetti rappresentano il filo conduttore alla base della modellizzazione del profiling quantitativo per l'Assessment in GOL degli utenti dei centri per l'impiego. Modellizzazione che il presente documento intende descrivere nel dettaglio.

3.1. Modelli con variabile dipendente dicotomica

Nell'applicazione classica della stima quantitativa del profiling si parte dall'individuazione della variabile dipendente, oggetto di stima, che rileva per ciascun individuo la presenza (=1) o l'assenza (=0) di una determinata condizione in relazione ad un determinato contesto spaziale e temporale. Nel presente contesto la condizione rilevata (presenza/assenza) dalla variabile dipendente sarà strettamente correlata alla difficoltà di (re)inserimento lavorativo.

L'utilizzo di una variabile dipendente dicotomica viene associata ad equazioni di regressione di tipo Logit o Probit. Per ciascun individuo il modello restituisce la probabilità dell'evento positivo⁴ ($y=1$) sulla base di alcune caratteristiche osservate/misurate:

$$p(y) = \vartheta(y = 1 | \mathbf{X}) \quad [1]$$

dove y rappresenta la variabile dipendente, $\vartheta(\dots)$ l'appropriata funzione di probabilità *assunta* dal modello utilizzato (normale nel caso di Probit, logistica nel caso della stima Logit), e \mathbf{X} rappresenta il vettore delle variabili/caratteristiche dell'individuo osservate e misurate in un dato momento temporale.

Tra i punti di forza di una tale modellistica vi è il fatto che i valori stimati sono sempre contenuti nell'intervallo $[0, 1]$ e, inoltre, il valore atteso del profiling stimato è *pari al valore atteso di y* :

$$E[\vartheta(Y = 1 | \mathbf{X})] = E(y)$$

Tra i punti di debolezza vi è comunque il forte legame con la scelta della variabile dipendente: ad esempio, definite due differenti variabili dipendenti (y_1, y_2), entrambe le quali danno una misura oggettiva della difficoltà di (re)inserimento lavorativo, per uno stesso individuo il modello [1] può restituire valori relativamente bassi/alti rispetto ad una variabile e relativamente alti/bassi rispetto all'altra.

3.2. La scelta della variabile dipendente/outcome

Punto di partenza è dunque l'individuazione dell'*outcome*: qual è la variabile indicatrice con cui si vuole misurare la difficoltà di (re)inserimento lavorativo di un soggetto che si reca presso un centro per l'impiego? Compito della variabile dipendente è quello di fornire una *misura* oggettiva del grado di difficoltà di

⁴ Nella presente modellizzazione del profiling l'evento "positivo" (presenza della condizione, $y=1$) è associata alla condizione di disoccupazione e non di occupazione: ad esempio, l'evento positivo corrisponde alla presenza (soddisfazione) della condizione di disoccupato di lunga durata. Questo aspetto può essere contro-intuitivo ma è in linea con le precedenti definizioni di profilazione già in uso.

(re)inserimento lavorativo del disoccupato entro un arco temporale che è stato definito in 365 giorni (un anno). Rispetto a questo intervallo di tempo sono state costruite due variabili dipendenti:

- y_0 : variabile dicotomica che assume valore “1” se l’individuo non ha rapporti di lavoro intervenuti nei 365 giorni successivi al rilascio della DID, e valore “0” altrimenti;
- y_1 : variabile dicotomica che assume valore “1” se l’individuo ha avuto al più 90 giorni complessivi di lavoro contrattualizzato nei 365 giorni dopo il rilascio della DID, estendendo l’orizzonte temporale fino al 454esimo giorno successivo se l’individuo era occupato al 365-esimo giorno per meno di 90 giorni, e valore “0” altrimenti.

Il primo modello (y_0) appare più restrittivo poiché considera una condizione di “estrema” debolezza/difficoltà di (re)inserimento lavorativo: il modello probabilistico stima, infatti, la probabilità che l’ i -esimo individuo, avente determinate caratteristiche descritte dall’insieme di variabili esplicative utilizzate nel modello X , non avrà giorni di lavoro contrattualizzati nell’anno successivo alla data di ingresso nella disoccupazione.

Il secondo modello utilizza la variabile dipendente y_1 che rispetto alla variabile precedente aderisce in maniera meno *radicale* ai concetti di maggiore o minore occupabilità: il modello probabilistico stima, infatti, la probabilità che l’ i -esimo individuo, avente determinate caratteristiche descritte dall’insieme di variabili esplicative utilizzate nel modello X , risulti aver effettuato al più 90 giorni di lavoro contrattualizzato nell’anno successivo.

3.3. Descrittive outcome (y) e variabili esplicative (X): eterogeneità regionali

L’universo di riferimento su cui è stato implementato il nuovo modello di profiling quantitativo è rappresentato dagli individui che hanno sottoscritto una DID negli anni 2018 e 2019. Nel complesso esso è pari a circa 3,4 milioni di individui.

Le principali differenze rispetto alla precedente stima del modello di profiling sono:

- dati amministrativi Vs. survey statistiche (Istat, Forze Lavoro);
- più variabili dipendenti che possono essere scelte per meglio adattarsi a più definizioni di occupabilità Vs. una sola variabile dipendente;
- ampliamento del set di variabili esplicative X

Per ciascun individuo dell’insieme di riferimento è stato possibile (integrazione degli archivi DID con l’archivio dei rapporti di lavoro di fonte MLPS-CO) ricostruire l’insieme dei rapporti di lavoro intervenuti nell’anno successivo alla data di rilascio della DID, e definire in tal modo i due *outcome* [y_0 , y_1] appena descritti.

Le variabili esplicative, come detto, descrivono le caratteristiche dell’individuo *possedute, osservate e misurate* al momento della sottoscrizione della DID. Queste informazioni coprono più dimensioni (anagrafiche, istruzione e formazione, esperienze lavorativa⁵, contesto familiare ecc.), e la loro scelta necessariamente *influisce* sulla bontà del modello utilizzato nel *predire* la più o meno accentuata difficoltà di (re)inserimento lavorativo. La capacità esplicativa delle singole variabili utilizzate nel modello segue un criterio generale di *parsimonia* e di sfruttamento integrale delle informazioni già censite negli archivi ANPAL (DID e SAP) e MLPS-CO.

La scelta di utilizzare i due anni 2018-2019 è maturata durante l’implementazione dei modelli, poiché essa si è mostrata più efficace nel produrre stime con migliori capacità predittive rispetto ai singoli contesti regionali. Il principale ostacolo di una tale scelta è il fatto che si viene in tal modo a coinvolgere nella misura dell’*outcome* (variabile dipendente) l’anno 2020, con tutta la sua eccezionalità derivante dalla crisi pandemica.

In termini relativi dal confronto tra il 2018 e il 2019 emerge per il 2019 un incremento di 4 punti percentuali (9,2%, tavola 3.1) per quel che riguarda la percentuale di LTU (*outcome* y_0) e di altrettanti 4 punti percentuali (8,6%, tavola 3.2) per quel che concerne il secondo *outcome* relativo alla percentuale di individui con al più 90 giorni di lavoro contrattualizzato nell’anno successivo all’ingresso in disoccupazione.

⁵ Alcune variabili, legate per lo più all’esperienza lavorativa ricostruita dalle Comunicazioni Obbligatorie, necessitano in ogni caso di un “calcolo” da Sistema.

A livello regionale, le variazioni in termini relativi più marcate si hanno nelle regioni del Nord ed in particolare nelle regioni più piccole: Valle d'Aosta, Liguria e P.A. di Bolzano.

Tavola 3.1: Quota LTU nei 365 giorni successivi alla DID (y_0). Confronto 2018-2019 per Regione

	Totale	2018	2019	Delta	Delta%
Piemonte	0,46	0,43	0,48	0,05	11,17
Valle d'Aosta	0,38	0,35	0,41	0,06	16,04
Lombardia	0,45	0,41	0,47	0,05	13,04
PA Bolzano	0,20	0,18	0,22	0,04	21,53
Pa Trento	0,24	0,23	0,25	0,03	12,05
Veneto	0,37	0,34	0,39	0,05	15,47
Friuli Venezia Giulia	0,38	0,36	0,40	0,03	9,51
Liguria	0,39	0,34	0,43	0,09	26,73
Emilia Romagna	0,40	0,39	0,41	0,02	4,47
Toscana	0,39	0,37	0,40	0,04	10,78
Umbria	0,47	0,46	0,48	0,02	4,08
Marche	0,43	0,42	0,45	0,03	7,68
Lazio	0,47	0,44	0,49	0,04	10,06
Abruzzo	0,40	0,39	0,41	0,02	5,71
Molise	0,45	0,44	0,46	0,02	5,29
Campania	0,46	0,44	0,48	0,04	8,18
Puglia	0,45	0,43	0,48	0,05	11,47
Basilicata	0,36	0,36	0,36	0,00	1,16
Calabria	0,49	0,47	0,51	0,04	8,93
Sicilia	0,52	0,51	0,54	0,03	5,87
Sardegna	0,36	0,34	0,38	0,04	10,29
Totale	0,44	0,42	0,46	0,04	9,18

Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Tavola 3.2: Quota di persone occupate per al più 90 giorni nei 365 giorni successivi alla DID (y_1). Confronto 2018-2019 per Regione

	Totale	2018	2019	Delta	Delta%
Piemonte	0,55	0,52	0,57	0,05	9,95
Valle d'Aosta	0,48	0,46	0,52	0,06	12,71
Lombardia	0,53	0,50	0,55	0,06	11,76
PA Bolzano	0,27	0,24	0,31	0,07	28,67
Pa Trento	0,34	0,31	0,36	0,05	17,07
Veneto	0,45	0,42	0,49	0,07	17,09
Friuli Venezia Giulia	0,47	0,45	0,49	0,04	9,59
Liguria	0,48	0,42	0,52	0,11	25,96
Emilia Romagna	0,50	0,49	0,51	0,02	5,13
Toscana	0,47	0,45	0,50	0,05	11,89
Umbria	0,55	0,54	0,56	0,02	4,48
Marche	0,53	0,52	0,55	0,03	6,49
Lazio	0,56	0,54	0,58	0,04	7,91
Abruzzo	0,51	0,50	0,53	0,03	5,72
Molise	0,55	0,54	0,57	0,03	5,04
Campania	0,55	0,53	0,57	0,04	7,65
Puglia	0,57	0,54	0,59	0,05	8,68
Basilicata	0,47	0,47	0,47	0,00	0,04
Calabria	0,58	0,56	0,60	0,04	6,62
Sicilia	0,61	0,60	0,63	0,03	5,50
Sardegna	0,47	0,44	0,50	0,06	12,74
Totale	0,53	0,51	0,55	0,04	8,62

Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Per la coorte 2018 il numero medio di giorni lavorati nell'anno successivo è pari a 111,9, superiore di circa 9,8 giorni (8,8%, tavola 3.3) rispetto al dato registrato per la coorte dei disoccupati del 2019.

Tavola 3.3: Giorni medi lavorati nei 365 giorni successivi alla DID. Confronto 2018-2019 per Regione

	Totale	2018	2019	Delta	Delta%
Piemonte	96,60	103,22	90,20	-13,02	-12,61
Valle d'Aosta	111,16	116,72	103,47	-13,25	-11,35
Lombardia	104,48	112,47	99,12	-13,36	-11,87
PA Bolzano	163,85	180,06	145,90	-34,16	-18,97
Pa Trento	158,01	166,10	150,15	-15,94	-9,60
Veneto	124,00	133,30	115,08	-18,22	-13,67
Friuli Venezia Giulia	121,77	127,64	116,05	-11,59	-9,08
Liguria	110,32	122,89	99,66	-23,23	-18,90
Emilia Romagna	110,06	112,53	107,88	-4,65	-4,13
Toscana	118,12	123,87	112,47	-11,40	-9,20
Umbria	100,34	103,99	96,95	-7,03	-6,76
Marche	102,33	105,25	99,47	-5,77	-5,49
Lazio	97,76	102,94	93,00	-9,94	-9,66
Abruzzo	115,42	117,82	113,19	-4,63	-3,93
Molise	105,94	107,68	104,06	-3,62	-3,37
Campania	109,25	113,45	105,21	-8,24	-7,26
Puglia	99,26	103,46	94,35	-9,11	-8,80
Basilicata	123,90	123,09	124,67	1,59	1,29
Calabria	96,45	99,00	93,82	-5,18	-5,24
Sicilia	89,55	92,64	86,08	-6,56	-7,08
Sardegna	118,70	124,55	112,87	-11,68	-9,38
Totale	106,88	111,86	102,06	-9,80	-8,76

Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Affinché si possa ammettere che la differenza nei livelli di *outcome* osservati tra le coorti 2018 e 2019 sia interamente attribuibile al mutato contesto di riferimento (leggi pandemia 2020), occorre che le stesse due coorti siano statisticamente simili rispetto ad un insieme di caratteristiche osservabili sui singoli individui.

Ad eccezione di alcune tra le Regioni più piccole (Basilicata, Valle d'Aosta) riguardo alla composizione per genere della platea dei disoccupati nelle due coorti (cfr. appendice tavola 3.4), le principali differenze si osservano per la Campania (donne +2,6 punti percentuali nel 2019) e la Lombardia (+1,5 p.p. nel 2019).

Anche l'età media per il 2019 (34,4 anni) si mostra significativamente diversa rispetto al 2018 (32,1 anni), con un incremento di +2,3 anni (cfr. appendice tavv. 3.5, 3.6 e 3.7). La dinamica riguarda sia le donne (+2,1 anni) che gli uomini (+2,4 anni). Per quanto riguarda il titolo di studio, il confronto tra il 2018 e il 2019, anche se con variazioni contenute in termini di punti percentuali, vede un generale aumento per il 2019 di disoccupati con titolo di studio basso e una contestuale riduzione dei disoccupati con titolo di studio medio. Questa dinamica è particolarmente evidente nelle regioni del Mezzogiorno (cfr. appendice tavv. 3.8, 3.9 e 3.10).

Rispetto alla presenza di disoccupati con una o più esperienze lavorative nei 12 mesi precedenti la DID, il confronto tra il 2019 e il 2018 mostra un aumento nelle regioni del Nord, in particolare la Lombardia (+16,5 p.p.) e un generale decremento nelle regioni del Mezzogiorno, in particolare in Campania e in Calabria (cfr. appendice tavola 3.11).

Come ultimo confronto si prende in esame la variabile relativa alla presenza di figli all'interno del nucleo familiare del disoccupato. Il 2019 presenta una percentuale sensibilmente più elevata di nuclei con figli con particolare riferimento alle Regioni Lombardia, Emilia-Romagna, Lazio, Calabria e Sicilia (cfr. appendice tavv. 3.12, 3.13 e 3.14).

In conclusione, la struttura delle due coorti di disoccupati 2018 e 2019 presenta delle differenze che diventano più o meno significative se calate nel contesto regionale. La Lombardia è la Regione che presenta le differenze più marcate rispetto alle caratteristiche dei soggetti analizzate.

Tenere insieme le due annualità nella stima del modello consente quindi di mediare le differenze con la conseguenza di ottenere stime più robuste e meno sensibili a variazioni di segno amministrativo che possono avere un peso non trascurabile in alcuni contesti territoriali soprattutto per il 2018.

3.4. Stime dei modelli

Sulla stessa platea di individui e utilizzando lo stesso insieme X di variabili esplicative, sono state implementate e stimate due regressioni logistiche in relazione alle due differenti variabili dipendenti analizzate: probabilità di diventare disoccupato di lunga durata (y_0) ovvero di lavorare per al più 90 giorni (y_1), (cfr. appendice tavv. 3.4.1 e 3.4.2). Il processo restituisce per ciascuna caratteristica individuale (genere, età, titolo di studio, esperienza lavorativa pregressa, ecc.) un coefficiente che indica il peso, e il relativo segno, che la corrispondente caratteristica ha nel determinare la probabilità positiva (=1) della variabile dipendente del modello (cfr. appendice figura 3.5.1).

Come già specificato, la scelta delle caratteristiche osservabili X da utilizzare nel modello si è basata sulla decisione di integrare e valorizzare molte delle informazioni già disponibili negli archivi amministrativi delle DID, delle MLPS-CO e del patrimonio informativo proveniente dalla profilazione quantitativa prevista dal D.lg. 150.

L'utilizzo di archivi amministrativi ha richiesto una fase di pre-processing molto importante, sia per quanto riguarda la presenza di dati mancanti, sia per la valutazione delle incoerenze intrarecord che spesso interessavano anche la stessa informazione registrata in archivi diversi, ma anche nella definizione degli aggregati da inserire nel modello.

Nel dettaglio, le informazioni anagrafiche quali, genere, età, anni presenza in Italia, ma anche il titolo di studio⁶, la condizione occupazionale dell'anno precedente dichiarata e le caratteristiche familiari (numero componenti, presenza di figli e di figli minori) sono quelle raccolte durante la profilazione quantitativa D.lg. 150, mentre le informazioni relative all'eventuale esperienza lavorativa pregressa, in termini anche di contratto, settore, qualifica prevalente e numero datori di lavoro cambiati, sono state ricostruite utilizzando l'archivio del MLPS CO, limitando l'analisi ai 24 mesi precedenti la sottoscrizione della DID, in modo da definire un arco temporale non troppo lontano dalla dichiarazione della disponibilità. Nel caso di nessun rapporto di lavoro alle dipendenze attivato nei due anni precedenti la sottoscrizione della DID, tali variabili sono valorizzate con la categoria "Mai lavorato". L'attività prevalente è determinata in termini di maggiore durata in giorni del rapporto di lavoro.

Dall'archivio delle SAP sono state definite alcune variabili relative alle competenze informatiche e linguistiche, e al possesso di una patente. Solo quest'ultima è risultata però significativa per il modello di stima.

Per quanto riguarda, invece, la ricostruzione dell'ambito territoriale, nel modello entra come variabile di controllo la provincia del CPI competente preferendola al domicilio/residenza dell'individuo poiché più aderente al contesto territoriale del mercato del lavoro di riferimento dell'individuo. La fase di pulizia del dataset di analisi ha, inoltre, previsto l'esclusione di individui: con età inferiore ai 16 anni e superiore ai 64 anni alla data di rilascio della DID; appartenenti alle liste speciali (L. 68); con valori mancanti in alcune variabili, tra cui il titolo di studio, gli anni di presenza in Italia per gli stranieri, e la qualifica prevalente.

Il modello restituisce la quasi totalità dei coefficienti statisticamente significativi, e anche il loro segno conferma quanto osservato a livello descrittivo: il rischio di diventare un disoccupato di lunga durata ($y_0=1$) è maggiore per le donne rispetto gli uomini, si riduce all'aumentare del livello di istruzione e in presenza di esperienze lavorative pregresse, mentre rispetto all'età il rischio di diventare disoccupati di lunga durata è dapprima decrescente per poi aumentare nelle fasce di età più adulte. Per quanto riguarda la dimensione territoriale, i disoccupati che risiedono nel Mezzogiorno hanno maggiore probabilità di diventare disoccupati di lunga durata, benché il dato sia sensibile alle diverse caratteristiche degli utenti dei CPI nelle singole Regioni. Il carico familiare, soprattutto la presenza di figli, aumenta molto il rischio di restare disoccupato di lunga durata per le donne.

⁶ Per il titolo di studio è stata applicata la stessa ricodifica prevista nell'attuale profiling D.lg. 150, ovvero tenendo distinto il tipo di scuola (qualifica professione, diploma istituto professionale, istituto tecnico, liceo) e il corso di laurea (triennale, magistrale, specialistica, vecchio ordinamento) per tipo facoltà.

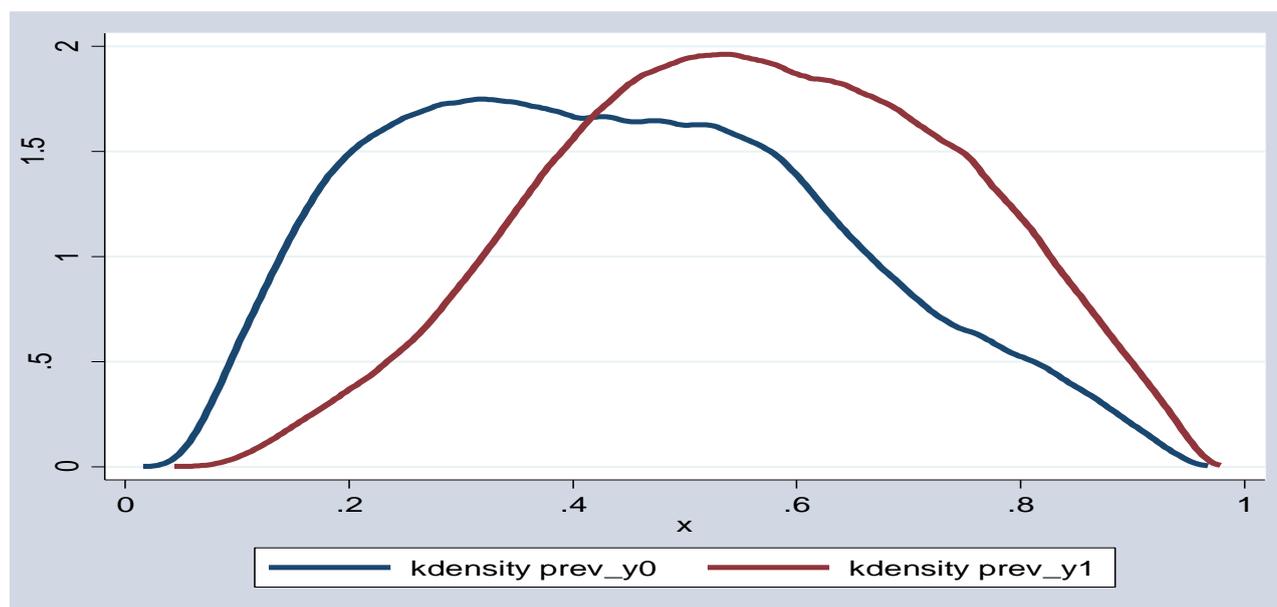
Con leggere differenze, le stesse conclusioni si estendono al caso della variabile y_1 , ovvero al rischio di lavorare al più 90 giorni nell'anno successivo alla DID.

4. UTILIZZO COMBINATO DEI MODELLI Y_0 E Y_1 : IDENTIFICAZIONE DELLE PLATEE PER FASCIA DI RISCHIO

I due modelli, $pr(y_0 = 1 | X)$ e $pr(y_1 = 1 | X)$, condividono come detto le stesse unità di analisi (coorti disoccupati 2018 e 2019) e lo stesso insieme di variabili esplicative.

Considerando le due distribuzioni di probabilità stimate, $\vartheta(y_0 = 1 | X)$, $\vartheta(y_1 = 1 | X)$, si osserva come per la variabile y_0 la forma sia più “panciuta” nella coda sinistra (bassi valori dell'indice), rispetto al modello y_1 che invece presenta una distribuzione più spostata verso la coda destra (figura 4.1).

Figura 4.1 - Distribuzione dei valori $P(y)$ per $y=y_0$ e $y=y_1$



	media	dev. St.	mediana	Q1	Q3
Y_0	0,436	0,197	0,424	0,278	0,578
Y_1	0,528	0,187	0,525	0,385	0,672

Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

L'idea, alla base del presente lavoro, è dunque quella di utilizzare congiuntamente entrambe le distribuzioni per meglio identificare l'insieme degli individui che *con un atteso margine di errore*:

- hanno migliori chances di occupazione (valori relativamente bassi di $p(y_0)$ e $p(y_1)$, ad indicare un più basso rischio di non trovare lavoro nell'anno a venire)
- hanno peggiori chances di occupazione (valori relativamente alti di $p(y_0)$ e $p(y_1)$, ad indicare un più alto rischio di non trovare lavoro nell'anno a venire)

Ogni considerazione/conclusione tratta dal modello vale in “media”, per cui ci si aspetta che individui che hanno un valore più basso dell'indice, per i due modelli, siano potenzialmente più occupabili, abbiano cioè caratteristiche di occupabilità migliori, rispetto ad individui dotati di livelli di probabilità più elevati. Tuttavia, come detto, questo aspetto vale in media mentre la reale “situazione” occupazionale dell'individuo può risentire di situazioni e/o caratteristiche non osservabili o misurabili e dunque non *adeguatamente* coperte dal modello.

Tutti i modelli di stima, siano essi lineari o non lineari, di regressione o di classificazione, sono soggetti ad errori nella capacità predittiva, e ciò che vale in media può tramutarsi in situazioni individuali critiche laddove si procedesse all'applicazione di regole automatiche di classificazione basate sull'indice attribuito al singolo

individuo.

I metodi quantitativi qui analizzati devono necessariamente essere utilizzati a supporto del processo complessivo di *assessment* in cui è rilevante l'integrazione con metodi qualitativi.

Muovendoci all'interno di questo paradigma, qui si propone un utilizzo del profiling quantitativo atto a individuare la parte di platea di disoccupati che si possono ritenere, con un margine di errore atteso ritenuto accettabile, particolarmente forti in termini di occupabilità (work-ready), e la platea di disoccupati per i quali invece si può ritenere che siano più distanti dal mercato di lavoro e dunque con maggiori difficoltà in termini di occupabilità. Queste due coorti si posizionano nelle due code distributive delle funzioni di probabilità stimate per y_0 e y_1 , con una limitata percentuale attesa di popolazione coinvolta.

4.1. Falsi Negativi e Falsi Positivi

Un criterio per valutare la capacità predittiva di un modello probabilistico di classificazione si basa sull'incidenza dei falsi negativi e dei falsi positivi per un dato livello (soglia) della probabilità stimata.

In generale i software per la stima dei modelli Logit/Probit forniscono alcune statistiche di sintesi sulla bontà predittiva del modello. Una di esse si basa sulla quota di individui correttamente classificati rispetto ad un valore soglia di $p(y)$ pari a 0,5: sono considerati correttamente classificati gli individui per i quali il valore stimato di $p(y) \geq 0,5$ e l'*outcome* osservato è pari a 1, e gli individui per i quali il valore stimato di $p(y) < 0,5$ e l'*outcome* osservato è pari a 0. Sulla base di questa statistica, complessivamente i due modelli classificano correttamente circa i due terzi della popolazione analizzata (tavola 4.1.1).

Tavola 4.1.1 – Statistiche sull'accuratezza dei modelli Logit per y_0 e y_1

Classified + if predicted $\Pr(D) \geq 0,5$

True D defined as $y \neq 0$

		y0	y1
Sensitivity	$\Pr(+D)$	55,83%	69,10%
Specificity	$\Pr(\sim D)$	76,76%	61,90%
Positive predictive value	$\Pr(D+)$	65,01%	67,03%
Negative predictive value	$\Pr(\sim D-)$	69,19%	64,12%
False + rate for true $\sim D$	$\Pr(+\sim D)$	23,24%	38,10%
False - rate for true D	$\Pr(-D)$	44,17%	30,90%
False + rate for classified +	$\Pr(\sim D+)$	34,99%	32,97%
False - rate for classified -	$\Pr(D-)$	30,81%	35,88%
Correctly classified		67,63%	65,70%
AUC (ROC curve)		0,729	0,715

Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Un ulteriore criterio per valutare la capacità predittiva del modello si basa sull'incidenza dei cosiddetti falsi negativi e falsi positivi (vedi oltre) al variare dei valori soglia di $p(y)$ da 0 a 1.

Definiamo con

$$\tau_i = \frac{i}{100} \quad (\text{per } i = 1, \dots, 100) \quad [2]$$

i valori soglia mobili della probabilità stimata (livello di profiling), valori che vanno da 0,01 (per $i = 1$) a 1 (per $i=100$), si incrementa cioè il valore soglia di un punto percentuale al crescere di i .

Per facilitare la scrittura indichiamo con " y " il generico *outcome* (variabile dipendente) che assume i valori 0 (negativo) e 1 (positivo). Indichiamo inoltre con $p(y)$ il valore (profiling) stimato dal modello logit che, come detto, può essere visto come una probabilità di osservare l'*outcome* positivo di y :

$$p(y) = \text{pr}(y=1|X)$$

Falsi negativi. Per ogni unità-individuo per cui vale:

$$p(y) \leq \tau_i,$$

viene fatta l'assunzione di ritenere l'individuo come un potenziale negativo rispetto ad y , cioè:

$$\text{Se } p(y) \leq \tau_i \rightarrow E(y)=0 \quad [3]$$

Si definisce "Falso negativo", per un dato valore soglia τ_i , l'unità-individuo che verifica la condizione:

$$p(y) \leq \tau_i \text{ e } y=1 \quad [4]$$

Chiaramente all'aumentare del valore soglia τ_i , cioè al crescere di i , aumenta il numero dei falsi negativi. In particolare, per $\tau_i = 1$, il numero dei falsi negativi corrisponde al numero complessivo dei positivi ($y=1$) nella popolazione. Definiamo come indice di tolleranza dei falsi negativi la quantità:

$$\alpha_i = \frac{\sum_{j \in D_i} I(p(y) \leq \tau_i; y=1)}{\sum_j I(D_i)} \quad [5]$$

Dove D_i rappresenta l'insieme caratterizzato da $p(y) \leq \tau_i$ e l'indice j identifica l'unità-individuo della popolazione. La [5] rappresenta l'incidenza dei falsi negativi (numeratore del rapporto) rispetto alla popolazione complessiva che soddisfa la condizione $p(y) \leq \tau_i$. Anche il valore α_i , presenta in generale un andamento crescente al crescere del valore soglia τ_i , fino ad arrivare al valore $E(y)$ per $i=1$.

Falsi positivi: Per ogni unità-individuo per cui vale $p(y) > \tau_i$, viene fatta l'assunzione di un *outcome* (y) positivo cioè:

$$\text{Se } p(y) > \tau_i \rightarrow E(y)=1 \quad [6]$$

Si definisce "Falso positivo" per un dato valore soglia τ_i , l'unità-individuo per la quale si verifica

$$p(y) > \tau_i \text{ e } y=0 \quad [7]$$

Chiaramente all'aumentare dei valori soglia τ_i , cioè al crescere di i , si riduce il numero dei falsi positivi. Per $\tau_i \sim 1$, il numero dei falsi positivi tende a zero. Definiamo come indice di tolleranza dei falsi positivi la quantità:

$$\gamma_i = \frac{\sum_{j \in G_i} I(p(y) > \tau_i; y=0)}{\sum_j I(G_i)} \quad [8]$$

Dove G_i rappresenta l'insieme caratterizzato da $p(y) > \tau_i$ e l'indice j identifica l'unità-individuo della popolazione. La [8] rappresenta l'incidenza dei falsi positivi (numeratore del rapporto) rispetto alla popolazione complessiva che soddisfa la condizione $p(y) \geq \tau_i$. Il valore γ_i , presenta un andamento decrescente, non necessariamente monotono, rispetto a τ_i , fino ad arrivare al valore 0 per $i=1$.

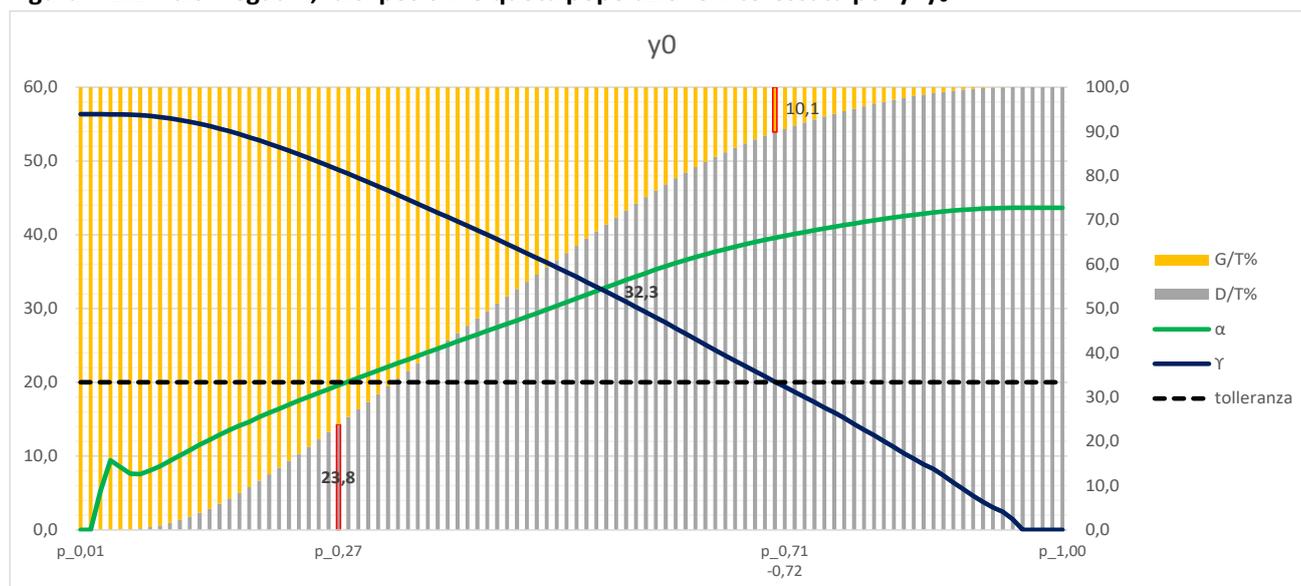
La figura 4.1.1 riproduce i valori α_i e γ_i , in base 100 (asse sx) e i valori della quota di popolazione interessata, sempre in base 100 (asse dx), al variare della soglia τ_i (asse delle ascisse). La quota di popolazione interessata

è pari al numero di unità in D_i (vale la condizione $p(y) \leq \tau_i$) sul totale della popolazione (istogramma grigio) e al numero di unità in G_i (vale la condizione $p(y) > \tau_i$) sul totale della popolazione (istogramma giallo). L'interesse per valori bassi di τ_i è rivolto ai falsi negativi (α_i e D_i/T), mentre per valori alti di τ_i l'interesse si sposta sui falsi positivi (γ_i e G_i/T). Che cosa osserviamo se fissiamo un valore comune per gli indici di tolleranza α_i e γ_i ? Supponiamo di fissare i margini di tolleranza ad un valore massimo di 0,2 (20%). Per quali valori di τ_i possiamo attenderci di osservare un'incidenza di falsi negativi e di falsi positivi, rispettivamente, non superiore al 20%?

Tracciando una retta parallela all'asse delle ascisse di ordinata (asse sinistra) pari a 0,2, si ottengono i valori cercati τ_i , nel punto di intersezione di questa retta con le rette α_i e γ_i rispettivamente.

La figura 4.1.1, in corrispondenza del margine di errore fissato a 0,2, mostra come per $y=y_0$ (LTU) i valori τ_i corrispondono rispettivamente a 0,27 e 0,71. Detto in altri termini, se volessimo considerare come "negativi" ($y_0=0$)⁷ gli individui che presentano un indice di profiling $pr(y_0) \leq 0,27$, allora ci si deve attendere che circa una persona su cinque (20%) sarebbe in realtà erroneamente classificata, potendo risultare invece come caso "positivo" ($y_0=1$); mentre se volessimo considerare come "positivi" ($y_0=1$) gli individui che presentano un indice di profiling $pr(y_0) > 0,71$, allora ci si deve attendere che circa una persona su cinque (20%) sarebbe in realtà erroneamente classificata, potendo risultare invece come caso "negativo" ($y_0=0$).

Figura 4.1.1: Falsi negativi, falsi positivi e quota popolazione interessata per $y=y_0$

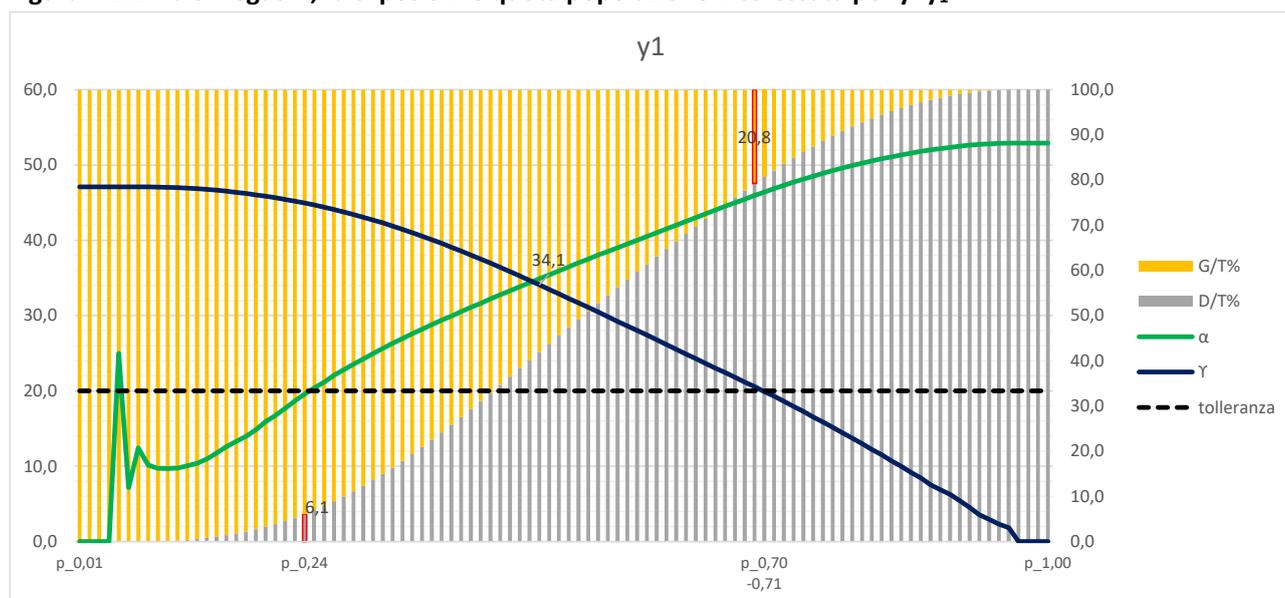


Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Allo stesso modo la figura 4.1.2 riporta l'andamento delle grandezze α_i , D_i/T , γ_i e G_i/T per l'outcome y_1 . I valori soglia per y_1 sono pari rispettivamente a 0,24 e 0,70.

⁷ L'accezione "negativo" attiene all'outcome utilizzato e dunque in questo caso i "negativi" sono gli individui più forti in termini di occupabilità, ovvero nella fattispecie individui con una bassa probabilità di essere LTU.

Figura 4.1.2: Falsi negativi, falsi positivi e quota popolazione interessata per y_1



Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Il confronto tra i due *outcome* mostra come, a parità di margine di tolleranza, per y_1 si ha un valore inferiore della popolazione più forte (D/T), pari al 6,1% contro il 23,8% visto per y_0 , e un valore superiore della quota di popolazione debole (G/T) pari al 20,8% contro il 10,1% osservato per y_0 .

4.2. Definizione delle fasce di rischio

Come visto nel paragrafo precedente, il modello LTU (y_0) consente, fissato un margine di errore atteso, di racchiudere una percentuale più alta di popolazione (23,8%) che può essere considerata più forte in termini di occupabilità, e una percentuale più bassa di popolazione (10,1%) che può essere classificata estremamente debole (tavola 4.2.1).

Tavola 4.2.1: Stima valore $p(y)$ per falsi negativi e falsi positivi, con tolleranza 20%. Modello y_0 e y_1

Modello	Falsi Negativi		Falsi Positivi	
	P(y)	Pop%	P(y)	Pop%
y_0	0,27	23,8	0,71	10,1
y_1	0,24	6,1	0,70	20,8

Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Ribaltando le stesse percentuali sul modello y_1 si ha una situazione opposta: una percentuale maggiore della popolazione (20,8%) è classificabile come più debole, e una parte esigua della popolazione (6,1%) è classificabile come più forte. È possibile allora sfruttare l'informazione congiunta dei due modelli per arrivare a degli insiemi *più* robusti, cercando il giusto compromesso tra il margine di errore e la percentuale di popolazione coinvolta. In particolare, volendo partire dal modello base $y=y_0$ (LTU), l'obiettivo è quello di *individuare* dalla popolazione $D_{27}(y_0)$ ⁸ coloro i quali, anche con riferimento alla variabile y_1 , appartengono alla parte sinistra della distribuzione del profiling, e analogamente a partire da $G_{71}(y_0)$ individuare la parte di popolazione che appartiene anche alla coda distributiva di destra della variabile y_1 .

⁸ In accordo con le notazioni utilizzate nel paragrafo 4.1 il termine D_{27} identifica la popolazione le cui unità-individuo soddisfano la condizione $pr(y_0=1|X) \leq 0,27$, ovvero il 23,8% di individui per i quali il livello di profiling stimato è non superiore a 0,27.

Livello basso di rischio (di non occupazione) (Ready to work)

Punto di partenza: $D_{27}(y_0) = \{j: pr(y_0) \leq 0,27\}$

La popolazione così individuata può mascherare tra i “veri” negativi, situazioni di “probabili” positivi rispetto alla variabile y_1 , e dunque situazioni comunque deboli in termini di occupabilità. Vale a dire, il fatto di non essere LTU potrebbe comunque associarsi a situazioni di scarsa occupazione ovvero presentare valori relativamente alti di $pr(y_1=1)$. Possiamo allora combinare le due distribuzioni di probabilità, $pr(y_0)$ e $pr(y_1)$, per meglio definire la categoria dei “work-ready”.

La proposta è quella di individuare la popolazione “work-ready” nella intersezione tra $D_{27}(y_0)$ e $D_{p^*}(y_1)$ in cui, in accordo con le notazioni già utilizzate, $D_{p^*}(y_1)$ rappresenta la popolazione con un livello di profiling stimato $pr(y_1) \leq 0,36$. Questo valore è tale che $D_{36}(y_1)/T \leq D_{27}(y_0)/T = 23,8\%$.

Complessivamente la popolazione attesa work-ready è pari a circa il 19,6% della platea complessiva degli utenti CPI che rilasciano la DID in un anno.

Livello alto di rischio (di non occupazione) alto

Il modello LTU (y_0), come detto, stima come casi positivi ($y_0=1$) situazioni di estrema difficoltà di inserimento lavorativo. Per questo motivo la coda destra della distribuzione di $pr(y_0)$ appare di per sé sufficiente per selezionare i casi più difficili. In ottica conservativa si può prendere a riferimento la popolazione $G_{71}(y_0)$ come fascia di popolazione più problematica, pari al 10,1% della popolazione complessiva. Un approccio meno conservativo potrebbe portare a considerare la popolazione $G_{70}(y_1)$ che invece raccoglie circa il 20,8% della popolazione. Nelle conclusioni (cfr. capitolo 5) verrà considerato quest’ultimo caso.

Livello di rischio (di non occupazione) medio

Tutti i restanti casi, non ricompresi nei precedenti, rientrano nella fascia intermedia di indeterminazione.

5. CONCLUSIONI

La finalità del nuovo sistema di profilazione quantitativa è di fornire all’operatore del CPI una prima indicazione sul livello di occupabilità dell’utente⁹, sulla base del rischio di diventare disoccupato di lunga durata. A tal fine sono definite tre classi di rischio:

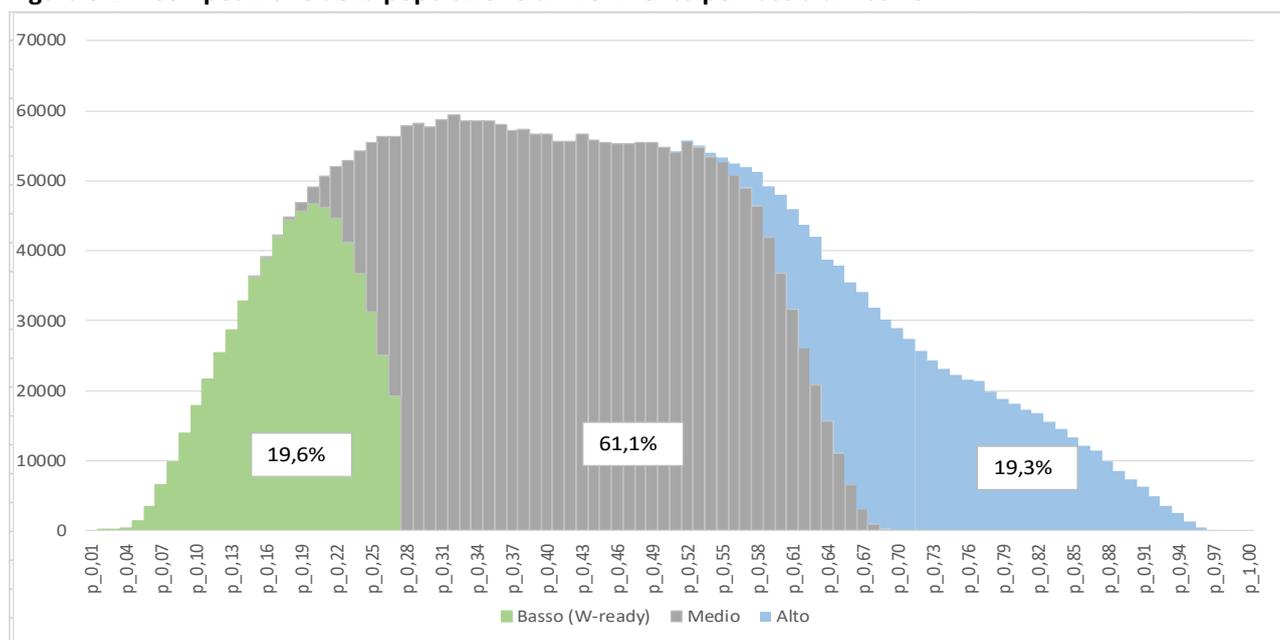
- 1 – basso (rischio)
- 2 – medio (rischio)
- 3 – alto (rischio)

In particolare, nella classe “1 – Basso” rientrano gli utenti per i quali la combinazione dei livelli di profiling dei due modelli y_0 e y_1 descrivono la situazione di “work-ready”. Nella classe “2 – Medio” rientrano gli utenti per i quali la combinazione dei livelli di profiling dei due modelli y_0 e y_1 dà luogo ad indeterminazione. Infine, nella classe “3 – Alto” rientrano gli utenti per i quali la combinazione dei livelli di profiling dei due modelli y_0 e y_1 dà luogo a situazioni di maggiore debolezza intesa come distanza dal mercato del lavoro.

Combinazione $p(y_0)$ e $p(y_1)$		Classe di profilazione
$P(y_0) \leq 0,27$	$p(y_1) \leq 0,36$	1. Basso
	$p(y_1) > 0,36$	2. Medio
$0,27 < P(y_0) \leq 0,71$	$p(y_1) \leq 0,70$	
$0,27 < P(y_0) \leq 0,71$	$p(y_1) > 0,70$	3. Alto
$p(y_0) > 0,71$		

⁹ Il nuovo sistema di profilazione quantitativa è soltanto uno dei complessivi strumenti di cui si compone la fase di assessment dell’utente. Non c’è dunque automatismo alcuno tra l’indice di profiling quantitativo e la definizione delle politiche e dei percorsi dell’utente.

Figura 5.1 – Composizione della popolazione di riferimento per fascia di rischio



Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Se guardiamo ai risultati della combinazione dei modelli per fascia di rischio (fig. 5,1), il 19,8% dell'intera popolazione dei disoccupati che rilasciano una DID in base all'indice di profilazione, risulta work ready, in quanto ricade nella classe di rischio bassa. Mentre il 19,3% ricade in quella alta, si tratta delle persone più lontane dal mercato del lavoro e quindi con caratteristiche di fragilità elevate. Nella classe intermedia ricade il 61% della popolazione.

Il modello presentato è stato adottato su tutto il territorio nazionale da Giugno 2022 e sarà oggetto di specifiche attività di monitoraggio e di aggiornamento anche sulla base dell'arricchimento del patrimonio informativo derivante dalla sperimentazione dello strumento di profilazione qualitativa.

Il paragrafo successivo presenta ulteriori spunti interpretativi dei risultati del nuovo profiling quantitativo per fascia di rischio rispetto ai diversi contesti territoriali e ai principali target del Programma GOL.

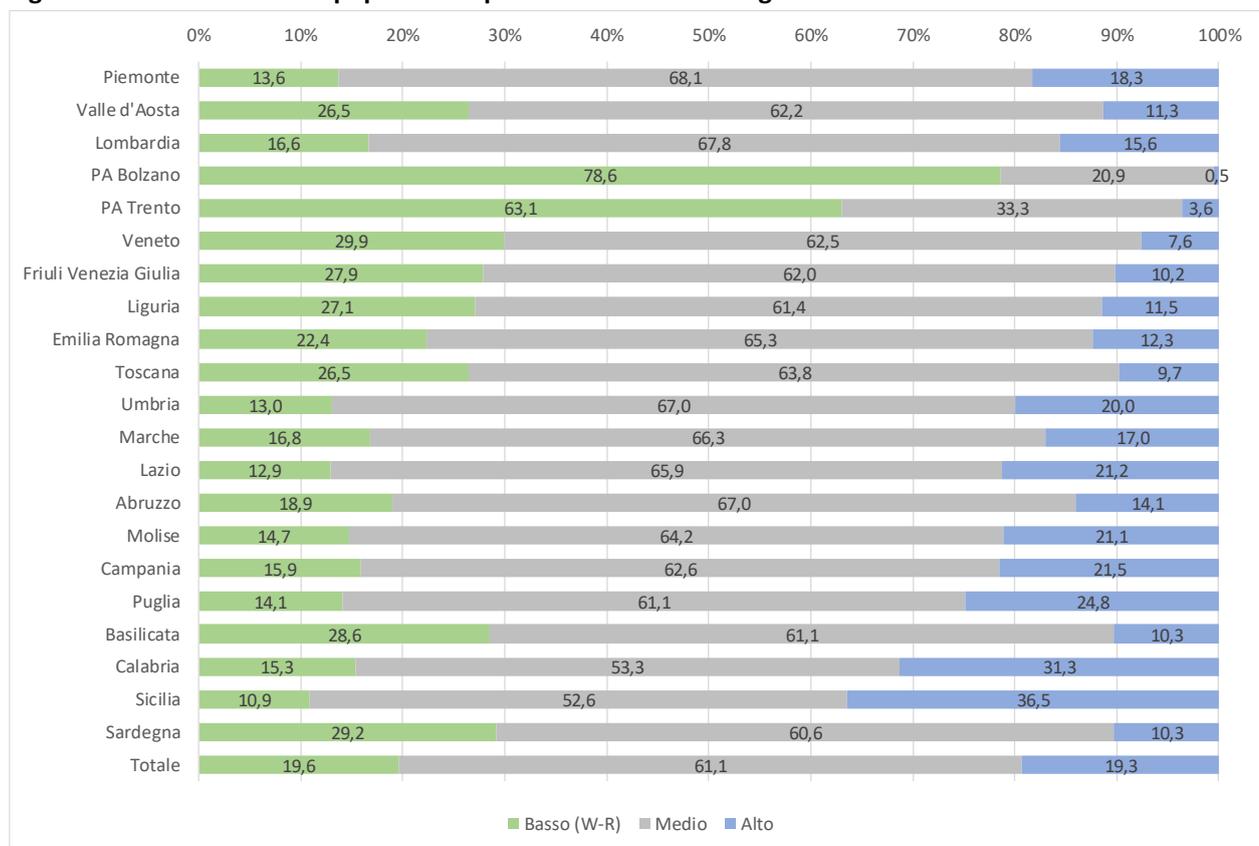
5.1. Simulazione – per alcuni strati della popolazione

In questo paragrafo si presentano le simulazioni relative al peso che ciascuna delle tre aree di rischio assume per alcune specifiche sotto-popolazioni. In particolare, si farà riferimento ai singoli contesti regionali e a specifici target del Programma GOL: donne, giovani under 30, beneficiari di Naspi, beneficiari di RDC, disoccupati da oltre 6 mesi e adulti Over 55.

Cominciando dai contesti regionali, la simulazione mostra una situazione molto eterogenea.

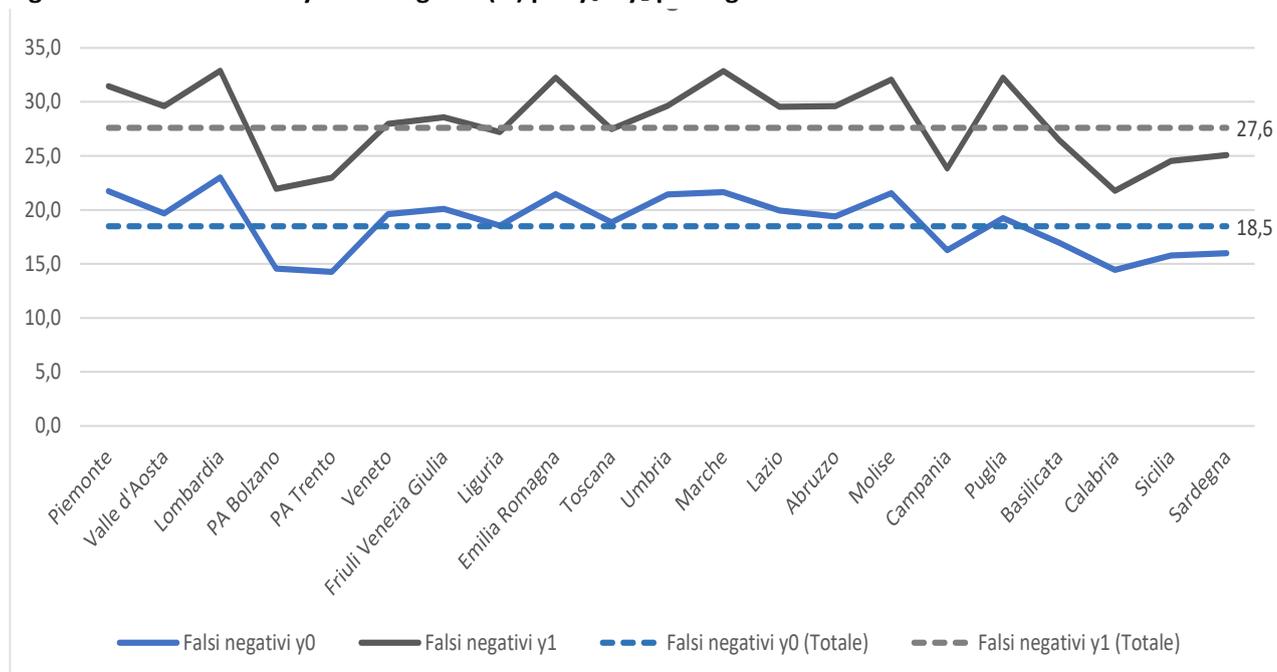
Prendendo ad esempio la popolazione a più basso rischio di disoccupazione (work ready), si passa da un minimo del 13% (Piemonte e Umbria) ad un massimo del 78,6% per la PA di Bolzano, o comunque trascurando le regioni più piccole, si arriva a valori prossimi al 30% nelle regioni Veneto e Sardegna. La percentuale di falsi negativi oscilla tra il 14% per le Province Autonome e il 23,2% per la Lombardia.

Figura 5.1.1 – Distribuzione popolazione per fascia di rischio e regione.



Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Figura 5.1.2 – Work Ready: Falsi negativi (%) per y₀ e y₁ per regione.



Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

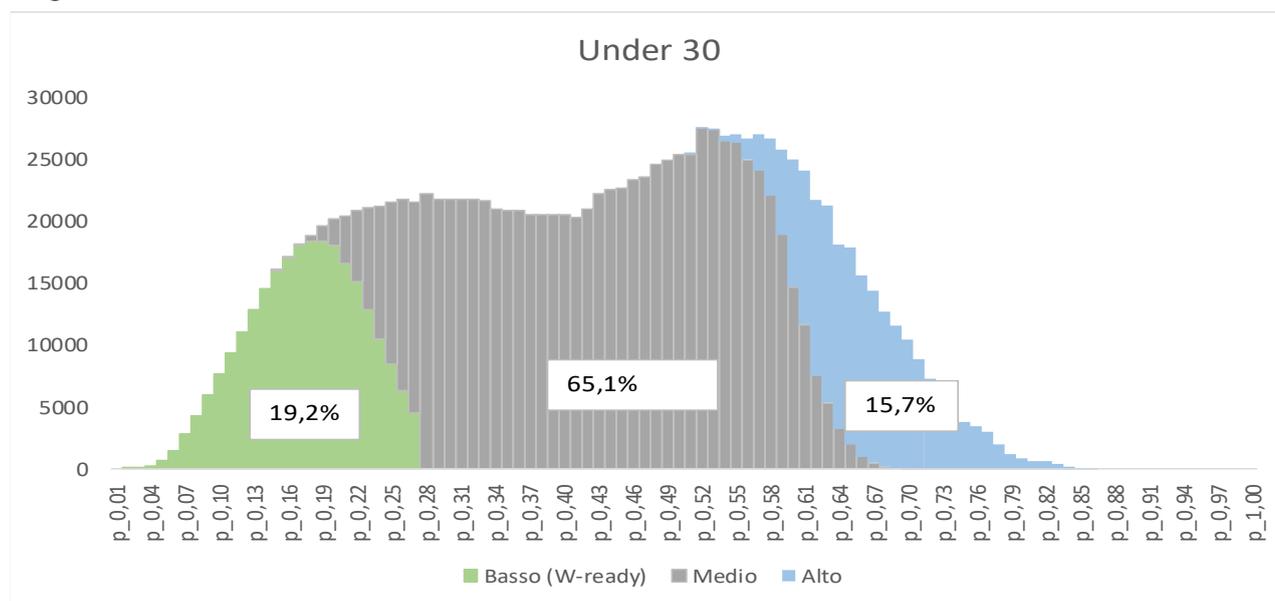
Consideriamo alcuni target del Programma GOL. L'individuazione dei gruppi target all'interno della platea dei disoccupati amministrativi avverrà per il tramite di proxy.

Per ciascun target è stata indicata la quota di popolazione in esso ricompresa, la suddivisione nelle tre aree di rischio - Basso, medio e alto.

In particolare il target dei beneficiari di Naspi presenta una percentuale di work ready molto superiore alla media dei disoccupati e una bassissima percentuale di soggetti ricadenti nella classe di rischio elevata. Questo perché questa tipologia di utenti è caratterizzata da esperienze di lavoro recenti che possono essere capitalizzate per un successivo e rapido reinserimento lavorativo.

Caratteristiche opposte hanno i disoccupati over 55 e i beneficiari di Reddito di Cittadinanza, che, sappiamo, rappresentano due tra i target più critici dal punto di vista dell'inserimento lavorativo. E il modello di profilazione quantitativa coglie molto bene questa criticità e infatti, circa il 50% dei beneficiari di reddito di cittadinanza e il 39% dei disoccupati over 55 ricade nella classe di rischio alta.

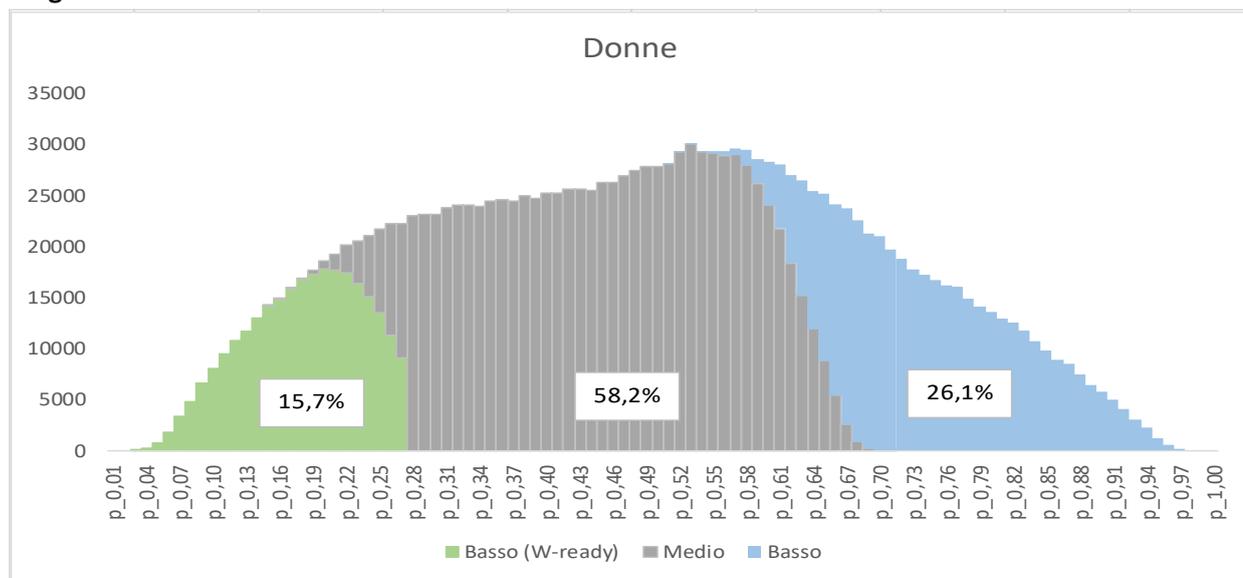
Target 1: Giovani under 30



	% su pop T	% Pop su pop U 30	Mean pr(y ₀)	Mean pr(y ₁)
Basso (W-R)	38,1	19,1	0,170	0,273
Medio	41,5	65,1	0,427	0,543
Alto	31,9	15,7	0,657	0,760
Totale	39,0	100,0	0,414	0,525

Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

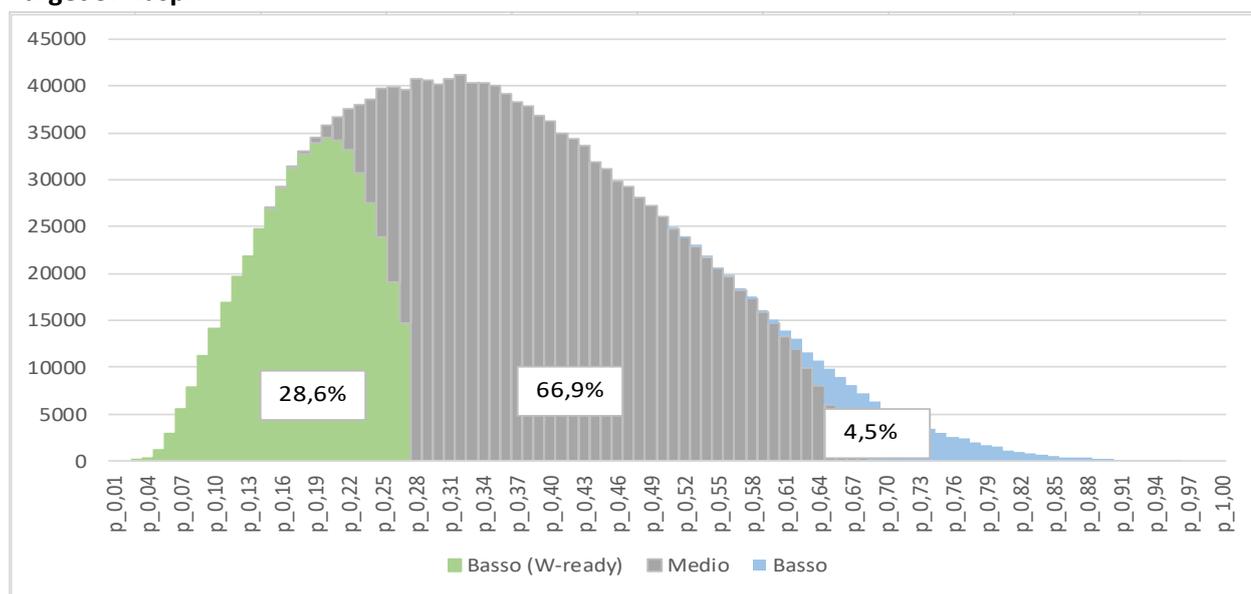
Target 2: Donne



	% su pop T	% Pop su pop Donne	Mean pr(y ₀)	Mean pr(y ₁)
Basso (W-R)	40,6	15,7	0,177	0,252
Medio	48,3	58,2	0,447	0,542
Alto	68,5	26,1	0,738	0,801
Totale	50,7	100,0	0,481	0,564

Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

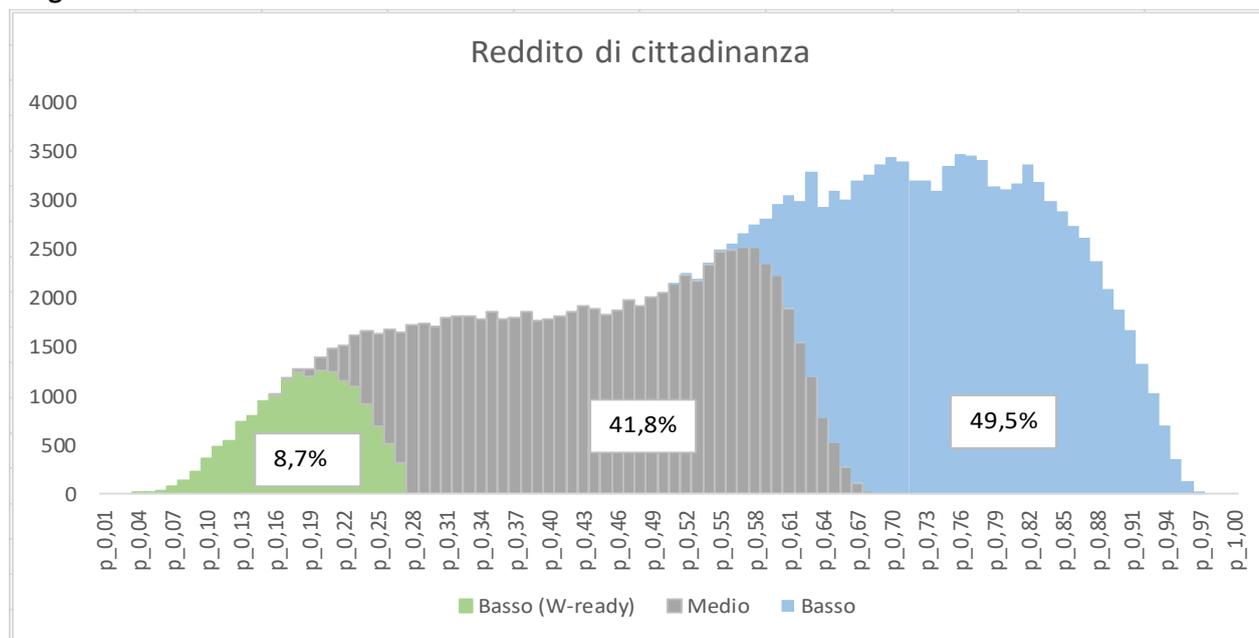
Target 3: Naspi



	% su pop T	% Pop su pop Naspi	Mean pr(y ₀)	Mean pr(y ₁)
Basso (W-R)	75,4	28,6	0,177	0,263
Medio	56,5	66,9	0,407	0,503
Alto	12,0	4,5	0,707	0,765
Totale	51,6	100,0	0,355	0,446

Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

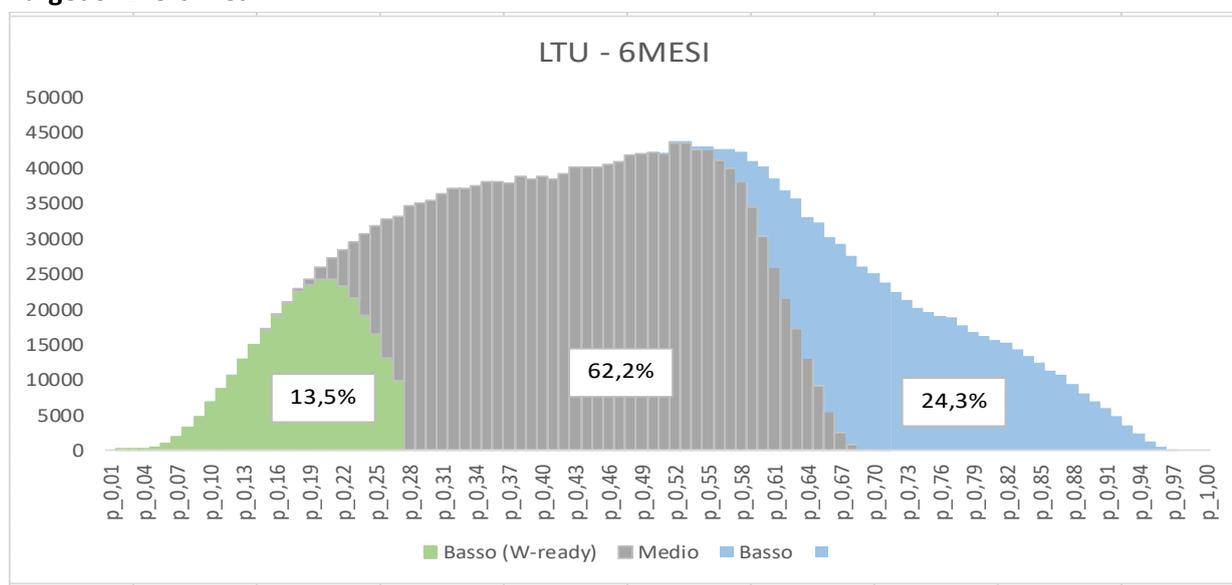
Target 4: Reddito di cittadinanza



	% su pop T	% Pop su pop RdC	Mean pr(y ₀)	Mean pr(y ₁)
Basso (W-R)	N.d.	8,7	0,180	0,287
Medio	N.d.	41,8	0,446	0,555
Alto	N.d.	49,5	0,760	0,820
Totale	N.d.	100,0	0,578	0,663

Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

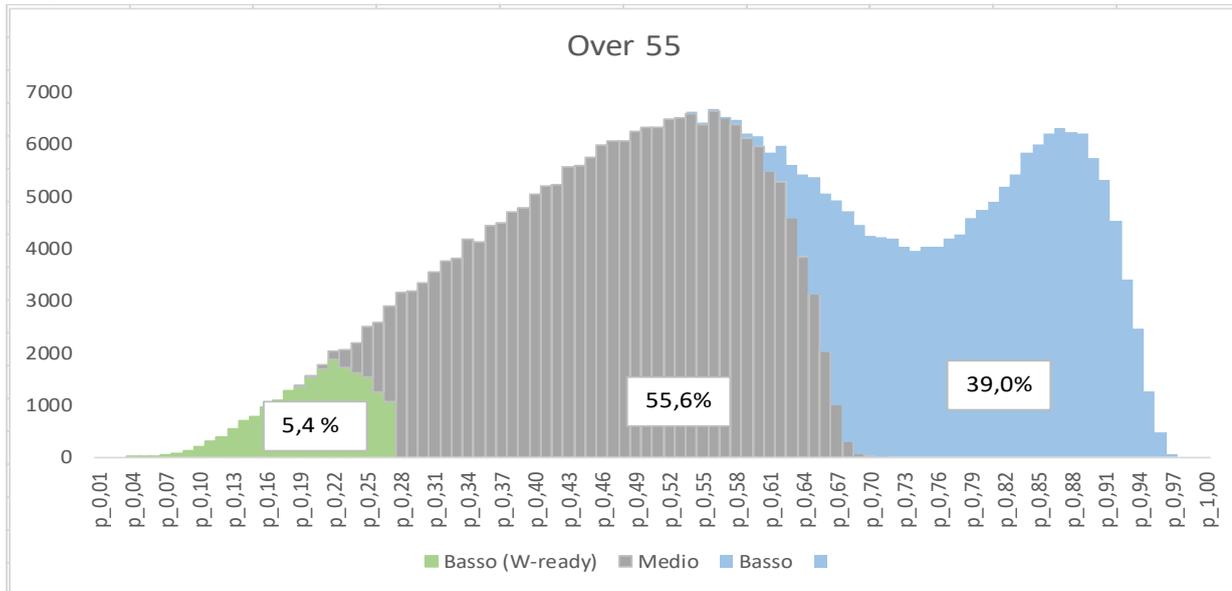
Target 5: LTU 6 mesi



	% su pop T	% Pop su pop Ltu	Mean pr(y ₀)	Mean pr(y ₁)
Basso (W-R)	48,6	13,5	0,183	0,281
Medio	71,6	62,2	0,438	0,539
Alto	88,7	24,3	0,730	0,798
Totale	70,4	100,0	0,475	0,567

Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Target 6: Over 55



	% su pop T	% Pop su pop Over55	Mean pr(y ₀)	Mean pr(y ₁)
Basso (W-R)	3,0	5,4	0,198	0,282
Medio	10,0	55,6	0,471	0,557
Alto	22,2	39,0	0,796	0,841
Totale	11,0	100,0	0,583	0,653

Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

APPENDICE

TAVOLE STASTICHE CAP. 3

Tavola 3.4: Quota di donne nella platea di riferimento.

Differenze 2018-2019 per Regione

	2018	2019	Diff	SE	t	P(t >t)
Piemonte	0,525	0,522	-0,0034	0,00	-1,58	0,11
Valle d'Aosta	0,509	0,530	0,0206	0,01	1,76	0,08
Lombardia	0,504	0,519	0,0149	0,00	7,83	0,00
PA Bolzano	0,574	0,583	0,0088	0,00	1,82	0,07
Pa Trento	0,573	0,576	0,0032	0,00	0,70	0,49
Veneto	0,568	0,553	-0,0141	0,00	-6,88	0,00
Friuli Venezia Giulia	0,545	0,556	0,0105	0,00	2,80	0,01
Liguria	0,550	0,546	-0,0043	0,00	-1,28	0,20
Emilia Romagna	0,538	0,542	0,0048	0,00	2,22	0,03
Toscana	0,549	0,549	0,0005	0,00	0,25	0,80
Umbria	0,556	0,555	-0,0009	0,00	-0,20	0,84
Marche	0,530	0,538	0,0085	0,00	2,52	0,01
Lazio	0,494	0,499	0,0040	0,00	1,79	0,07
Abruzzo	0,473	0,479	0,0059	0,00	1,66	0,10
Molise	0,437	0,455	0,0179	0,01	2,40	0,02
Campania	0,455	0,478	0,0223	0,00	14,36	0,00
Puglia	0,467	0,485	0,0181	0,00	9,80	0,00
Basilicata	0,438	0,464	0,0264	0,00	5,59	0,00
Calabria	0,495	0,505	0,0097	0,00	4,07	0,00
Sicilia	0,462	0,479	0,0167	0,00	11,20	0,00
Sardegna	0,493	0,493	-0,0002	0,00	-0,08	0,93

Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Tavola 3.5: Età media nella platea di riferimento.

Differenze 2018-2019 per Regione

Età	2018	2019	Diff	SE	t	P(t >t)
Piemonte	35,791	36,237	0,4463	0,06	7,88	0,00
Valle d'Aosta	36,546	36,960	0,4141	0,31	1,34	0,18
Lombardia	32,099	34,359	2,2602	0,05	45,48	0,00
PA Bolzano	37,334	36,972	-0,3628	0,12	-2,99	0,00
Pa Trento	37,225	36,871	-0,3543	0,12	-2,95	0,00
Veneto	37,086	37,103	0,0172	0,05	0,32	0,75
Friuli Venezia Giulia	37,015	37,087	0,0722	0,10	0,73	0,47
Liguria	38,542	38,456	-0,0857	0,09	-0,99	0,32
Emilia Romagna	36,061	36,781	0,7199	0,06	12,56	0,00
Toscana	36,801	37,042	0,2407	0,05	4,48	0,00
Umbria	37,442	37,035	-0,4068	0,12	-3,43	0,00
Marche	36,223	36,042	-0,1806	0,09	-2,05	0,04
Lazio	34,955	35,708	0,7536	0,06	12,96	0,00
Abruzzo	36,576	36,893	0,3166	0,09	3,41	0,00
Molise	37,865	37,295	-0,5706	0,19	-2,95	0,00
Campania	35,572	35,666	0,0940	0,04	2,31	0,02
Puglia	34,799	34,818	0,0183	0,05	0,38	0,70
Basilicata	37,382	38,044	0,6627	0,13	5,22	0,00
Calabria	37,057	37,288	0,2306	0,06	3,80	0,00
Sicilia	36,193	36,704	0,5106	0,04	13,13	0,00
Sardegna	36,885	37,105	0,2200	0,07	3,05	0,00

Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Tavola 3.6: Età media per le donne nella platea di riferimento.**Differenze 2018-2019 per Regione**

Età Donne	2018	2019	Diff	SE	t	P(t >t)
Piemonte	36,652	36,926	0,2739	0,08	3,57	0,00
Valle d'Aosta	37,575	37,719	0,1445	0,42	0,34	0,73
Lombardia	32,772	34,893	2,1212	0,07	31,17	0,00
PA Bolzano	38,187	37,711	-0,4757	0,16	-2,98	0,00
Pa Trento	37,835	37,552	-0,2824	0,15	-1,83	0,07
Veneto	37,807	37,746	-0,0619	0,07	-0,88	0,38
Friuli Venezia Giulia	37,831	37,968	0,1374	0,13	1,06	0,29
Liguria	39,330	39,280	-0,0503	0,11	-0,44	0,66
Emilia Romagna	37,034	37,669	0,6349	0,08	8,24	0,00
Toscana	37,712	37,969	0,2574	0,07	3,63	0,00
Umbria	38,509	38,265	-0,2442	0,15	-1,58	0,11
Marche	37,233	36,975	-0,2578	0,12	-2,19	0,03
Lazio	35,178	36,062	0,8849	0,08	11,16	0,00
Abruzzo	36,848	37,229	0,3809	0,13	2,91	0,00
Molise	37,271	37,027	-0,2441	0,27	-0,89	0,37
Campania	35,331	35,691	0,3603	0,06	6,24	0,00
Puglia	34,640	34,710	0,0697	0,07	1,05	0,29
Basilicata	37,437	38,841	1,4034	0,19	7,57	0,00
Calabria	37,380	37,687	0,3075	0,08	3,72	0,00
Sicilia	36,065	36,910	0,8448	0,05	15,48	0,00
Sardegna	37,157	37,369	0,2118	0,10	2,12	0,03

Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Tavola 3.7: Età media per gli uomini nella platea di riferimento.**Differenze 2018-2019 per Regione**

Età Uomini	2018	2019	Diff	SE	t	P(t >t)
Piemonte	34,838	35,486	0,6475	0,08	7,75	0,00
Valle d'Aosta	35,479	36,105	0,6261	0,45	1,38	0,17
Lombardia	31,416	33,784	2,3680	0,07	32,67	0,00
PA Bolzano	36,187	35,939	-0,2473	0,18	-1,34	0,18
Pa Trento	36,408	35,945	-0,4631	0,19	-2,43	0,02
Veneto	36,139	36,307	0,1678	0,08	2,02	0,04
Friuli Venezia Giulia	36,037	35,985	-0,0517	0,15	-0,34	0,73
Liguria	37,576	37,464	-0,1118	0,13	-0,84	0,40
Emilia Romagna	34,931	35,729	0,7985	0,09	9,38	0,00
Toscana	35,694	35,912	0,2180	0,08	2,67	0,01
Umbria	36,103	35,498	-0,6050	0,18	-3,32	0,00
Marche	35,086	34,956	-0,1300	0,13	-0,98	0,32
Lazio	34,737	35,357	0,6195	0,08	7,29	0,00
Abruzzo	36,332	36,584	0,2516	0,13	1,92	0,05
Molise	38,327	37,518	-0,8085	0,27	-2,99	0,00
Campania	35,774	35,643	-0,1306	0,06	-2,28	0,02
Puglia	34,939	34,919	-0,0196	0,07	-0,28	0,78
Basilicata	37,339	37,354	0,0159	0,17	0,09	0,93
Calabria	36,741	36,881	0,1397	0,09	1,57	0,12
Sicilia	36,303	36,514	0,2110	0,06	3,83	0,00
Sardegna	36,620	36,849	0,2283	0,10	2,20	0,03

Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Tavola 3.8: Quota di disoccupati con titolo di studio basso.**Differenze 2018-2019 per Regione**

Titolo studio Basso	2018	2019	Diff	SE	t	P(t >t)
Piemonte	0,497	0,504	0,0069	0,00	3,24	0,00
Valle d'Aosta	0,583	0,534	-0,0482	0,01	-4,15	0,00
Lombardia	0,390	0,397	0,0068	0,00	3,64	0,00
PA Bolzano	0,533	0,520	-0,0131	0,00	-2,69	0,01
Pa Trento	0,462	0,467	0,0058	0,00	1,24	0,22
Veneto	0,443	0,454	0,0110	0,00	5,34	0,00
Friuli Venezia Giulia	0,431	0,446	0,0147	0,00	3,94	0,00
Liguria	0,455	0,446	-0,0092	0,00	-2,76	0,01
Emilia Romagna	0,496	0,486	-0,0107	0,00	-4,90	0,00
Toscana	0,468	0,465	-0,0035	0,00	-1,69	0,09
Umbria	0,421	0,439	0,0182	0,00	4,02	0,00
Marche	0,465	0,446	-0,0188	0,00	-5,58	0,00
Lazio	0,387	0,398	0,0110	0,00	5,00	0,00
Abruzzo	0,493	0,498	0,0051	0,00	1,43	0,15
Molise	0,428	0,459	0,0315	0,01	4,25	0,00
Campania	0,492	0,505	0,0127	0,00	8,14	0,00
Puglia	0,455	0,482	0,0268	0,00	14,52	0,00
Basilicata	0,476	0,504	0,0283	0,00	5,97	0,00
Calabria	0,422	0,436	0,0145	0,00	6,19	0,00
Sicilia	0,493	0,521	0,0282	0,00	18,83	0,00
Sardegna	0,516	0,521	0,0049	0,00	1,74	0,08

Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Tavola 3.9: Quota di disoccupati con titolo di studio medio.**Differenze 2018-2019 per Regione**

Titolo studio Medio	2018	2019	Diff	SE	t	P(t >t)
Piemonte	0,383	0,380	-0,0032	0,00	-1,52	0,13
Valle d'Aosta	0,324	0,359	0,0349	0,01	3,14	0,00
Lombardia	0,425	0,428	0,0026	0,00	1,36	0,17
PA Bolzano	0,410	0,420	0,0108	0,00	2,25	0,02
Pa Trento	0,404	0,397	-0,0077	0,00	-1,66	0,10
Veneto	0,396	0,401	0,0042	0,00	2,09	0,04
Friuli Venezia Giulia	0,424	0,410	-0,0132	0,00	-3,55	0,00
Liguria	0,411	0,425	0,0136	0,00	4,11	0,00
Emilia Romagna	0,366	0,367	0,0012	0,00	0,59	0,56
Toscana	0,380	0,384	0,0048	0,00	2,41	0,02
Umbria	0,433	0,415	-0,0178	0,00	-3,95	0,00
Marche	0,402	0,403	0,0010	0,00	0,30	0,77
Lazio	0,449	0,445	-0,0045	0,00	-2,03	0,04
Abruzzo	0,378	0,383	0,0050	0,00	1,46	0,15
Molise	0,428	0,413	-0,0150	0,01	-2,04	0,04
Campania	0,398	0,389	-0,0094	0,00	-6,14	0,00
Puglia	0,440	0,417	-0,0231	0,00	-12,64	0,00
Basilicata	0,410	0,387	-0,0235	0,00	-5,06	0,00
Calabria	0,443	0,422	-0,0206	0,00	-8,76	0,00
Sicilia	0,400	0,373	-0,0266	0,00	-18,25	0,00
Sardegna	0,362	0,365	0,0035	0,00	1,28	0,20

Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Tavola 3.10: Quota di disoccupati con titolo di studio alto.**Differenze 2018-2019 per Regione**

Titolo studio Alto	2018	2019	Diff	SE	t	P(t >t)
Piemonte	0,120	0,116	-0,0038	0,00	-2,74	0,01
Valle d'Aosta	0,093	0,106	0,0133	0,01	1,91	0,06
Lombardia	0,185	0,175	-0,0093	0,00	-6,39	0,00
PA Bolzano	0,057	0,060	0,0023	0,00	1,00	0,32
Pa Trento	0,134	0,136	0,0019	0,00	0,58	0,56
Veneto	0,161	0,146	-0,0152	0,00	-10,21	0,00
Friuli Venezia Giulia	0,145	0,143	-0,0015	0,00	-0,59	0,56
Liguria	0,134	0,129	-0,0044	0,00	-1,94	0,05
Emilia Romagna	0,138	0,148	0,0095	0,00	6,18	0,00
Toscana	0,152	0,151	-0,0013	0,00	-0,91	0,36
Umbria	0,146	0,146	-0,0004	0,00	-0,12	0,91
Marche	0,133	0,151	0,0178	0,00	7,54	0,00
Lazio	0,164	0,157	-0,0065	0,00	-3,91	0,00
Abruzzo	0,129	0,119	-0,0101	0,00	-4,31	0,00
Molise	0,144	0,127	-0,0165	0,01	-3,22	0,00
Campania	0,110	0,107	-0,0033	0,00	-3,43	0,00
Puglia	0,105	0,101	-0,0036	0,00	-3,24	0,00
Basilicata	0,114	0,109	-0,0048	0,00	-1,61	0,11
Calabria	0,135	0,142	0,0061	0,00	3,70	0,00
Sicilia	0,107	0,106	-0,0016	0,00	-1,69	0,09
Sardegna	0,123	0,114	-0,0083	0,00	-4,59	0,00

Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Tavola 3.11: Quota di disoccupati con esperienze lavorative nell'anno precedente la DID.**Differenze 2018-2019 per Regione**

Lavorato ultimo anno	2018	2019	Diff	SE	t	P(t >t)
Piemonte	0,694	0,700	0,0066	0,00	3,36	0,00
Valle d'Aosta	0,685	0,708	0,0225	0,01	2,09	0,04
Lombardia	0,431	0,596	0,1652	0,00	87,91	0,00
PA Bolzano	0,904	0,924	0,0199	0,00	7,26	0,00
Pa Trento	0,809	0,837	0,0280	0,00	7,80	0,00
Veneto	0,751	0,776	0,0247	0,00	14,11	0,00
Friuli Venezia Giulia	0,702	0,728	0,0257	0,00	7,56	0,00
Liguria	0,775	0,736	-0,0391	0,00	-13,60	0,00
Emilia Romagna	0,680	0,731	0,0509	0,00	25,52	0,00
Toscana	0,734	0,752	0,0185	0,00	10,24	0,00
Umbria	0,682	0,674	-0,0083	0,00	-1,94	0,05
Marche	0,612	0,610	-0,0021	0,00	-0,64	0,52
Lazio	0,603	0,621	0,0186	0,00	8,48	0,00
Abruzzo	0,689	0,694	0,0056	0,00	1,69	0,09
Molise	0,612	0,601	-0,0106	0,01	-1,45	0,15
Campania	0,639	0,598	-0,0401	0,00	-26,47	0,00
Puglia	0,598	0,576	-0,0221	0,00	-12,15	0,00
Basilicata	0,723	0,738	0,0145	0,00	3,45	0,00
Calabria	0,614	0,571	-0,0433	0,00	-18,58	0,00
Sicilia	0,550	0,532	-0,0184	0,00	-12,33	0,00
Sardegna	0,729	0,721	-0,0079	0,00	-3,16	0,00

Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

**Tavola 3.12: Quota di disoccupati in nuclei famigliari con figli.
Differenze 2018-2019 per Regione**

Presenza figli	2018	2019	Diff	SE	t	P(t >t)
Piemonte	0,293	0,295	0,0014	0,00	0,73	0,47
Valle d'Aosta	0,313	0,305	-0,0078	0,01	-0,72	0,47
Lombardia	0,247	0,284	0,0367	0,00	21,70	0,00
PA Bolzano	0,325	0,324	-0,0010	0,00	-0,22	0,82
Pa Trento	0,326	0,332	0,0054	0,00	1,22	0,22
Veneto	0,356	0,349	-0,0071	0,00	-3,59	0,00
Friuli Venezia Giulia	0,288	0,295	0,0063	0,00	1,85	0,06
Liguria	0,356	0,367	0,0115	0,00	3,57	0,00
Emilia Romagna	0,285	0,320	0,0353	0,00	17,52	0,00
Toscana	0,279	0,274	-0,0052	0,00	-2,79	0,01
Umbria	0,320	0,334	0,0140	0,00	3,28	0,00
Marche	0,338	0,360	0,0214	0,00	6,65	0,00
Lazio	0,323	0,373	0,0504	0,00	23,49	0,00
Abruzzo	0,240	0,252	0,0123	0,00	4,01	0,00
Molise	0,413	0,429	0,0163	0,01	2,21	0,03
Campania	0,250	0,245	-0,0052	0,00	-3,86	0,00
Puglia	0,347	0,347	0,0001	0,00	0,03	0,98
Basilicata	0,217	0,223	0,0054	0,00	1,38	0,17
Calabria	0,342	0,369	0,0268	0,00	11,79	0,00
Sicilia	0,269	0,291	0,0224	0,00	16,65	0,00
Sardegna	0,325	0,333	0,0084	0,00	3,18	0,00

Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

**Tavola 3.13: Quota di disoccupati donne in nuclei famigliari con figli.
Differenze 2018-2019 per Regione**

Presenza figli Donne	2018	2019	Diff	SE	t	P(t >t)
Piemonte	0,342	0,342	0,0001	0,00	0,03	0,98
Valle d'Aosta	0,383	0,376	-0,0068	0,02	-0,43	0,67
Lombardia	0,322	0,343	0,0208	0,00	8,26	0,00
PA Bolzano	0,375	0,382	0,0070	0,01	1,12	0,26
Pa Trento	0,382	0,387	0,0045	0,01	0,75	0,45
Veneto	0,417	0,411	-0,0057	0,00	-2,11	0,03
Friuli Venezia Giulia	0,348	0,360	0,0117	0,00	2,41	0,02
Liguria	0,418	0,442	0,0238	0,00	5,34	0,00
Emilia Romagna	0,331	0,373	0,0426	0,00	14,93	0,00
Toscana	0,308	0,303	-0,0049	0,00	-1,89	0,06
Umbria	0,362	0,386	0,0236	0,01	4,00	0,00
Marche	0,384	0,417	0,0331	0,00	7,30	0,00
Lazio	0,363	0,422	0,0585	0,00	18,77	0,00
Abruzzo	0,239	0,254	0,0154	0,00	3,45	0,00
Molise	0,456	0,467	0,0116	0,01	1,04	0,30
Campania	0,221	0,230	0,0091	0,00	4,76	0,00
Puglia	0,377	0,383	0,0060	0,00	2,31	0,02
Basilicata	0,204	0,216	0,0120	0,01	2,08	0,04
Calabria	0,366	0,409	0,0438	0,00	13,39	0,00
Sicilia	0,263	0,292	0,0292	0,00	14,97	0,00
Sardegna	0,361	0,374	0,0132	0,00	3,42	0,00

Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

**Tavola 3.14: Quota di disoccupati uomini in nuclei familiari con figli.
Differenze 2018-2019 per Regione**

Presenza figli Uomini	2018	2019	Diff	SE	t	P(t >t)
Piemonte	0,240	0,243	0,0036	0,00	1,36	0,18
Valle d'Aosta	0,241	0,226	-0,0150	0,01	-1,05	0,30
Lombardia	0,171	0,220	0,0491	0,00	22,56	0,00
PA Bolzano	0,258	0,243	-0,0146	0,01	-2,25	0,02
Pa Trento	0,251	0,256	0,0056	0,01	0,89	0,38
Veneto	0,276	0,272	-0,0043	0,00	-1,55	0,12
Friuli Venezia Giulia	0,216	0,213	-0,0035	0,00	-0,76	0,45
Liguria	0,280	0,278	-0,0021	0,00	-0,47	0,64
Emilia Romagna	0,232	0,258	0,0257	0,00	9,24	0,00
Toscana	0,244	0,238	-0,0056	0,00	-2,13	0,03
Umbria	0,266	0,269	0,0022	0,01	0,37	0,71
Marche	0,287	0,293	0,0060	0,00	1,34	0,18
Lazio	0,283	0,325	0,0416	0,00	14,26	0,00
Abruzzo	0,241	0,250	0,0095	0,00	2,25	0,02
Molise	0,380	0,397	0,0177	0,01	1,81	0,07
Campania	0,274	0,258	-0,0160	0,00	-8,47	0,00
Puglia	0,321	0,313	-0,0075	0,00	-3,17	0,00
Basilicata	0,227	0,228	0,0008	0,01	0,15	0,88
Calabria	0,319	0,328	0,0086	0,00	2,73	0,01
Sicilia	0,274	0,291	0,0164	0,00	8,87	0,00
Sardegna	0,290	0,294	0,0038	0,00	1,06	0,29

Fonte: Elaborazioni su dati ANPAL (SAP, DID) e MLPS – Comunicazioni Obbligatorie

STIME DEL MODELLO DI PROFILING

Tavola 3.4.1 – Stima modello logit $y=y_0$ (*)

Logistic regression		N. osservazioni = 3368727		
LR chi2(177)= 567851.35				
Prob>chi2=0				
Log likelihood= -2023606.7				
Pseudo R2 = 0.1230				
Y ₀	X	Coeff.	P>z	Std. Err.
	donna	0.35681	***	0.00537
	età	-0.07314	***	0.00154
	età al quadrato	0.00125	***	0.00002
Interazione donna classe di età	25 a 29 anni	0.06064	***	0.00903
	30 a 39 anni	0.30868	***	0.01211
	40 a 49 anni	0.48346	***	0.01584
	over 50 anni	0.45142	***	0.01835
	donna - 25 a 29 anni	0.08254	***	0.00812
	donna - 30 a 39 anni	0.12333	***	0.00769
	donna - 40 a 49 anni	-0.09954	***	0.00810
	donna - over 50 anni	-0.12999	***	0.00798
Durata presenza in Italia (rif. Cittadino italiano)	Nato In Italia	-0.11087	***	0.00907
	Fino a 12 mesi	-0.02722	***	0.00700
	Da 1 a 2 anni	-0.07374	***	0.01121
	Oltre 2 anni	0.07119	***	0.00470
	Occupato alla data DID	0.20041	***	0.00533
Interazione classe età e precedente esperienza di lavoro 12 mesi pre DID	Prec. Esperienza - inferiore 24 anni	-0.32165	***	0.00724
	Prec. Esperienza - 25 a 29 anni	0.08490	***	0.00837
	Prec. Esperienza - 30 a 39 anni	-0.12505	***	0.00808
	Prec. Esperienza - 40 a 49 anni	-0.49783	***	0.00850
	Prec. Esperienza - over 50	-0.69592	***	0.00882
Titolo di studio (cat. Rif. Fino licenza elementare)	Licenza media	-0.07299	***	0.00471
	Qualifica professionale	-0.25288	***	0.00710
	Istituto professionale	-0.28140	***	0.00663
	Istituto tecnico	-0.18764	***	0.00563
	Liceo	-0.02406	***	0.00705
	Diploma: altro	-0.17521	***	0.00639
Diploma o triennale	Scienze umanistiche	-0.35588	***	0.01217
	Scienze sociali	-0.20022	***	0.01106
	Scienze della salute	-0.51349	***	0.01571
	Ingegneria, informatica e trasporti	-0.54653	***	0.01927
	Altro diploma o triennale	-0.23862	***	0.01337
	Scienze umanistiche	-0.55807	***	0.01115
Laurea magistrale, specialistica, vecchio ordinamento	Scienze sociali	-0.25735	***	0.00893
	Scienze della salute	-0.66594	***	0.01648
	Ingegneria, informatica e trasporti	-0.74078	***	0.01574
	Scienze naturali	-0.57046	***	0.02125
	Architettura	-0.04156	**	0.02023
	Altra laurea magistrale, specialistica, vecchio ordinamento	-0.34771	***	0.01081
Possedere una patente		-0.18909	***	0.00293
Condizione prevalente anno precedente dichiarata (cat. Rif. Altro Inattivo)	Occupato	-0.00667		0.00474
	In cerca di nuova occupazione	-0.33050	***	0.00486
	In cerca di prima occupazione	0.06591	***	0.00588
	Studente	-0.14136	***	0.00587
	Ha svolto tirocinio nei 12 mesi precedenti DID	-0.13303	***	0.00800

Logistic regression
 LR chi2(177)= 567851.35
 Prob>chi2=0
 Log likelihood= -2023606.7
 Pseudo R2 = 0.1230

N. osservazioni = 3368727

Y ₀	X	Coeff.	P>z	Std. Err.
Professione prevalente 24 mesi precedenti DID (cat. Rif. Non ha lavorato nel periodo)	Bassa qualifica	-1.15914	***	0.01089
	Media qualifica	-1.12805	***	0.01073
	Alta qualifica	-1.26750	***	0.01150
Interazione donna e Settore prevalente 24 mesi precedenti DID (cat. Rif. Non ha lavorato nel periodo)	Agricoltura	-0.82309	***	0.01254
	Industria in senso stretto	0.02307	**	0.00917
	Costruzioni	-0.21646	***	0.00934
	Commercio	0.12146	***	0.00964
	Alloggio e Ristorazione	-0.45065	***	0.00934
	Trasporto Immagazzinaggio e Altri servizi di mercato	-0.14047	***	0.00874
	P.A., Istruzione e Sanità	-0.35793	***	0.01267
	donna - Agricoltura	-0.19668	***	0.01685
	donna - Industria in senso stretto	-0.08779	***	0.00968
	donna - Costruzioni	0.61878	***	0.02061
	donna - Commercio	-0.13387	***	0.00910
	donna - Alloggio e Ristorazione	-0.05079	***	0.00843
	donna - Trasporto Immagazzinaggio e Altri servizi di mercato	0.01362	*	0.00790
donna - P.A., Istruzione e Sanità	-0.54571	***	0.01265	
donna - Altri servizi pubblici, sociali e personali	0.06949	***	0.01015	
Numero datori cambiati 24 mesi precedenti DID (cat. Rif. Più di due datori)	1 datore di lavoro	0.86323	***	0.00433
	2 datori di lavoro	0.49742	***	0.00467
	Più di 2 datori di lavoro	0.00000		0.00000
Impegno familiare	Famiglia con figli	-0.05840	***	0.00459
	Donna con figli	0.37329	***	0.00578
Numero componenti famiglia (cat. Rif. Unico componente)	2 componenti	0.01955	***	0.00436
	3 componenti	-0.06298	***	0.00434
	4 componenti	-0.13755	***	0.00434
	5 componenti	-0.10155	***	0.00534
	Più di 5 componenti	0.00871		0.00747
Costante		1.65022	***	0.02659

(*) non si riportano i coefficienti provinciali

Figura 3.5.1 - Curva ROC, grafico sensitivity-specificity. Modello y_0

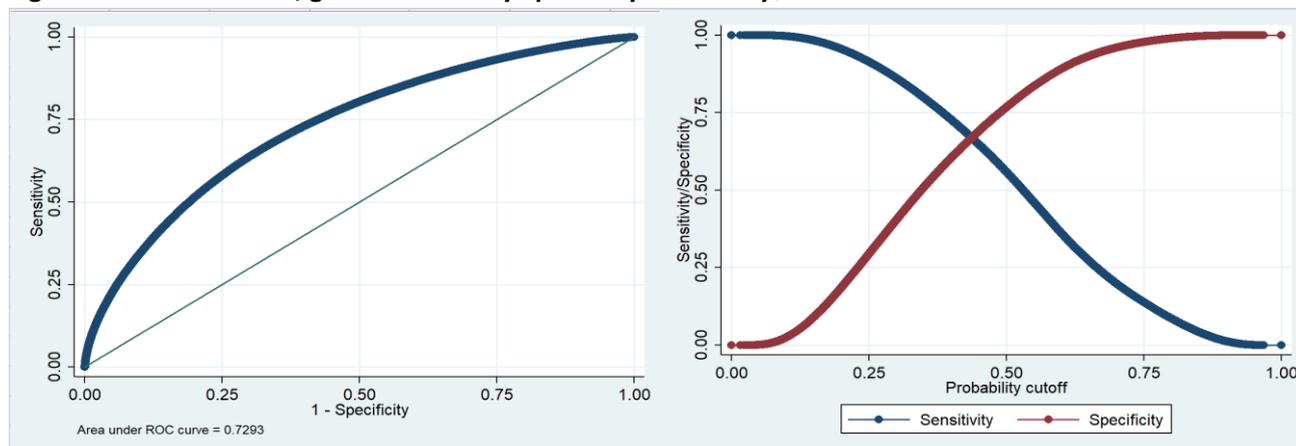


Tavola 3.4.2 – Stima modello logit $y=y_1$ (*)

Logistic regression		N. osservazioni = 3368727		
LR chi2(177)= 504774.02				
Prob>chi2=0				
Log likelihood= -2077162.7				
Pseudo R2 = 0.1083				
Y ₁	X	Coeff.	P>z	Std. Err.
	donna	0.33548	***	0.00558
	età	-0.11543	***	0.00154
	età al quadrato	0.00162	***	0.00002
Interazione donna classe di età	25 a 29 anni	0.10157	***	0.00895
	30 a 39 anni	0.51155	***	0.01203
	40 a 49 anni	0.79982	***	0.01572
	over 50 anni	0.80380	***	0.01827
	donna - 25 a 29 anni	0.05075	***	0.00789
	donna - 30 a 39 anni	0.08164	***	0.00752
	donna - 40 a 49 anni	-0.13385	***	0.00795
	donna - over 50 anni	-0.17954	***	0.00794
Durata presenza in Italia (cat. rif. Cittadino italiano)	Nato In Italia	-0.04788	***	0.00902
	Fino a 12 mesi	0.05647	***	0.00693
	Da 1 a 2 anni	-0.06590	***	0.01100
	Oltre 2 anni	0.05690	***	0.00462
	Occupato alla data DID	0.13278	***	0.00511
Interazione classe età e precedente esperienza di lavoro 12 mesi pre DID	Prec. Esperienza - inferiore 24 anni	-0.22344	***	0.00702
	Prec. Esperienza - 25 a 29 anni	0.03269	***	0.00815
	Prec. Esperienza - 30 a 39 anni	-0.22839	***	0.00804
	Prec. Esperienza - 40 a 49 anni	-0.59969	***	0.00858
	Prec. Esperienza - over 50	-0.81599	***	0.00908
Titolo di studio (cat. Rif. Fino licenza elementare)	Licenza media	-0.06629	***	0.00473
	Qualifica professionale	-0.28030	***	0.00694
	Istituto professionale	-0.31526	***	0.00652
	Istituto tecnico	-0.24783	***	0.00561
	Liceo	-0.05439	***	0.00705
	Diploma: altro	-0.20784	***	0.00637
Diploma o triennale	Scienze umanistiche	-0.38502	***	0.01173
	Scienze sociali	-0.30349	***	0.01087
	Scienze della salute	-0.59178	***	0.01512
	Ingegneria, informatica e trasporti	-0.74215	***	0.01872
	Altro diploma o triennale	-0.30420	***	0.01306
Laurea magistrale, specialistica, vecchio ordinamento	Scienze umanistiche	-0.60221	***	0.01066
	Scienze sociali	-0.37196	***	0.00881
	Scienze della salute	-0.78932	***	0.01595
	Ingegneria, informatica e trasporti	-0.94553	***	0.01531
	Scienze naturali	-0.68047	***	0.02040
	Architettura	-0.18987	***	0.02008
	Altra laurea magistrale, specialistica, vecchio ordinamento	-0.46654	***	0.01060
Possedere una patente		-0.17072	***	0.00289
Condizione prevalente anno precedente dichiarata (cat. Rif. Altro Inattivo)	Occupato	-0.08474	***	0.00473
	In cerca di nuova occupazione	-0.34093	***	0.00487
	In cerca di prima occupazione	0.04654	***	0.00611
	Studente	-0.14670	***	0.00595
Ha svolto tirocinio nei 12 mesi precedenti DID		-0.19066	***	0.00764
Professione prevalente 24 mesi precedenti DID (cat. Non ha lavorato)	Bassa qualifica	-0.80958	***	0.01050
	Media qualifica	-0.79852	***	0.01034
	Alta qualifica	-0.96198	***	0.01105

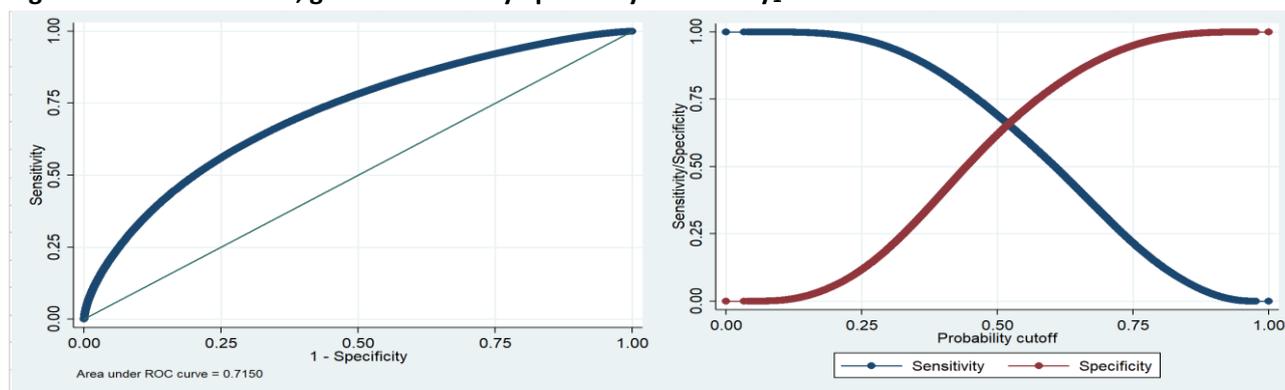
Logistic regression
 LR chi2(177)= 504774.02
 Prob>chi2=0
 Log likelihood= -2077162.7
 Pseudo R2 = 0.1083

N. osservazioni = 3368727

Y ₁	X	Coeff.	P>z	Std. Err.
nel periodo)				
Interazione donna e Settore prevalente 24 mesi precedenti DID (cat. Non ha lavorato nel periodo)	Agricoltura	-0.76503	***	0.01132
	Industria in senso stretto	-0.08283	***	0.00879
	Costruzioni	-0.22451	***	0.00891
	Commercio	0.01988	**	0.00926
	Alloggio e Ristorazione	-0.38577	***	0.00872
	Trasporto Immagazzinaggio e Altri servizi di mercato	-0.17384	***	0.00831
	P.A., Istruzione e Sanità	-0.49687	***	0.01192
	donna - Agricoltura	-0.21871	***	0.01530
	donna - Industria in senso stretto	-0.02847	***	0.00967
	donna - Costruzioni	0.58425	***	0.02115
	donna - Commercio	-0.07447	***	0.00908
	donna - Alloggio e Ristorazione	-0.07571	***	0.00803
	donna - Trasporto Immagazzinaggio e Altri servizi di mercato	0.03109	***	0.00783
	donna - P.A., Istruzione e Sanità	-0.51840	***	0.01202
donna - Altri servizi pubblici, sociali e personali	0.00643		0.01003	
N. datori cambiati 24 mesi precedenti DID (cat. Rif. Più di due datori)	1 datore di lavoro	0.63592	***	0.00390
	2 datori di lavoro	0.33012	***	0.00421
	Più di 2 datori di lavoro	0.00000		0.00000
Impegno familiare	Famiglia con figli	-0.09423	***	0.00443
	Donna con figli	0.38334	***	0.00568
N. componenti famiglia (cat. Rif. Unico componente)	2 componenti	-0.00100		0.00429
	3 componenti	-0.08281	***	0.00425
	4 componenti	-0.16043	***	0.00424
	5 componenti	-0.11227	***	0.00524
	Più di 5 componenti	0.00811		0.00743
Costante		2.87448	***	0.02662

(*) non si riportano i coefficienti provinciali

Figura 3.5.2 - Curva ROC, grafico sensitivity-specificity. Modello y₁



BIBLIOGRAFIA

- AlgorithmWatch (2019). Automating society: Taking stock of a automated decision making in the EU. Berlin, AlgorithmWatch in cooperation with Bertelsmann Stiftung.
- Altman, DG, Bland, JM. Diagnostic tests 2: Predictive values. *BMJ*. 1994 Jul 9;309(6947):102. doi: 10.1136/bmj.309.6947.102. PMID: 8038641; PMCID: PMC2540558.
- Altman, DG, Bland, JM. Diagnostic tests. 1: Sensitivity and specificity. *BMJ*. 1994 Jun 11;308(6943):1552. doi: 10.1136/bmj.308.6943.1552. PMID: 8019315; PMCID: PMC2540489
- Barnes, S-A., Wright, S., Irving, P., Deganis, I. (2015), "Identification of latest trends and current developments in methods to profile jobseekers in European public employment services". Final report, Brussels: Directorate-General for Employment, Social Affairs and Inclusion. European Commission.
- Barnes, S.-A., Wright, S., Irving, P. and Deganis, I. (2015), Identification of latest trends and current developments in methods to profile jobseekers in European public employment services: final report.
- Black, D.A., Smith, J.A., Plesca, M. and Shannon, S. (2003), Profiling UI claimants to allocate reemployment services: Evidence and Recommendations for States. Final Report to United States Department of Labor.
- Breiman, L. Random Forests. *Machine Learning* 45, 5–32 (2001)
- Centra, M., De Minicis, M., Marocco, M., Gualtieri, V., 2016, Profiling e garanzia giovani, in Canal T. (a cura di), *L'Italia fra Jobs Act ed Europa 2020, Isfol, I libri del Fondo Sociale Europeo*, Spoleto (PG), Del Gallo Editori
- Desiere, S., Langenbucher, K. and Struyven, L. (2019), Statistical profiling in public employment services. *OECD Working Paper*.
- Eberts, R.W., O'Leary, C.J. and Wandner, S.A. (2002), Targeting employment services, WE Upjohn Institute.
- Eubanks, V. (2018), Automating inequality: How high-tech tools profile, police, and punish the poor, St. Martin's Press.
- Henman, P. (2004). 'Targeted! Population segmentation, electronic surveillance and governing the unemployed in Australia.' *International Sociology*, 19(2), 173–191.
- Kleinberg, J., Mullainathan, S. and Raghavan, M. (2016), 'Inherent trade-offs in the fair determination of risk scores.' *arXiv preprint arXiv:1609.05807*.
- Loxha, A. and Morgandi, M. (2014), Profiling the unemployed: a review of OECD experiences and implications for emerging economies. *Social Protection and labor discussion paper*. World Bank Group, Washington, DC.
- Martin, K. (2018), 'Ethical implications and accountability of algorithms.' *Journal of Business Ethics*, 1–16.
- McCullagh P. & Nelder (1989), *Generalized linear models* 2nd edition, Chapman and Hall, London.
- OECD (2018), "Profiling tools for early identification of jobseekers who need extra support." *Policy Brief on Activation Policies*, OECD Publishing, Paris
- OECD (1998), *Early identification of jobseekers at risk of long-term unemployment: the role of profiling*, OECD.
- Pope, D.G. and Sydnor, J.R. (2011), 'Implementing anti-discrimination policies in statistical profiling models.' *American Economic Journal: Economic Policy*, 3, 3, 206–231.
- Schwab, S. (1986), 'Is statistical discrimination efficient?' *The American Economic Review*, 76, 1, 228–234.
- Weber, T., (2011), "Profiling systems for effective labour market integration". *Thematic Synthesis Paper*. European Commission. Directorate-General for Employment, Social Affairs and Inclusion
- Wijnhoven, M. and Havinga, H. (2014), 'The Work Profiler: A digital instrument for selection and diagnosis of the unemployed.' *Local Economy*, 29, 6-7, 740–749.

COLLANA
FOCUS
ANPAL