
Opportunità lavorative e contrattuali: un'analisi sulle differenze territoriali e tra gli ambiti di studio

Riccardo Merlo¹, Tommaso Redaelli², Leonardo Riva³

¹ Università degli Studi Milano-Bicocca, CdL Data Science, Matricola 829805, r.merlo2@campus.unimib.it

² Università degli Studi Milano-Bicocca, CdL Data Science, Matricola 830442, t.redaelli7@campus.unimib.it

³ Università degli Studi Milano-Bicocca, CdL Data Science, Matricola 830647, l.riva37@campus.unimib.it

Abstract:

Il raggiungimento di un certo titolo di studio dovrebbe garantire pari soddisfazioni e opportunità lavorative all'interno dell'intera nazione, qualunque sia la tipologia del titolo conseguito e la provenienza territoriale dello studente. Ma è davvero così? Questa analisi si basa sul dataset, fornito dall'ISTAT, dei laureati intervistati nel 2015.

Sono state individuate delle differenze significative tra le zone d'Italia, in particolare riguardo allo spostamento dal sud verso il nord, il tutto incentivato dalla minor tempistica nel trovare lavoro. Inoltre, non sono emerse delle disparità contrattuali e/o di percorso di studio che giustifichino in questo senso la partenza dei neolaureati in cerca di una migliore carriera. Infine, è stata verificata la relazione tra soddisfazione e tipologia contrattuale, rapportando poi questo legame sulle diverse tipologie di corsi di laurea.

1. Introduzione	2
1.1 Descrizione dataset	2
1.2. Attributi utilizzati	3
2. Inferenze	6
2.1 Disponibilità a trasferirsi per lavoro in un'altra regione (tra nord, centro, sud e isole)	6
2.2 Effettivi trasferimenti in un'altra regione (tra nord, centro, sud e isole)	7
2.3 Mesi necessari per trovare lavoro	9
2.4 Soddisfazione della stabilità e sicurezza del lavoro e tipologia di contratto lavorativo.	10
2.5 Soddisfazione stabilità e sicurezza del lavoro delle diverse tipologie di contratto lavorativo nei diversi gruppi di laurea.	11
3. Conclusioni	13
Appendice	13

1. Introduzione

1.1 Descrizione dataset

“Inserimento professionale dei laureati”¹ è il nome del dataset utilizzato per le indagini, pubblicato nel 2015 da ISTAT. Fanno parte di esso anche altre due rilevazioni: l’Indagine campionaria sui “Percorsi di studio e di lavoro dei diplomati di scuola secondaria di secondo grado” e l’indagine totale sull’“Inserimento professionale dei dottori di ricerca”. Ha tra i principali obiettivi quello di fornire informazioni per valutare la condizione occupazionale dei giovani a una prefissata distanza dal conseguimento del titolo di studio (rispettivamente diploma, laurea, dottorato di ricerca). La scelta di analizzare il fenomeno a una certa distanza dal titolo è motivata sia dall’esigenza di tener conto dell’eventuale conclusione di ulteriori attività di qualificazione, sia dal fatto che i tempi di inserimento nel mercato del lavoro non sono brevi, soprattutto se si vuole indagare anche sulla qualità del lavoro svolto. I dati raccolti comprendono informazioni relative alla popolazione dell’intero territorio italiano, e pertanto permettono di studiare l’impatto dell’origine territoriale e sociale, sia sul processo di scelta universitaria, sia sul processo di entrata nel mondo lavorativo.

Popolazione di riferimento

È rappresentata da quanti nell’anno 2011 hanno conseguito la laurea in un ateneo italiano. Si tratta di 299.449 individui, 58.9% donne, appartenenti a 90 atenei.

L’indagine ha riguardato:

- I 169 232 laureati nei corsi di laurea (di primo livello)
- I 43 624 laureati nei corsi “a ciclo unico” (che comprendono, oltre alle lauree specialistiche/magistrali a ciclo unico, anche le lauree “tradizionali”, della durata 4-6 anni, del vecchio ordinamento)
- Gli 86 593 laureati nei corsi di laurea specialistica/magistrale (di durata biennale).

Campionamento

L’indagine ha come universo due popolazioni distinte: la prima è costituita dai laureati nei corsi di laurea a ciclo unico e specialistica biennale, la seconda è composta da coloro che hanno conseguito una laurea triennale del nuovo ordinamento.

I domini di interesse sono l’intero territorio nazionale, i gruppi di corsi di laurea e i corsi di laurea (o le classi) incrociati con il sesso, gli atenei e gli atenei incrociati con l’area didattica.

Il campione è stato selezionato mediante due distinti disegni di campionamento, il primo progettato per la popolazione dei laureati nei corsi di laurea a ciclo unico e specialistica biennale, il secondo per i laureati nei corsi triennali.

Per ciascuna delle due popolazioni di riferimento è stato definito un disegno a stratificazione incompleta, un caso speciale di un campione bilanciato dove i vincoli di bilanciamento sono il risultato dell’allocazione nei domini definiti dagli incroci corso di laurea per sesso e ateneo per area tematica, dalla cui aggregazione si ottengono tutti i domini di interesse dell’indagine sopra definiti. Il processo di allocazione a Stratificazione incompleta, a differenza del campionamento casuale semplice stratificato (SSRS) utilizzato nelle scorse edizioni d’indagine, non ha vincoli di numerosità negli strati. Il vantaggio a cui porta tale procedura è la riduzione della dimensione campionaria complessiva, a parità di precisione attesa delle stime nei domini d’interesse.

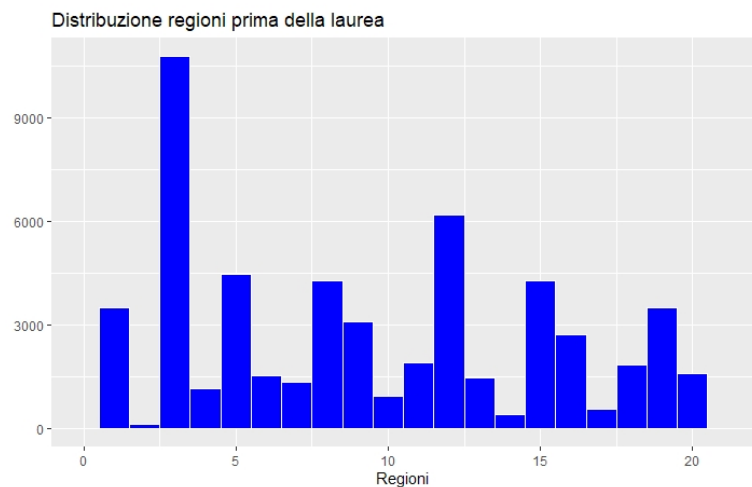
¹ “L’indagine sui percorsi di studio e di lavoro dei diplomati: microdati ad uso pubblico”, ISTAT, 2017, <https://www.istat.it/it/archivio/96042>

1.2. Attributi utilizzati

REG_O_EST_RES_mIcro

Rappresenta le regioni di residenza prima del conseguimento della laurea. Le regioni sono contrassegnate da un codice, in ordine crescente da nord a sud. È un attributo categorico nominale.

Modalità	Descrizione
1	Piemonte
2	Valle d'Aosta
3	Lombardia
4	Trentino-Alto Adige
5	Veneto
6	Friuli-Venezia Giulia
7	Liguria
8	Emilia-Romagna
9	Toscana
10	Umbria
11	Marche
12	Lazio
13	Abruzzo
14	Molise
15	Campania
16	Puglia
17	Basilicata
18	Calabria
19	Sicilia
20	Sardegna
99	Estero



In seguito, per le inferenze, verranno spesso effettuati dei raggruppamenti sulle regioni, in nord (regioni da 1 a 8), centro (da 9 a 12), sud (da 13 a 18) e isole (19 e 20). Il raggruppamento in zone² segue lo schema di seguito rappresentato.



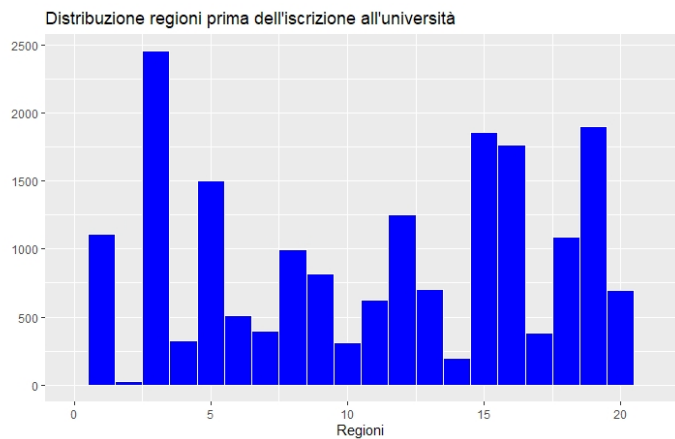
² “Gruppi di regioni d’Italia”, Wikipedia, 2020 https://it.wikipedia.org/wiki/Gruppi_di_regioni_d'Italia]

REG_O_EST_PRIMA_mIcro

Rappresenta la regione di residenza prima dell'iscrizione all'università le quali sono contrassegnate da un codice, in ordine crescente da nord a sud. È un attributo categorico ordinale.

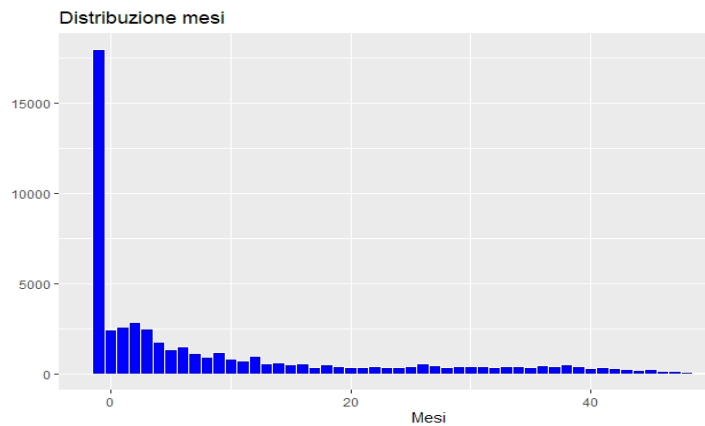
I codici delle regioni seguono lo stesso schema dell'attributo REG_O_EST_RES_mIcro.

Anche la suddivisione in zone è eseguita con lo stesso procedimento.



mesi_I_lavoro

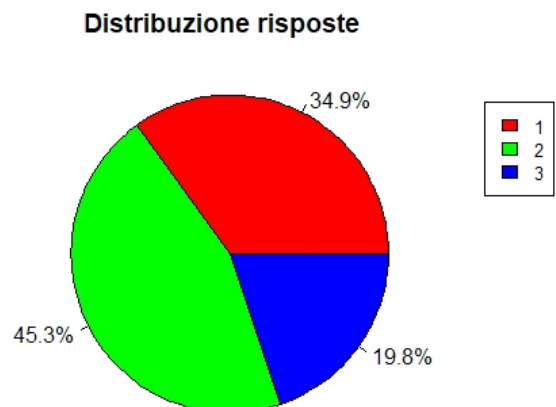
Rappresenta i mesi passati prima di trovare il primo lavoro.



L3_10

Rappresenta la disponibilità a cambiare città per lavorare. Le opzioni di risposta sono contrassegnate da dei codici. È un attributo categorico nominale, con valori in {1, 2, 3, <NA>}.

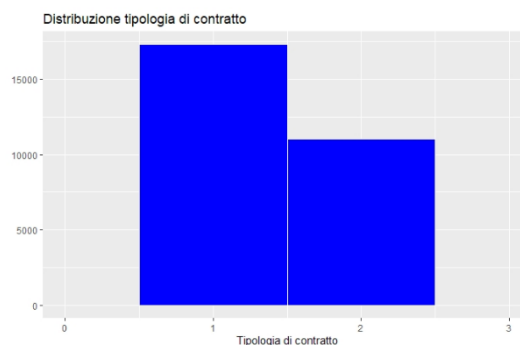
Modalità	Descrizione
1	“Sì, ma solo nell'ambito della regione in cui vivo”
2	“Sì, anche fuori della regione in cui vivo”
3	“No”



L2_17

Indica la tipologia di contratto. È un attributo categorico nominale, con valori in {1, 2, <NA>}.

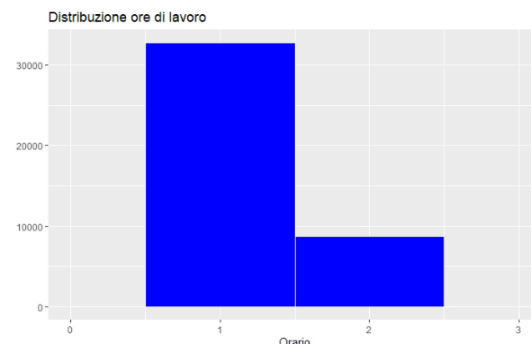
Modalità	Descrizione
1	Lavoro a tempo indeterminato
2	Lavoro a tempo determinato



L2_22

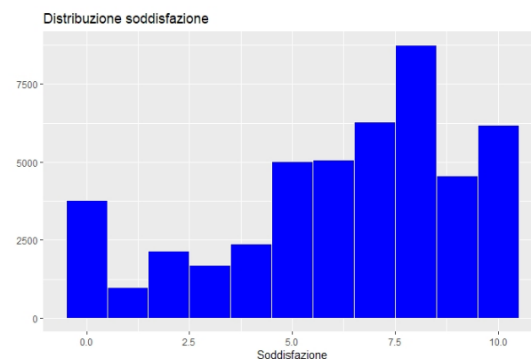
Indica le ore di lavoro. È un attributo categorico nominale, con valori in {1, 2, <NA>}.

Modalità	Descrizione
1	Lavoro a tempo pieno
2	Lavoro part-time



L2_54B

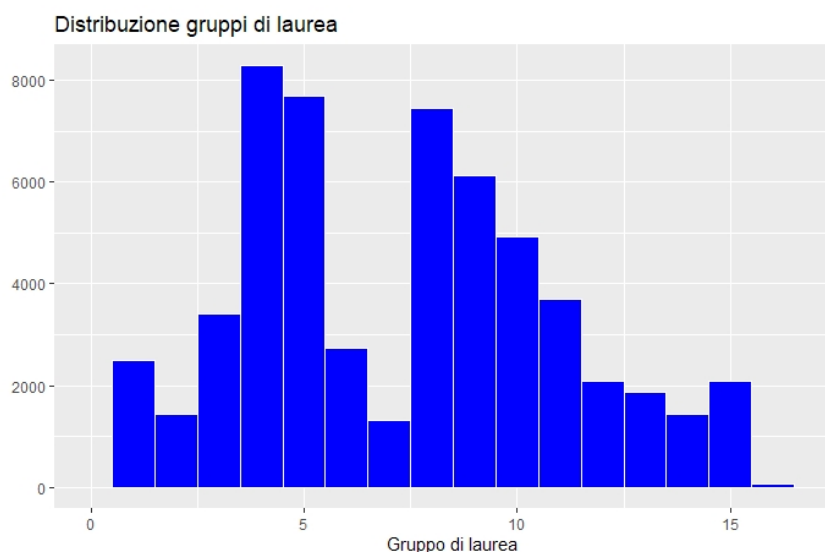
È un indice di soddisfazione e prospettive di stabilità e di sicurezza sul lavoro. È un attributo categorico ordinale, con valori nell'intervallo [0,10].



GRUPPO_mIcro

Indica il gruppo di laurea. È un attributo categorico nominale.

1. Scientifico
2. Chimico-farmaceutico
3. Geo-biologico
4. Medico
5. Ingegneria
6. Architettura
7. Agrario
8. Economico-statistico
9. Politico-sociale
10. Giuridico
11. Letterario
12. Linguistico
13. Insegnamento
14. Psicologico
15. Educazione fisica
16. Difesa e sicurezza



2. Inferenze

2.1 Disponibilità a trasferirsi per lavoro in un'altra regione (tra nord, centro, sud e isole)

Indagine

Si vuole analizzare la disponibilità delle persone a trasferirsi per lavoro in un'altra regione. Verranno confrontati i dati rispetto a dei gruppi di regioni: nord, centro, sud e isole.

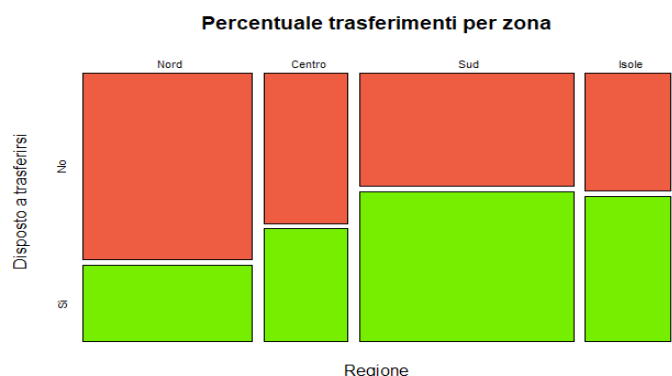
Cleaning e preprocessing

Il procedimento di raccolta e pulizia dei dati consiste in:

1. Selezione, dal dataset iniziale, dei soli attributi "REG_O_EST_RES_mIcro" e "L3_10".
2. Rimozione dei record con regione "99" (estero) ed "NA" (valori mancanti).
3. Rimozione dei record con disponibilità al trasferimento "NA".
4. Modifica dei record con disponibilità al trasferimento "3" in "1", ottenendo quindi:
 - "1" per chi non sarebbe disposto a trasferirsi al di fuori della regione per lavoro.
 - "2" per chi sarebbe disposto a farlo.
5. Raggruppamento delle regioni in nord, centro, sud, isole.

Dati pronti per il test

Zona	No	Sì	Totali	Percentuale
Nord	570	234	804	29.10 %
Centro	227	171	398	42.96 %
Sud	436	579	1015	57.04 %
Isole	182	225	407	55.28 %



Inferenza

Dopo aver osservato i dati, si nota come al sud e isole le persone siano più disponibili a trasferirsi al di fuori della regione per lavoro. Vengono confrontati quindi sud-isole e successivamente isole-centro. Per farlo, viene utilizzato un proportion-test.

Ipotesi nulla H_0 : "coloro che vivono al sud sono allo stesso modo disponibili a trasferirsi di quanto lo sono quelli che vivono nelle isole".

Ipotesi alternativa H_1 : "c'è una differenza statisticamente significativa nella disponibilità a trasferirsi".

Viene effettuato un proportion-test a due code con confidenza al 95%, $\alpha=0.05$.

Risultati:

X-squared con 1 gdl	0.299
P-value	0.5847

Non viene rifiutata l'ipotesi nulla, confermando ciò che si desumeva dal grafico: al sud e nelle isole la disponibilità a trasferirsi è la stessa.

Ipotesi nulla H_0 : coloro che vivono nelle isole sono allo stesso modo disponibili a trasferirsi di quanto lo siano quelli che vivono al centro.

Ipotesi alternativa H_1 : chi vive nelle isole è statisticamente più disponibile a trasferirsi.
Viene effettuato un proportion-test a una coda con confidenza al 95%, $\alpha=0.05$.

Risultati:

X-squared con 1 gdl	11.728
P-value	0.00003

Viene rifiutata l'ipotesi nulla: in questo caso, la disponibilità nelle isole è statisticamente maggiore rispetto al centro Italia.

Possiamo di conseguenza desumere che la disponibilità nelle isole è statisticamente maggiore anche rispetto al nord Italia.

2.2 Effettivi trasferimenti in un'altra regione (tra nord, centro, sud e isole)

Indagine

Dopo aver analizzato la disponibilità al trasferimento, si vuole verificare l'effettivo riscontro con i reali trasferimenti dei laureati, avvenuti nel periodo tra l'iscrizione all'Università e il successivo periodo in cui sono stati acquisiti i dati attuali, in un'altra regione. Verranno confrontati i dati rispetto a dei gruppi di regioni: nord, centro, sud e isole.

Dati non stratificati

		Regione Attuale			
		1. Nord	2. Centro	3. Sud	4. Isole
Regione Nascita	1. Nord	5607	348	59	35
	2. Centro	818	1554	135	30
	3. Sud	2334	1673	1422	70
	4. Isole	1000	377	50	919

Cleaning e preprocessing

Procedimento di raccolta e pulizia dei dati:

1. Selezione, dal dataset iniziale, dei soli attributi "REG_O_EST_PRIMA_mIcro" (Regione Nascita) e "REG_O_EST_RES_mIcro" (Regione Attuale).
2. Rimozione dei record con regione "NA" o "99" da entrambi gli attributi.
3. Raggruppamento delle regioni in quattro zone: nord, centro, sud e isole.
4. Campionamento stratificato per l'attributo "Regione Attuale" sui quattro gruppi di regioni.

N.B.: Si suppone che la zona d'Italia di residenza prima dell'iscrizione all'Università coincida con quella di nascita.

Dati pronti per il test

		Regione Attuale			
		1. Nord	2. Centro	3. Sud	4. Isole
Regione Nascita	1. Nord	2175	135	23	14
	2. Centro	756	1437	125	28
	3. Sud	996	714	607	30
	4. Isole	990	373	50	910

Sulla diagonale sono presenti le persone che non hanno cambiato regione dopo essersi iscritti all'Università.

Inferenza

Dopo aver osservato i dati, si nota come la maggior parte delle persone che si sono trasferite si è spostata nel nord Italia mentre pochissime verso sud o nelle isole.

Inoltre, al sud e alle isole sono più le persone che si sono trasferite rispetto a quelle che sono rimaste. Al contrario chi appartiene ad una regione del nord difficilmente si è spostato. È stato, quindi, effettuato un test sulle proporzioni a due code (confidenza al 95%, $\alpha = 0.05$) al fine di dare significatività statistica a queste osservazioni.

Un primo test è stato calcolato tra i trasferiti che erano residenti al sud contro quelli delle isole per verificare quale tra le due zone possiede più nativi che si sono trasferiti in altre parti d'Italia.

Ipotesi nulla H_0 : "i nativi del sud e delle isole si trasferiscono con la stessa frequenza in un'altra zona d'Italia".

Ipotesi alternativa H_1 : "i nativi del sud si trasferiscono maggiormente in un'altra zona d'Italia rispetto ai nativi delle isole".

Viene fortemente rifiutata l'ipotesi nulla con p-value $< 2.2e-16$ (X-squared = 138.59, gdl = 1).

È così dimostrato che i laureati nativi del sud si trasferiscono maggiormente rispetto a quelli delle isole, in particolare il 74% del totale contro il 61%.

Un secondo test è stato calcolato tra i trasferiti che erano residenti nelle isole contro quelli del centro.

Ipotesi nulla H_0 : "i nativi delle isole e del centro si trasferiscono con la stessa frequenza in un'altra zona d'Italia".

Ipotesi alternativa H_1 : "i nativi delle isole si trasferiscono maggiormente in un'altra zona d'Italia rispetto ai nativi del centro".

Viene fortemente rifiutata l'ipotesi nulla con p-value $< 2.2e-16$ (X-squared = 236.86, gdl = 1).

Alle isole si trasferiscono il 61% del totale contro il 39% dei laureati al centro.

Un terzo test è stato calcolato tra i trasferiti che erano residenti al nord contro quelli del centro.

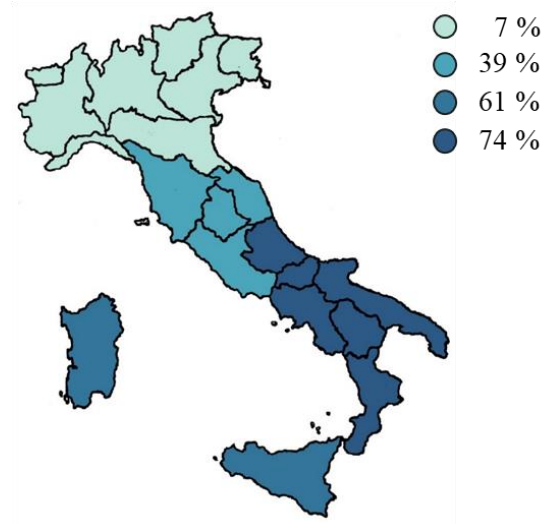
Ipotesi nulla H_0 : "i nativi del nord e del centro si trasferiscono con la stessa frequenza in un'altra zona d'Italia".

Ipotesi alternativa H_1 : "i nativi del centro si trasferiscono maggiormente in un'altra zona d'Italia rispetto ai nativi del nord".

Viene fortemente rifiutata l'ipotesi nulla con p-value $< 2.2e-16$ (X-squared = 1274, gdl = 1).

Al centro si trasferiscono il 39% del totale contro il 7% dei laureati al nord.

Possiamo dunque concretare che le zone d'Italia, i cui nativi si sono trasferiti maggiormente sono: Sud > Isole > Centro > Nord.



2.3 Mesi necessari per trovare lavoro

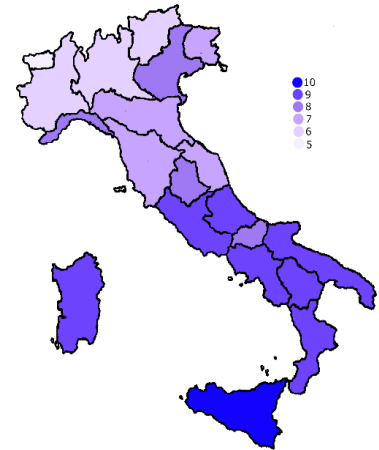
Indagine

Dopo aver notato differenze significative nel trasferimento tra zone d'Italia, si vogliono trovare dei fattori che possano causarlo. Perciò, verranno analizzati i mesi necessari a trovare il primo lavoro in ogni regione d'Italia, al fine di verificare una possibile correlazione tra i due dati.

Cleaning e preprocessing

Procedimento di raccolta e pulizia dei dati:

1. Selezione, dal dataset iniziale, dei soli attributi "REG_O_EST_PRIMA_mIcro" e "mesi_I_lavoro".
2. Rimozione dei record con regione "99" (estero) o "NA" (valori mancanti).
3. Rimozione dei record con mese "-1" (già impiegati).
4. Rimozione dei record outlier che hanno un mese con un valore minore di $\text{Inter Quantile Range} * 1.5 + 75\text{th Quantile}$. Formula derivata dal calcolo dei whiskers nella funzione di boxplot in R.
5. Raggruppamento dei record per regione e assegnazione del valore mediano dell'attributo mese.



Inferenza

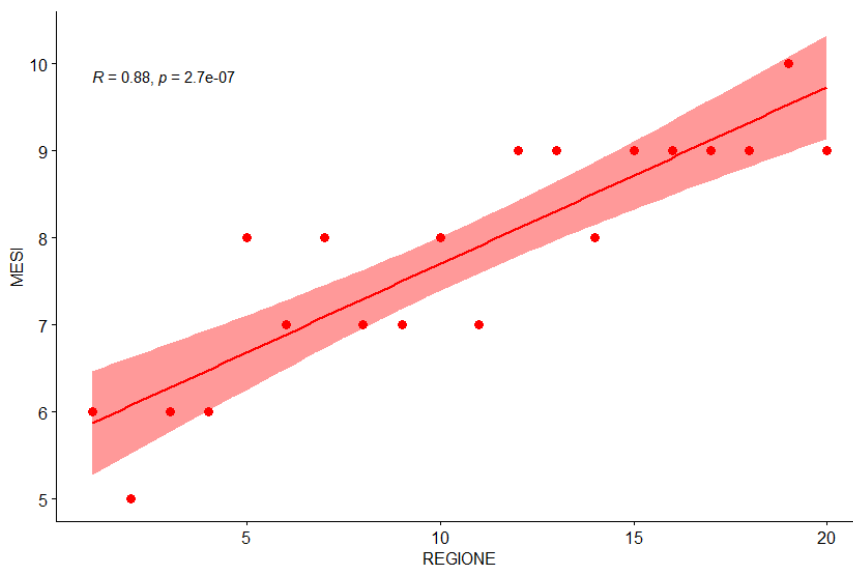
Dopo aver osservato i dati [tabella Italia colorata], si nota come attraversando l'Italia da nord a sud i mesi necessari a trovare il primo lavoro aumentano. Al fine di dimostrare questa relazione è stata effettuata una regressione lineare semplice tra la variabile delle regioni e la variabile dei mesi.

Ipotesi nulla H_0 : il tempo necessario a trovare il primo lavoro rimane costante per ogni regione.

Ipotesi alternativa H_1 : il tempo necessario a trovare il primo lavoro aumenta spostandosi da nord verso sud.

Il test F per la correlazione, calcolato con 1 e 18 gdl, ottenendo un valore $63.29 > 15.38$, rifiuta fortemente l'ipotesi nulla con un p-value di $2.652e-07$ (con confidenza al 95%, $\alpha = 0.05$) dimostrando anche quanto significativo questo legame sia.

$$\text{Mesi}_i = 5.668 + 0.203 \text{ Regione}_i$$



Parametri di regressione:

RSE (18 gdl)	0.658
R ²	0.7786
R ² corretto	0.7663
F (1 e 18 gdl)	63.29
P-value	2.562e-07

Residui:

Min	- 1.07444
1 st Qt.	-0.48421
Mediana	-0.01805
3 rd Qt.	0.34474
Max	1.31654

Da questa statistica si evince che i mesi passati per trovare il primo lavoro aumentano significativamente spostandosi da nord verso sud, come mostrato dall'infografica.

2.4 Soddisfazione della stabilità e sicurezza del lavoro e tipologia di contratto lavorativo.

Un altro importante aspetto di pari opportunità sul quale ci siamo voluti soffermare è la possibilità di ottenere una condizione lavorativa stabile e soddisfacente indipendentemente dal titolo conseguito tramite laurea universitaria.

A monte di questa indagine è stato verificato lo stato di dipendenza tra regione e tipo di contratto a seguito di un test del Chi Quadro per l'indipendenza ($p\text{-value} = 1.326e-05$, $\alpha = 0.05$) e tra regione e gruppo di laurea ($p\text{-value} = 2.2e-16$, $\alpha = 0.05$). Questo legame sottolinea, con una buona probabilità, una differenza, in termini lavorativi, tra le zone d'Italia. [vedi appendice]

Indagine

Si vuole verificare se l'indice di soddisfazione sulla stabilità e sicurezza del lavoro sono indipendenti dagli attributi riguardanti il tipo di contratto, successivamente confrontare la relazione tra questi due parametri sulle differenti classi di titoli di studio.

Cleaning e preprocessing

Di seguito le operazioni iniziali di preparazione dei dati per i successivi test:

1. Vengono scartati tutti i record contenenti almeno un valore pari a <NA> nelle colonne L2_17, L2_22, L2_54B e GRUPPO_mIcro.
2. Le due colonne L2_17 e L2_22, entrambe descriventi le caratteristiche del contratto sono unite in un'unica colonna, i cui valori appartengono all'intervallo [1,4].
3. Viene effettuato un campionamento stratificato rispetto alla soddisfazione per ognuno dei quattro contratti di lavoro, ottenendo per ognuno, campioni di stessa numerosità pari al minore di questi.
4. Divisione in quattro intervalli dell'indice di soddisfazione:

Modalità	Descrizione
0	Soddisfazione pessima $\rightarrow \{0, 1, 2\}$
1	Soddisfazione bassa $\rightarrow \{3, 4, 5\}$
2	Soddisfazione media $\rightarrow \{6, 7\}$
3	Soddisfazione alta $\rightarrow \{8, 9, 10\}$

Inferenza

Essendo i due caratteri sotto indagine, contratto e soddisfazione, divisi entrambi in 4 classi, si vuole osservare se un certo livello di soddisfazione è indipendente dal gruppo di appartenenza della tipologia del contratto. Date le condizioni del problema, il test Chi-Quadro per l'indipendenza si rivela uno strumento adeguato. La seguente tabella mostra le frequenze relative degli individui appartenenti ad una certa classe di contratto e l'indice di soddisfazione da loro fornito.

		Indice di soddisfazione			
		0. Pessima	1. Bassa	2. Media	3. Alta
Tipologia di contratto	1. Indeterminato tempo pieno	0.02	0.07	0.18	0.73
	2. Indeterminato part time	0.09	0.18	0.28	0.45
	3. Determinato tempo pieno	0.13	0.20	0.30	0.37
	4. Determinato part time	0.26	0.31	0.25	0.18

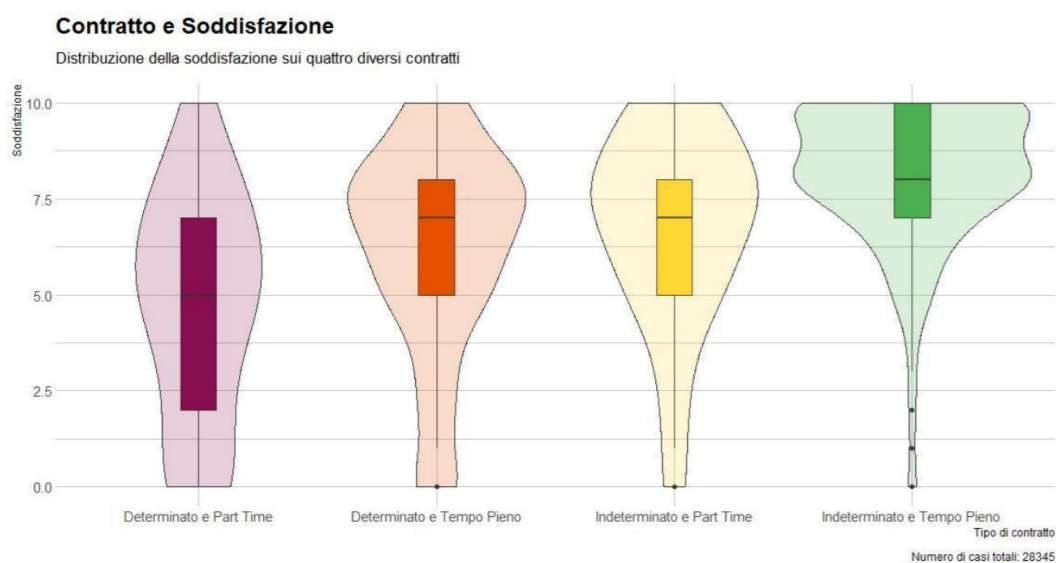
Ipotesi nulla H_0 : Il tipo di contratto lavorativo e l'indice di soddisfazione sulla sicurezza e stabilità lavorativa sono due caratteri indipendenti.

Ipotesi alternativa H_1 : L'indice di soddisfazione sulla sicurezza e stabilità lavorativa è condizionato dalla tipologia contrattuale.

Il test Chi-Quadro, condotto a 9 gradi di libertà e con livello di significatività $\alpha = 0.05$, fornisce i seguenti risultati:

X-squared	1677.5
C regione critica	(3.325, $+\infty$)
P-value	$2,2 \cdot 10^{-16}$

È evidente che la statistica X-squared cade ben oltre l'inizio della regione critica, così come il p-value è estremamente inferiore al livello di significatività α . Pertanto, l'ipotesi nulla va fortemente rifiutata essendo nettamente in contrasto con i risultati osservabili nel test.



La differenza attesa ed evidente si ha infatti tra il gruppo “contratto indeterminato a tempo pieno” e “contratto determinato part time”, si può notare in aggiunta che la maggior parte della distribuzione, all'interno dell'interquartile, per quanto riguarda il primo è superiore alla mediana, mentre per il secondo accade l'opposto.

2.5 Soddisfazione stabilità e sicurezza del lavoro delle diverse tipologie di contratto lavorativo nei diversi gruppi di laurea.

Indagine

Effettuata questa dimostrazione ci si è chiesto se questa relazione sussista tra tutti i gruppi di laurea di provenienza presi separatamente; e dunque se tutti i titoli di studio portino ad una soddisfazione simile dato il proprio contratto lavorativo.

Si è scelto quindi di confrontare le media del grado di soddisfazioni tra i sedici gruppi di laurea, dividendo i campioni per tipo di contratto in quattro categorie.

Cleaning e preprocessing

1. Per ognuna delle categorie contrattuali, vengono estratte le soddisfazioni dei campioni di ogni gruppo di laurea.
2. Per ognuna delle categorie contrattuali, viene eseguito un test ANOVA sulle medie di soddisfazione di ogni corso di laurea, ognuno con livello di confidenza α pari a 0.05.

- Essendo il test molto sensibile alle variazioni nella numerosità dei campioni, prima di eseguire ciascun test, ogni gruppo di laurea è sotto campionato in modo stratificato rispetto alla soddisfazione, ad un numero di campioni eguale tra i gruppi, pari al numero dei campioni appartenente al minore di questi. (Sono stati esclusi i laureati del gruppo “Difesa e Sicurezza” in quanto si hanno davvero pochissimi campioni).

Inferenza

Ipotesi nulla H_0 : il grado di non differisce in media tra le diverse classi di laurea.

Ipotesi alternativa H_1 : il grado di soddisfazione è mediamente diverso tra le diverse classi di laurea.

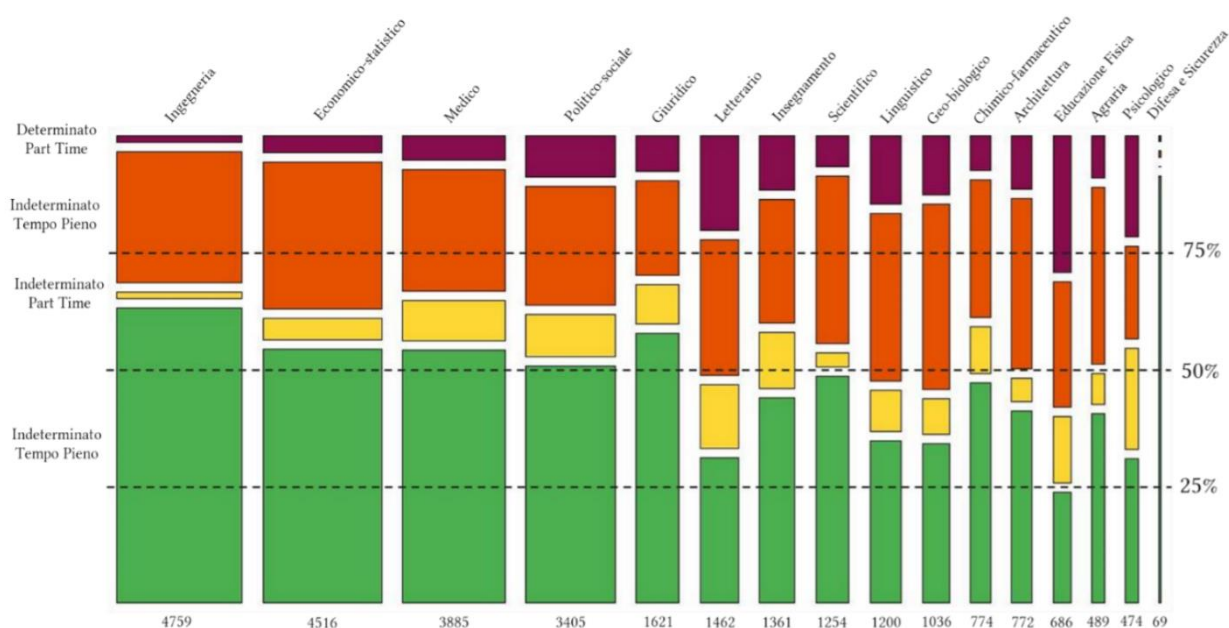
Di seguito la tabella dei risultati:

	Indeterminato tempo pieno	Indeterminato part time	Determinato tempo pieno	Determinato part time
Min	7.65	5.83	5.18	4.23
Mean	8.06	6.64	5.95	4.72
Max	8.37	7.35	6.85	5.53
P-value	0.226	0.384	1.15 e-4	0.627

Si può notare che riguardo al test eseguito sul primo, secondo e quarto tipo di contratto, non possiamo rifiutare l'ipotesi nulla che le medie di soddisfazione relative al contratto di lavoro siano uguali tra i diversi gruppi di laurea. Il test eseguito sul terzo tipo di contratto invece porta al rifiuto di questa.

In conclusione, si può dire che generalmente non vi è differenza nella soddisfazione relativa al contratto di lavoro al variare del titolo conseguito. Il caso escluso è quello relativo al contratto “Determinato a Tempo Pieno”, la cui soddisfazione, non si può dire uguale tra gruppi di laurea, fattore dovuto a un'alta dispersione dei valori medi di soddisfazione tra differenti classi, visibile nella dimensione del range, il che potrebbe condurre a osservare che questo tipo di contratto intacca la soddisfazione in maniera più o meno intensa a seconda della laurea di provenienza (supponendo impieghi inerenti al titolo conseguito).

Pur essendo vero che tutti i laureati, indipendentemente dal loro titolo, una volta ottenuto quel contratto hanno soddisfazione simile, non è detto che tutti i laureati, sempre indipendentemente dal loro titolo, abbiano la stessa probabilità di ottenere quel tipo di contratto soddisfacente. Con la seguente grafica si è cercato proprio di rappresentare questa situazione.



Proporzioni dei diversi tipi di contratti ottenuti dai diversi gruppi di laureati.

Dal grafico si nota come le distribuzioni dei contratti siano sproporzionate nei diversi tipi di laurea, vi sono casi in cui il contratto a tempo indeterminato e tempo pieno supera il 50% e casi in cui non va oltre il 25%. Inoltre, è possibile intravedere come i corsi di laurea più scelti siano anche quelli con maggiore frequenza relativa al contratto “Indeterminato e Tempo Pieno”, denotando un possibile rischio di non avere più persone che posseggono titoli di studio assolutamente necessari per alcuni tipi di impieghi per il fatto che questi non offrono al momento un buon livello di soddisfazione legato alla sicurezza e alla stabilità del posto lavorativo.

3. Conclusioni

Combinando i risultati ottenuti nei precedenti test, possiamo dedurre che la causa maggiore alla base della voglia di trasferirsi, sia una minor presenza di lavoro al di fuori del nord Italia alla quale consegue un minor numero di mesi necessari a trovare un primo impiego (vedi test mesi per trovare il primo lavoro). Altre possibili cause derivano dalla disparità dei contratti tra settentrione e meridione, le quali potrebbero essere una possibile spiegazione delle differenze relative ai percorsi formativi intrapresi nelle varie regioni ma non causa incidente sulla soddisfazione che il particolare corso di laurea garantisce ai suoi studenti.

Assumendo quindi che gli spostamenti siano una conseguenza del ridotto numero di posti di lavoro disponibili, il quale è direttamente dipendente al numero di persone che risiedono in quella stessa regione di conseguenza i trasferimenti dei laureati, alla ricerca di lavoro, graveranno in maniera sempre maggiore sul numero stesso di lavori presenti nella regione di partenza e sulla crescita e lo sviluppo della specifica zona d’Italia.

Appendice

Dati non stratificati

Regione condizionata al contratto

		Contratto				Totale
		Indeterminato tempo pieno	Indeterminato part time	Determinato tempo pieno	Determinato part time	
Regione	Nord	3477	244	2023	324	6068
	Centro	1028	86	652	162	1928
	Sud	370	74	169	51	664
	Isole	229	39	104	42	414
Totale		5104	443	2948	579	9704

Regione condizionata al gruppo di laurea

		Gruppo di laurea																Totale
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
Regione	Nord	1232	667	1504	3679	3853	1267	530	3990	2487	2301	1682	976	723	699	1060	15	26665
	Centro	394	267	706	1840	1648	625	240	1444	1486	905	816	416	350	260	445	32	11874
	Sud	405	275	676	1602	1330	382	343	1266	1226	1113	701	332	565	274	371	22	10883
	Isole	132	128	277	1039	342	291	173	419	574	545	293	120	211	190	192	7	4933
Totale		2163	1337	3163	8160	7173	2565	1286	7119	5773	4864	3492	1844	1849	1423	2068	76	54335

Test di indipendenza tra regione e contratto e tra regione e gruppo di laurea

Cleaning e preprocessing

Procedimento di raccolta e pulizia dei dati, applicato a contratto ("L2_17" e "L2_22") e gruppo di laurea ("GRUPPO_mIcro"):

1. Rimozione dei record con regione "99" (estero) o "NA" (valori mancanti).
2. Rimozione dei record nei quali gli attributi "L2_17" (indeterminato – determinato), "L2_22" (pieno – part-time) o "GRUPPO_mIcro" avevano valore "NA".
3. Raggruppamento delle regioni in quattro zone: nord, centro, sud e isole.
4. Campionamento stratificato dell' attributo "contratto" e "gruppo di laurea" rispetto ai gruppi di regioni.

Dati stratificati

		Contratto				Totale
		Indeterminato tempo pieno	Indeterminato part time	Determinato tempo pieno	Determinato part time	
Regione	Nord	237	17	138	22	414
	Centro	221	18	140	35	414
	Sud	231	46	105	32	414
	Isole	227	39	103	42	411
Totale		916	120	486	131	1653

		Gruppo di laurea																Totale
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
Regione	Nord	228	123	278	681	713	234	98	738	460	426	311	181	134	129	196	3	4933
	Centro	164	111	293	764	685	260	100	600	617	376	339	173	145	108	185	13	4933
	Sud	184	125	306	726	603	173	155	574	556	504	318	150	256	124	168	10	4932
	Isole	131	127	274	1029	339	288	171	415	568	540	290	119	209	188	190	7	4885
Totale		707	486	1151	3200	2340	955	524	2327	2201	1846	1258	623	744	549	739	33	19683

Contratto:

Ipotesi nulla H_0 : "regione e contratto sono indipendenti".

Ovvero, sono distribuiti in maniera indipendente in ogni zona d'Italia.

Ipotesi alternativa H_1 : "regione e contratto sono dipendenti".

Il test Chi Quadro di Pearson, calcolato con 9 gdl, ottenendo un valore $38.662 > 3.325$, rifiuta fortemente l'ipotesi nulla con un p-value di $1.326e-05$ (con confidenza al 95%, $\alpha = 0.05$).

Gruppo di laurea:

Ipotesi nulla H_0 : "regione e laureati per gruppo di laurea sono indipendenti".

Ipotesi alternativa H_1 : "regione e laureati per gruppo di laurea sono dipendenti".

Il test Chi Quadro di Pearson, calcolato con 45 gdl, ottenendo un valore $593 > 61.66$, rifiuta fortemente l'ipotesi nulla con un p-value di $2.2e-16$ (con confidenza al 95%, $\alpha = 0.05$).

Osservazioni:

Da queste statistiche si evince che le distribuzioni, dei contratti e dei gruppi di laurea, sono uguali in ogni zona d'Italia senza maggiori differenze; sono in misura nettamente differente, invece, le quantità di laureati presenti in ogni zona d'Italia. [vedi tabella marginali]

Formulario:

Chisquare per l'indipendenza

Valore atteso delle due variabili

$$E_{i,j} = \frac{\sum_{k=1}^c O_{i,j} \sum_{k=1}^r O_{k,j}}{N}$$

X² chi quadro per l'indipendenza

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}$$

O_{i,j} = valore osservato

gdl = (r - 1)(c - 1)

Proportion test (z test)

$$\text{Statistica } z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

\hat{p} (valore) proporzione osservata

p_0 (valore) proporzione attesa

One way ANOVA

Statistica per ipotesi H₀:

$$F = \frac{SSBG / k - 1}{SSWG / n - k} \sim F_{k-1; n-k}$$

SSBG: variabilità dovuta al fattore = $\sum_{i=1}^s (\bar{y}_i - \bar{y})^2$

SSWG: è la variabilità legata al campionamento casuale (errore) = $\sum_{i=1}^s \sum_{k=1}^{n_i} (y_{ik} - \bar{y})^2$