

**UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA**

Scuola di Scienze

Corso di laurea magistrale in Data Science

**Relazione Foundation of Statistics**

**Lavoro svolto da:**

*Cesare Colzani 821253*

*Gabriele Strano 866563*

*Monica Vivace 820470*

**Anno Accademico 2020-2021**

# Contents

<b>1</b>	<b>Introduzione</b>	<b>1</b>
1.1	Dataset . . . . .	1
1.2	Obiettivi . . . . .	2
<b>2</b>	<b>Analisi relazione tra variabili</b>	<b>3</b>
2.1	Voto di maturità e cittadinanza . . . . .	3
2.1.1	ANOVA ad una via . . . . .	7
2.2	Regioni italiane e cittadinanza . . . . .	8
2.3	Scuola pubblica e privata e iscrizione universitaria . . . . .	9
<b>3</b>	<b>Analisi del reddito netto mensile</b>	<b>12</b>
3.1	Reddito netto mensile e cittadinanza . . . . .	12
3.2	Analisi di regressione lineare sui redditi . . . . .	17
3.3	Analisi di regressione multipla sui redditi . . . . .	18
<b>4</b>	<b>Conclusioni</b>	<b>19</b>

# Chapter 1

## Introduzione

### 1.1 Dataset

Per svolgere i nostri studi, abbiamo utilizzato il dataset pubblicato dall'ISTAT relativo all'indagine sui percorsi di studio e di lavoro dei diplomati.

L'indagine campionaria è parte integrante del Sistema di rilevazioni sulla transizione dal mondo dell'istruzione a quello del lavoro.

Fanno parte del Sistema anche altre due rilevazioni: l'Indagine campionaria sull' "Inserimento professionale dei laureati" e l'indagine totale sull' "Inserimento professionale dei dottori di ricerca". Le tre rilevazioni hanno come principale obiettivo quello di rilevare la condizione occupazionale dei giovani a una prefissata distanza dal conseguimento del titolo (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.

La popolazione di riferimento è rappresentata dagli individui che nell'anno scolastico 2010-11 hanno conseguito un diploma di scuola superiore in una scuola italiana, si tratta di 450.869 individui appartenenti a 6.471 scuole diverse.

Da questa popolazione è stato estratto il campione di diplomati intervistati a circa quattro anni dal conseguimento del titolo, che utilizziamo in questo elaborato.

Il dataset è composto da 10880 record e 161 variabili sia quantitative che qualitative (ordinali o nominali).

Le variabili principalmente utilizzate in questo elaborato sono state:

- Cittadinanza: variabile qualitativa nominale dicotomica, le cui possibili modalità sono "Italiana" o "Straniera o apolide".
- Iscrizione università: variabile qualitativa nominale dicotomica, le cui possibili modalità sono "Sì" o "No".
- Reddito netto: variabile quantitativa continua misurata in Euro, è la somma del reddito principale e di eventuali redditi secondari.
- Regione di residenza: variabile qualitativa nominale, le cui modalità sono 21, comprendendo così al suo interno, ogni regione italiana più la modalità "Esteri".
- Tipo scuola di provenienza: variabile qualitativa nominale dicotomica, le cui possibili modalità sono "Scuola pubblica" o "Scuola privata".

- Voto diploma: variabile quantitativa discreta, che può assumere valori compresi tra 60 e 101 che corrisponde al 100 e Lode.

## 1.2 Obiettivi

Nel corso di questo elaborato svolgeremo alcune analisi che risultano particolarmente interessanti e utili per capire il contesto italiano relativo ai giovani neo diplomati. In particolare andremo a studiare quale relazione intercorre fra la cittadinanza e il voto di maturità, la distribuzione degli studenti stranieri fra le varie scuole italiane e il rapporto fra la scuola di provenienza (pubblica o privata) e la volontà dei giovani neo-diplomati di proseguire il percorso di studi.

L'obiettivo di questo elaborato è inoltre quello di individuare quali variabili tra quelle a nostra disposizione riescano maggiormente ad influenzare il reddito netto mensile dei giovani a quattro anni dal conseguimento del diploma. Studieremo in modo distinto il reddito mensile netto degli studenti stranieri o apolidi e degli italiani cercando di mettere in luce eventuali differenze di opportunità offerte dal mondo del lavoro ad entrambi i gruppi considerati.

Proveremo inoltre a verificare l'esistenza di una relazione tra il reddito netto ed il voto di diploma tramite l'analisi di regressione lineare. Dato il diffuso convincimento che a voti di diploma crescenti corrispondano redditi netti mensili crescenti ci aspettiamo dei valori che confermino questa teoria. Svolgeremo poi una ulteriore analisi di regressione multipla in cui includeremo come variabili indipendenti il voto di diploma e la dimestichezza in internet.

Infine svolgeremo l'analisi della varianza (ANOVA) ad una via per confermare o contraddire i risultati ottenuti dallo studio della relazione fra il voto di diploma e la cittadinanza.

# Chapter 2

## Analisi relazione tra variabili

### 2.1 Voto di maturità e cittadinanza

La presenza di alunni con cittadinanza non italiana sul territorio nazionale rappresenta un fenomeno strutturale e ormai consolidato. Secondo uno studio della Fondazione Ismu, la presenza di questi ultimi si è registrata inizialmente alla fine degli anni '80 e si incrementa con sempre maggiore rapidità. Si è infatti passati da circa 59.000 alunni nell'anno scolastico 1996/1997 con un'incidenza sulla popolazione scolastica complessiva dello 0.7% ai 711.000 nell'anno scolastico 2010/2011 con un'incidenza pari al 7.9%.

Il ritardo scolastico degli studenti stranieri è tutt'oggi significativo, tuttavia sembra annullarsi per gli studenti con diversa cittadinanza che però sono nati nel nostro Paese.

Utilizzando i dati a nostra disposizione vogliamo svolgere un'analisi prendendo in considerazione il voto di maturità e la cittadinanza su un campione di 10880 studenti. Nello specifico la variabile “voto di maturità” è quantitativa discreta e può assumere valori da 60 a 101 compresi, la variabile cittadinanza invece è qualitativa e può assumere solo due categorie: “Italiana” e “Straniera o apolide”.

Nel nostro campione gli studenti italiani sono 10.567 circa il 97% del campione mentre gli studenti stranieri o apolidi sono 313 e rappresentano una piccola percentuale pari al 3%.

Per poter effettuare un confronto oggettivo sulla bravura degli italiani e degli stranieri è stato necessario calcolare le frequenze relative.

I dati emersi dai calcoli effettuati sono i seguenti:

Indici	Media	Moda	Mediana	Varianza	Dev. Standard
	73.91814	60	72	113.7765	10.66661

Table 2.1: Indici di posizione voto di diploma studenti italiani

Indici	Media	Moda	Mediana	Varianza	Dev. Standard
	73.21406	60	72	108.3408	10.40869

Table 2.2: Indici di posizione voto di diploma studenti stranieri o apolidi

Quantili	Min	0.25	0.50	0.75	Max
	60	65	72	80	101

Table 2.3: Quantili voto di diploma studenti italiani

Quantili	Min	0.25	0.50	0.75	Max
	60	65	72	80	100

Table 2.4: Quantili voto di diploma studenti stranieri o apolidi

Ciò che emerge dai dati ottenuti è una somiglianza fra gli studenti italiani e gli studenti stranieri o apolidi. Come è possibile notare dalle tabelle 2.1 e 2.3 sia gli studenti italiani che quelli stranieri nell'anno scolastico 2010/2011 all'esame di maturità hanno ottenuto una votazione media pari a 73.

Anche la varianza sembra essere molto simile, il voto di diploma degli italiani presenta una varianza pari a 113, mentre il voto di diploma degli stranieri presenta una varianza pari a 108, i primi sembrano quindi presentare una variabilità dei dati maggiore rispetto ai secondi e tale comportamento viene seguito di conseguenza dalla deviazione standard. La variabilità maggiore può essere spiegata anche dal range, i primi presentano infatti un range leggermente superiore rispetto ai secondi, rispettivamente pari a 41 e 40. Per entrambi i gruppi il voto più frequente è 60 mentre il valore centrale o mediana è pari a 72.

Anche per ciò che concerne i quantili entrambi i gruppi sembrano essere caratterizzati da una distribuzione molto simile, il 25% degli studenti sia italiani che stranieri ha preso un voto inferiore o uguale a 65, il 50% ha preso un voto inferiore uguale a 72 mentre il 75% ha preso una votazione inferiore o pari a 80.

Il grafico seguente descrive le distribuzioni di frequenze relative. La linea rossa rappresenta la media dei voti di diploma.

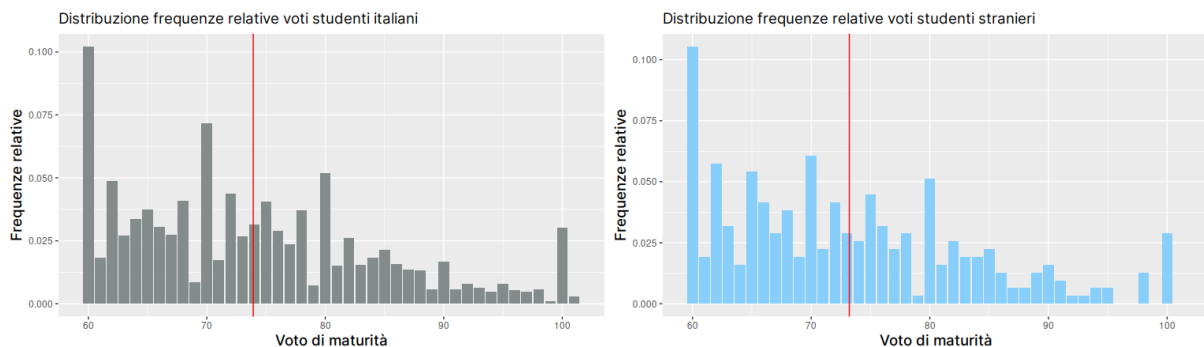


Figure 2.1: Distribuzione frequenze relative voti di maturità studenti italiani e stranieri o apolidi

Per entrambi i gruppi di studenti notiamo che a voti bassi corrispondono frequenze molto alte mentre ad alti voti corrispondono frequenze basse.

Da ciò si può facilmente desumere che entrambe le distribuzioni sono caratterizzate da una asimmetria positiva con coda lunga a destra. Il calcolo dell'indice di asimmetria

conferma questa intuizione grafica. L'indice di asimmetria della distribuzione relativa agli studenti italiani e della distribuzione relativa agli studenti stranieri è pari rispettivamente a 1.69 e 1.54.

Sia il box plot sia il grafico che rappresenta la distribuzione congiunta delle variabili sembrano confermare quanto è possibile leggere da un ulteriore rapporto stilato da Miur in collaborazione con la Fondazione Ismu, i quali dichiarano che gli studenti stranieri nati in Italia sono non solo più bravi di quelli nati all'estero, ma raggiungono livelli non lontani da quelli dei loro compagni italiani da generazioni. Il grafico 2.2 mostra solo lievi differenze nelle frequenze tra gli studenti con cittadinanza italiana e gli studenti con cittadinanza straniera o apolide.

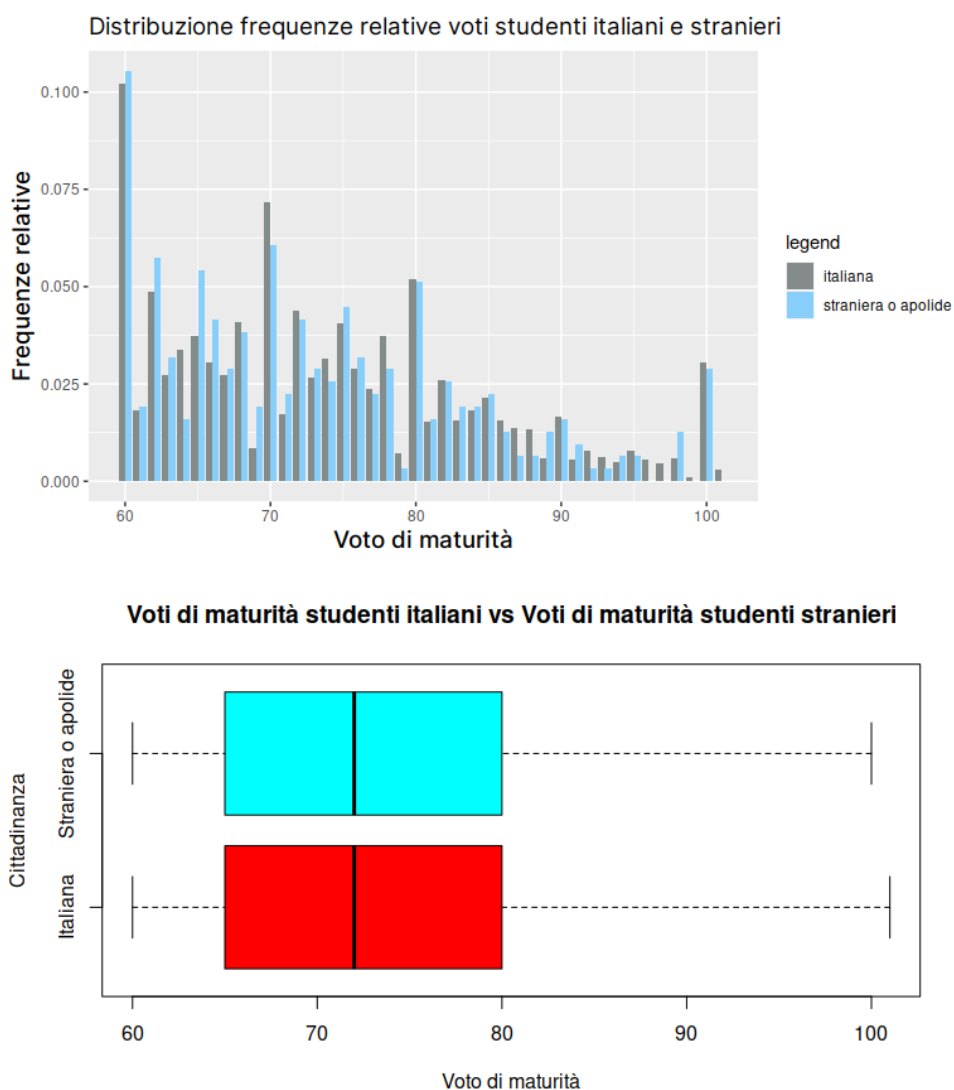


Figure 2.2: Distribuzione bivariata e box-plot

Abbiamo proseguito l'analisi della relazione fra voti e cittadinanza costruendo una tabella a doppia entrata che includesse i voti di diploma e la cittadinanza. Di seguito vengono riportate le frequenze congiunte e le relative contingenze.

Voto diploma	Italiana	Straniera o apolide	Italiana	Straniera o apolide
60	1080	33	-0.029830960	0.173328876
61	193	6	-0.019787452	0.114972391
62	516	18	-0.115821904	0.672967957
63	287	10	-0.085715474	0.498038501
64	357	5	0.288745315	-1.677716715
65	396	17	-0.255575777	1.484989470
66	324	13	-0.182684775	1.061465880
67	289	9	-0.025100407	0.145842620
68	431	12	0.035887172	-0.208517697
69	91	6	-0.330663021	1.921274039
70	756	19	0.120118133	-0.697930631
71	183	7	-0.112924775	0.656134569
72	463	13	0.032265479	-0.187474325
73	282	9	-0.037379095	0.217186321
74	333	8	0.099459069	-0.577893853
75	430	14	-0.059079174	0.343271776
76	307	10	-0.050176486	0.291543878
77	251	7	0.026674165	-0.154986732
78	393	9	0.129805849	-0.754219830
79	76	1	0.140516926	-0.816455138
80	547	16	0.008407489	-0.048850608
81	161	5	-0.017676705	0.102708172
82	276	8	0.010249236	-0.059551842
83	164	6	-0.086336116	0.501644659
84	194	6	-0.017673782	0.102691188
85	226	7	-0.019740979	0.114702367
86	165	4	0.067271368	-0.390871448
87	145	2	0.186543741	-1.083887899
88	140	2	0.177551136	-1.031637545
89	61	4	-0.268085153	1.557673559
90	176	5	0.015618222	-0.090747629
91	59	3	-0.156749074	0.910769900
92	83	1	0.156829969	-0.911239928
93	66	1	0.114975817	-0.668051877
94	51	2	-0.066243962	0.384901839
95	84	2	0.051873053	-0.301401560
96	58	0	0.222314625	-1.291729916
97	49	0	0.204339382	-1.187287133
98	61	4	-0.268085153	1.557673559
99	11	0	0.096816723	-0.562540848
100	321	9	0.027569381	-0.160188266
101	31	0	0.162530504	-0.944362141

Table 2.5: Frequenze congiunte e contingenze



Dalla tabella delle contingenze (a destra) si nota che circa metà delle contingenze presenta valori molto bassi positivi e metà valori molto bassi negativi. Il fatto che le contingenze siano basse e diverse da zero ci induce a ritenere che vi possa essere una bassa connessione fra i due caratteri e che pertanto le frequenze condizionate dei voti siano in minima parte influenzate dalle modalità dell'altro carattere (la cittadinanza).

Svolgendo il test di Pearson Chi-Quadrato, che misura la dipendenza fra due caratteri, emergono dei risultati che confermano quanto sopra affermato.

Chi-Squared Test	p-value	Chi-Squared Test Norm
28.46	0.96	0.002616

Il  $X^2$  normalizzato risulta pari a 0.002616 suggerendo quindi una bassissima connessione fra i due caratteri. Utilizzando le tavole chi quadro vediamo che definita  $\alpha$  pari a 0.05 il valore critico del chi quadro con 41 gradi di libertà è pari 56.94. Essendo il Chi quadro del nostro test pari a 28.46 e inferiore a quel valore, accettiamo l'ipotesi nulla. Pertanto i due caratteri sono indipendenti.

Un ulteriore conferma di quanto emerso deriva dallo svolgimento dell'ANOVA.

### 2.1.1 ANOVA ad una via

L'Anova è un'indagine statistica finalizzata ad individuare le differenze tra le medie campionarie attraverso l'analisi delle rispettive varianze. Possiamo confrontare le medie e le varianze tra la variabile indipendente categorica (cittadinanza) e la variabile dipendente quantitativa (voto diploma) per valutare se tali differenze sono statisticamente significative.

In questo caso avendo una sola variabile dipendente, parleremo di Anova a una via.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Cittadinanza	1	151	150.7	1.326	0.25
Residuals	10878	1236187	113.6		

L'ipotesi nulla  $H_0$  è che le varianze siano uguali fra di loro, dunque la variabile indipendente non produce effetti sulla variabile dipendente, mentre l'ipotesi alternativa  $H_1$  è che la variabile dipendente  $y$  produca effetti sulla variabile dipendente.

Il valore Pr è il cosiddetto p-value associato alla statistica F, cioè fornisce una misura della probabilità che la statistica F assuma valori estremi rispetto a quelli osservati assumendo vera l'ipotesi nulla. Fornisce pertanto la probabilità che l'influenza o l'effetto della variabile indipendente sulla variabile dipendente sia dovuta alla scelta casuale del campione e non ad una vera relazione fra le due.

Stabilendo una soglia di  $\alpha$  pari a 0.05, il p-value risulta superiore pertanto l'ipotesi nulla viene accettata.

Dal risultato ottenuto concludiamo che la relazione non è significativa, ovvero non esiste una relazione significativa tra  $y$  e  $x$ , portandoci ad accettare  $H_0$  (le due variabili sono indipendenti tra di loro).

## 2.2 Regioni italiane e cittadinanza

Secondo un rapporto dell'Istat del 2011 gli stranieri residenti in Italia tendono a distribuirsi sul territorio in maniera molto disomogenea, soprattutto in relazione alla cittadinanza di appartenenza. Considerando la popolazione straniera nel complesso, la maggior parte si concentra nel Nord (55%) e, in misura inferiore, nel Centro (25,2%), mentre nel Mezzogiorno è presente soltanto il 13,5% dei residenti stranieri.

Un'ulteriore analisi che è stata affrontata nel rapporto della Fondazione Ismu riguarda proprio la presenza degli stranieri nelle scuole diffuse sul territorio italiano.

Dal rapporto emerge che il 24,3% di essi è iscritto nelle scuole lombarde, seguono il Veneto ovvero l'11,9% (del totale della popolazione scolastica con cittadinanza non italiana), l'Emilia Romagna 11,6%, il Piemonte 9,6% e il Lazio 9,5% mentre nelle regioni meridionali, tranne che in Abruzzo, le percentuali si mantengono sotto il 3%.

Abbiamo quindi approfondito questo aspetto considerando il nostro campione di studenti. La distribuzione degli italiani e degli stranieri nelle regioni italiane è rappresentata dai seguenti grafici.

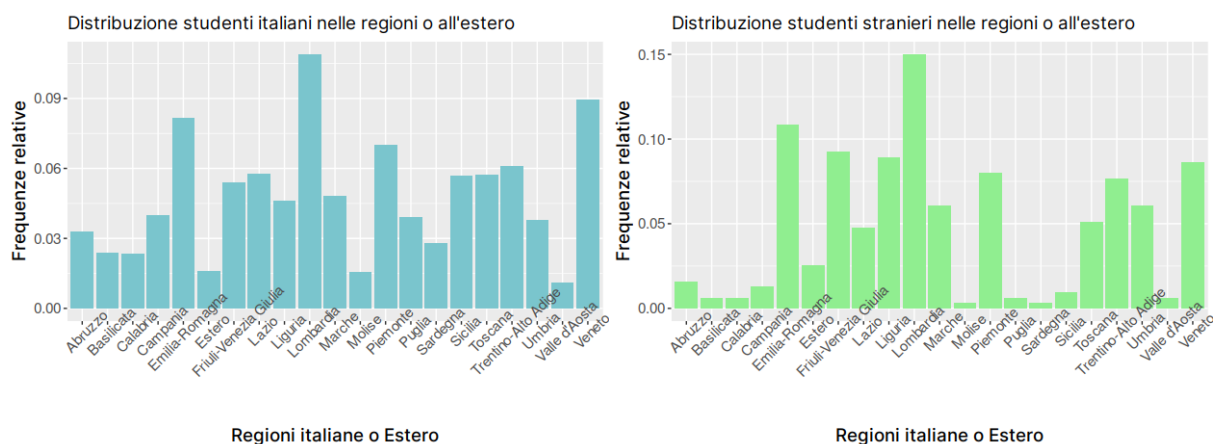


Figure 2.3: Distribuzione frequenze relative degli italiani e stranieri o apolidi nelle diverse regioni italiane o all'estero

Ciò che emerge è sicuramente una supremazia delle regioni del nord-Italia. Notiamo infatti che in Lombardia è presente il 15% (del totale del campione scolastico con cittadinanza non italiana), seguita dall'Emilia Romagna con il 10%, dal Friuli-Venezia Giulia con il 9%, Liguria con 8.9%, Veneto con l'8.6% e Piemonte con l'8%. Mentre nelle regioni del sud-Italia la presenza di stranieri è inferiore all'1.5%.

I nostri risultati sembrano pertanto confermare quanto emerge dai rapporti del Miur e della Fondazione Ismu. Questo fenomeno è dovuto essenzialmente al fatto che i flussi migratori hanno interessato e interessano tutt'oggi come destinazione residenziale i comuni di maggiori dimensioni e i comuni che presentano una maggiore offerta di lavoro e una migliore qualità della vita. In secondo luogo, un fattore determinante è la presenza di reti di relazioni e di comunità che favoriscono i ricongiungimenti familiari, pertanto in alcune regioni l'incidenza degli stranieri è maggiore.

## 2.3 Scuola pubblica e privata e iscrizione universitaria

Analizzando le due variabili qualitative (tipo di scuola di provenienza, iscrizione in università) sul campione a nostra disposizione, si evince che il numero di studenti provenienti da una scuola pubblica che si iscrivono in università sia maggiore del numero di quelli provenienti da una scuola privata, rispettivamente il 95.3% (10370 studenti) contro il 4.7% (510 studenti).

La figura 2.4 fornisce una rappresentazione grafica delle frequenze percentuali.

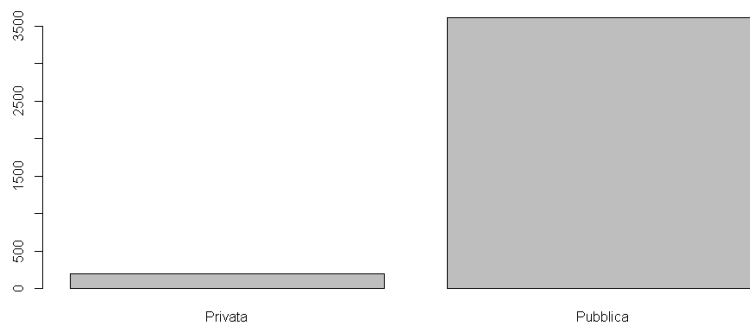


Figure 2.4: Frequenze relative degli studenti che si iscrivono all'università

Considerando le frequenze percentuali, è possibile evidenziare che chi proviene da una scuola privata è più propenso a frequentare l'università rispetto a chi proviene da una scuola pubblica.

Nonostante questo, la differenza tra le due variabili analizzate non è così elevata, quindi si può dedurre che non è dato per certo che chi frequenta una scuola privata, in futuro voglia proseguire gli studi. A dimostrazione di questo possiamo vedere nel grafico 2.5 che il 39.4% degli studenti che provenendo da una scuola privata si iscrivono all'università, contro il 34.8% degli studenti provenienti da un tipo di istruzione pubblica.

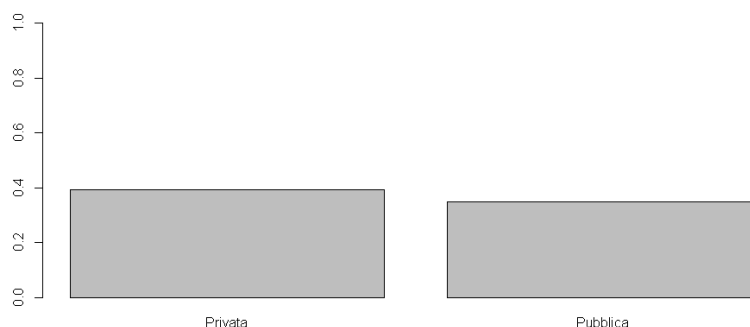


Figure 2.5: Frequenze relative degli studenti che non si iscrivono all'università

Dai valori del dataset possiamo affermare che sul totale del campione di 10880 individui, solo il 35% (3809) di esso, sia che abbiano frequentato una scuola pubblica o privata

si iscrivono successivamente in università, contro il 75% (7071) che decide di invece di terminare gli studi.

Nella rappresentazione sottostante sono riportati due diversi grafici a barre, dove si dimostra quanto detto precedentemente.

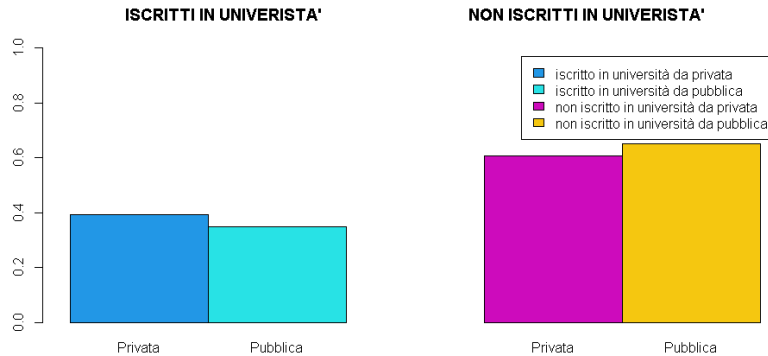


Figure 2.6: Tasso di iscrizione e di impiego

Ora possiamo andare a dimostrare mediante il test Chi Quadro se le due variabili analizzate sono statisticamente dipendenti o indipendenti.

Partiamo definendo  $H_0$ , ovvero l'ipotesi di indipendenza tra le due variabili (iscrizione in università e tipo di scuola), e l'ipotesi alternativa  $H_1$  di dipendenza tra le due variabili.

Definiamo il livello di significatività a priori del test  $\alpha=0.05$  e in seguito calcoliamo il p-value.

Utilizzando la tabella delle contingenze calcolata partendo dalle frequenze congiunte possiamo implementare il test Chi Quadro, con cui andremo ad ottenere il p-value:

	Italiani	Stranieri	tot
Privata freq. assoluta	201	309	510
Pubblica freq. assoluta	3608	6762	10370
Tot. assoluta	3809	7071	10880

Table 2.6: Frequenze congiunte assolute iscrizione universitaria degli studenti italiani e stranieri

	Italiani	Stranieri	tot
Privata freq. relativa	0.394	0.606	0.047
Pubblica freq. relativa	0.348	0.652	0.953
Tot. relativa	0.35	0.65	1

Table 2.7: Frequenze congiunte relative iscrizione universitaria degli studenti italiani e stranieri

Chi-Squared Test	df	Chi-Squared Test Norm
4.558265	1	0.0328

Essendo il p-value  $\prec 0.05$  possiamo rifiutare l'ipotesi nulla e confermare che le due variabili analizzate sono statisticamente dipendenti.

Partendo dal campione abbiamo provato l'esistenza di relazione di dipendenza ad un livello di significatività del 5%. Nonostante questo si può affermare che sulla base del campione la presunta relazione di dipendenza non è così forte, poiché se avessimo scelto un livello di significatività del test inferiore al 3.8% avremmo ottenuto un p-value  $\succ \alpha$ , in tal caso avremmo accettato l'ipotesi di indipendenza tra le due variabili.

# Chapter 3

## Analisi del reddito netto mensile

### 3.1 Reddito netto mensile e cittadinanza

Andremo ora ad analizzare la distribuzione dei redditi netti mensili dei giovani a 4 anni dal conseguimento del diploma. Nello specifico cercheremo di verificare l'esistenza di eventuali differenze tra i cittadini italiani ed i cittadini stranieri o apolidi residenti in Italia.

Secondo un'indagine ISTAT quasi due terzi degli stranieri nati all'estero e residenti in Italia sono immigrati per motivi di lavoro.

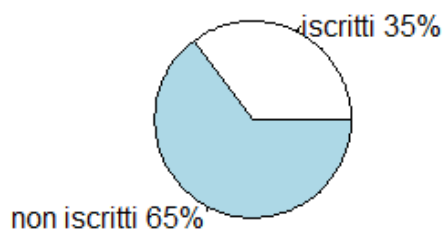
In tutte le classi di età, la quota di stranieri percettori di redditi da lavoro è più elevata di quella degli italiani (75,4% contro 66,3% tra i 15-64enni).

Nel 2019, la proporzione di giovani stranieri nella forza lavoro era superiore di quasi 11 punti percentuali a quella dei giovani italiani.

Questo maggior tasso di impiego degli stranieri nei 4 anni successivi al conseguimento del diploma, può essere in parte giustificato dal loro minor tasso di iscrizione universitaria, come mostrato dai grafici seguenti, che probabilmente si traduce in un ingresso nel mondo del lavoro anticipato rispetto ai giovani italiani.

Il grafico a destra della fig 3.1 mostra che la percentuale degli italiani (35%) che si iscrivono all'università è maggiore di 4 punti percentuali rispetto alla percentuale degli stranieri (31%).

**Tasso iscrizione universitaria  
studenti italiani**



**Tasso iscrizione universitaria  
studenti stranieri o apolidi**

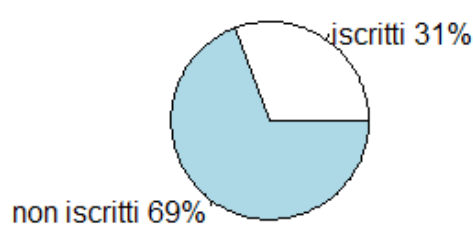


Figure 3.1: Tasso di iscrizione universitaria tra stranieri o apolidi e italiani

Un'ulteriore conferma di quanto esposto sopra l'abbiamo andando a guardare i Boxplot nella fig 3.2. Infatti è possibile notare che il baffo inferiore relativo al Boxplot degli italiani si trova più in basso rispetto a quello degli stranieri, potremmo quindi ipotizzare che il

minimo zero sia dovuto anche al fatto che nei 4 anni successivi al diploma vi sia un minor tasso d'impiego degli italiani rispetto agli stranieri.

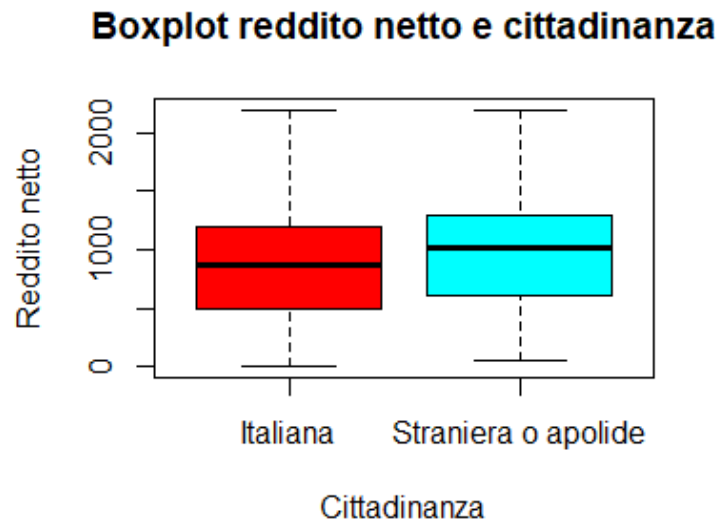


Figure 3.2: Reddito netto mensile e cittadinanza

Riguardo ai principali indici di posizione possiamo affermare che il reddito netto prodotto dagli stranieri a 4 anni dal conseguimento del diploma è mediamente superiore a quello prodotto dagli italiani (969 contro 892), questo dato viene confermato anche dalla mediana (1020 contro 870) che rispetto alla media è meno sensibile alla presenza di outliers.

Passando invece ad analizzare gli indici di variabilità, vediamo che la deviazione standard dei redditi prodotti dagli stranieri (447) è leggermente inferiore rispetto a quella relativa ai redditi prodotti dagli italiani (455).

Ciò significa che, i singoli redditi prodotti dagli stranieri tendono a discostarsi meno dalla media rispetto ai singoli redditi prodotti dagli italiani.

Nella stessa direzione va una recente indagine ISTAT, secondo la quale, tra gli italiani, i redditi da lavoro delle persone laureate sono maggiori del 75% rispetto ai redditi delle persone con licenza elementare, mentre tra gli stranieri questa differenza si riduce all'8%.

Questo dato viene inoltre confermato dal minor coefficiente di variazione relativo agli stranieri ( $cv = 0.4606$ ), rispetto a quello relativo agli italiani ( $cv = 0.5099$ ).

Indici	Media	IQR	Mediana	Varianza	Dev. Standard
	892.8485	700	870	207216.0	455.2099

Table 3.1: Indici di posizione reddito netto mensile studenti italiani

Indici	Media	IQR	Mediana	Varianza	Dev. Standard
	969.1693	700	1020.00	199908.3	447.1110

Table 3.2: Indici di posizione reddito netto mensile studenti stranieri o apolidi

Entrambe le distribuzioni presentano lo stesso valore di IQR (range interquartile) tuttavia, guardando ai quantili delle due distribuzioni, notiamo che sia il primo quantile (0.25) che il terzo quantile (0.75) relativi ai redditi prodotti dagli stranieri sono maggiori di 100 euro rispetto a quelli prodotti dagli italiani, ciò significa che il 25% degli stranieri guadagna 100 euro in più rispetto al 25% degli italiani e che tale divario resta immutato anche guardando il 75% delle due distribuzioni.

Divario che però non è costante in tutta la distribuzione, infatti è addirittura maggiore (di 150 euro) se guardiamo al 50% delle due distribuzioni (cioè le due mediane).

Quantili	Min	0.25	0.50	0.75	Max
	0	500	870	1200	2200

Table 3.3: Quantili reddito netto mensile studenti italiani

Quantili	Min	0.25	0.50	0.75	Max
	50	600	1020	1300	2200

Table 3.4: Quantili reddito netto mensile studenti non italiani

Guardando i grafici nella fig 3.3, raffiguranti le distribuzioni delle frequenze assolute dei redditi netti di italiani (in basso a sinistra) e stranieri o apolidi (in basso a destra), emergono subito delle differenze, in quanto nonostante la forma suggerisce che entrambe le distribuzioni siano asimmetriche positive (con le code di destra più lunghe), il valore dell'indice di asimmetria (o Skewness) relativo alla distribuzione del reddito netto prodotto dagli italiani è pari a 0.3953 mentre quello relativo agli stranieri è pari a 0.0713, cioè per questi ultimi, vi è un'asimmetria molto lieve, il che viene confermato anche dalla forma “quasi simmetrica” della distribuzione.



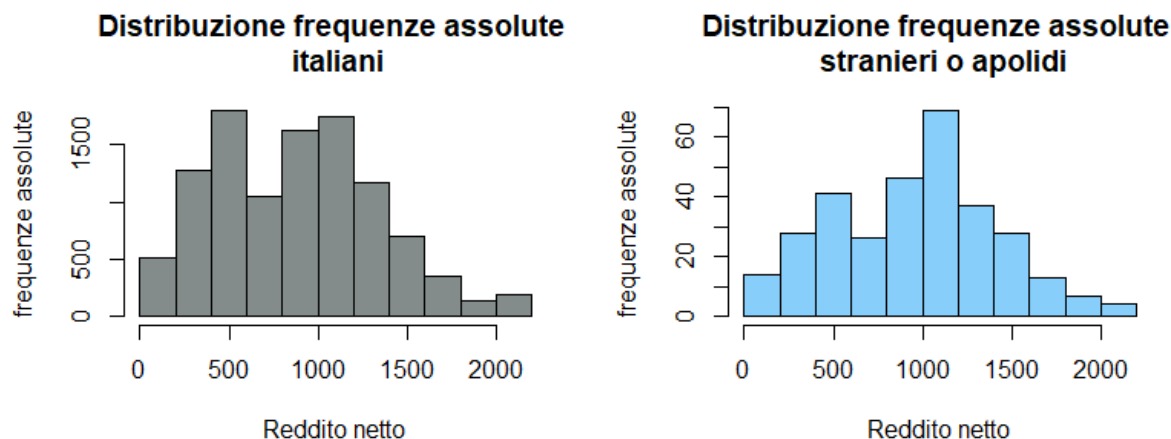


Figure 3.3: Distribuzione redditi tra studenti italiani e stranieri o apolidi

Proseguiamo la nostra analisi andando a discretizzare la variabile reddito netto, dividendola in 4 fasce di reddito di uguale ampiezza. La classe modale è quella delle 4 che presenta maggior frequenza cioè la fascia numero 2 che va da 550 euro a 1099 euro e questo dato viene confermato anche andando a calcolare la classe modale riguardante rispettivamente, i soli italiani ed i soli stranieri.

Nella figura 3.4 invece riportiamo il grafico delle frequenze relative riguardante la variabile reddito netto discretizzata e la variabile cittadinanza.

In entrambe le distribuzioni la fascia di reddito meno popolosa è quella più alta, il che era prevedibile essendo il nostro campione composto da giovani da poco diplomati.

Quello che è utile notare invece è che per gli stranieri le due fasce di reddito più rappresentative sono la seconda e la terza con frequenze relative molto vicine fra loro, mentre per gli italiani abbiamo una fascia dominante cioè la seconda e altre due fasce che presentano frequenze relative molto simili.

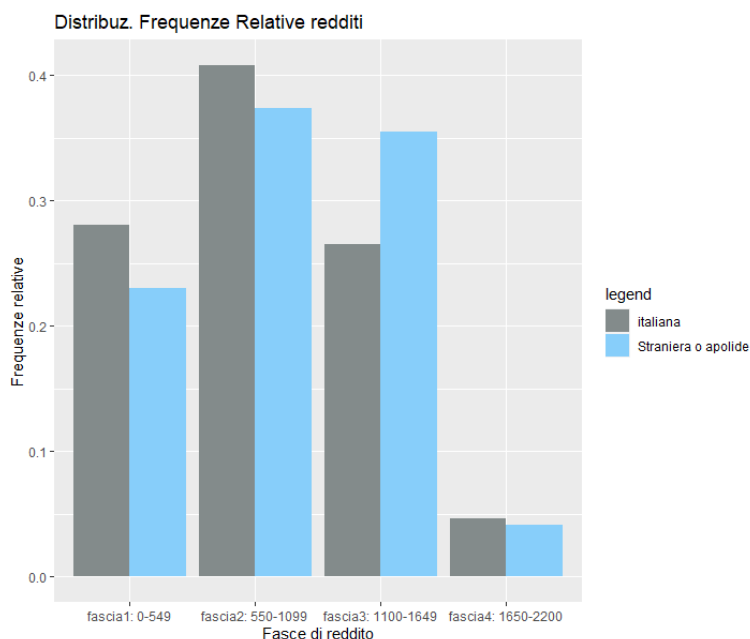


Figure 3.4: Distribuzione frequenze relative dei redditi

Di seguito abbiamo inserito una tabella a doppia entrata che mostra la relazione tra il reddito netto e la cittadinanza, in particolare vengono riportate le frequenze congiunte e le relative contingenze.

Fascia di reddito	Italiana	Straniera o apolide
Fascia1: 0-549	0.28154457	-1.63587771
Fascia2: 550-1049	0.15795893	-0.91779958
Fascia3: 1100-1649	-0.51007238	2.96370853
Fascia4: 1650-2200	0.06776441	-0.39373618

Table 3.5: Tabella delle contingenze

Andare a guardare la tabella delle contingenze (frequenze congiunte osservate – frequenze teoriche) ci permette di mettere in evidenza le attrazioni (valori positivi) o le repulsioni (valori negativi) fra le varie coppie di modalità.

Svolgendo il test Chi-Quadro ed interpretandone i risultati, rifiutiamo l'ipotesi nulla ossia l'indipendenza fra i due caratteri.

Il nostro Chi-Quadrato normalizzato risulta pari a 0.001178863 suggerendo quindi la presenza di una connessione di bassa entità fra i due caratteri.

Viene quindi confermato che, anche se in minima parte, la cittadinanza in una fascia di età giovanile influenza il reddito netto.

Avendo inoltre un p-value inferiore a 0.5 possiamo affermare che questa differenza di reddito netto tra i giovani italiani e i giovani stranieri o apolidi si presenta nel 95% dei campioni estratti dalla popolazione, pertanto il test è significativo.

Chi-Squared Test	p-value	Chi-Squared Test Norm
12.826	0.005028	0.001178863

A confermare ulteriormente la presenza di una relazione tra le due variabili, riportiamo infine un'ulteriore indagine ISTAT, secondo la quale in Italia la percentuale di persone italiane che ha guadagnato redditi da lavoro è maggiore rispetto alla percentuale di persone straniere che ha guadagnato redditi da lavoro. Ciò vale sia nel caso degli uomini, sia in quello delle donne: i maschi stranieri tra i 15 e i 24 anni sono percettori di reddito da lavoro nel 65,8% dei casi, contro il 36,5% dei maschi italiani nella stessa classe di età; per le femmine il differenziale è più limitato (38,9%, contro 27,3%).

Questa circostanza sembra inoltre coerente con una diversa velocità di ingresso nel mercato del lavoro da parte degli stranieri, impegnati fin da giovanissimi nel procurarsi il reddito per vivere, rispetto agli italiani che, grazie a condizioni economiche familiari relativamente più agiate, hanno percorsi di studio più lunghi.

Va in conclusione, segnalato che mentre la quota di stranieri percettori di redditi da lavoro è più elevata di quella degli italiani anche in riferimento a fasce d'età maggiori, la differenza delle medie dei redditi netti percepiti viene invece ribaltata, infatti sono gli italiani ad avere redditi medi maggiori rispetto agli stranieri, ciò è giustificato dal maggior livello d'istruzione, conseguente al maggior tasso d'iscrizione universitaria, che permette solitamente agli italiani di raggiungere posizioni lavorative più alte.

## 3.2 Analisi di regressione lineare sui redditi

In ultima analisi abbiamo svolto una regressione lineare sul reddito netto mensile per verificare se vi è una relazione tra quest'ultimo e il voto di diploma, i risultati emersi sono i seguenti:

	Estimate	Std. Error	t-value	P >   t
Intercept	1049.3556	30.5266	34.375	$< 2e - 16$
Slope	-2.0882	0.4089	-5.107	3.32e-07

Multiple R-squared	0.002392
Adjusted R-squared	0.0023

In primo luogo l'intercetta di una retta di regressione è il valore che assume la variabile dipendente nel caso in cui il coefficiente angolare (il voto di diploma) sia pari a 0. In generale affinché un modello di regressione lineare sia un buon modello predittivo in grado di spiegare la variabile dipendente in funzione di una variabile indipendente è necessario che l'intercetta sia pari o molto vicino a 0, in tal caso vorrebbe dire che il coefficiente angolare della variabile indipendente e la variabile indipendente (voto di diploma) sono in grado di spiegare la variabile dipendente (reddito netto mensile).

Dalla tabella sopra possiamo notare che il valore dell'intercetta è molto alto e pari a 1049.3556, mentre il valore del coefficiente angolare è pari a -2.0882, ciò significa che gran parte delle variazioni nel reddito non vengono catturate dal voto di diploma ma non solo, il valore negativo del coefficiente ci dice che il voto di diploma ha lievi effetti negativi sul reddito, quindi un aumento del voto sembrerebbe diminuire il livello del reddito.

Inoltre possiamo notare che l'errore standard ossia l'errore nella stima di questi valori è pari a 34.375 per l'intercetta e -5.107 per il coefficiente angolare e che il p-value è molto basso e inferiore alla soglia dello 0.05. Entrambi questi parametri insieme al t value che è il rapporto tra il valore dell'intercetta o del coefficiente angolare e il loro standard error confermano che le stime sono statisticamente significative. Da ciò è possibile confermare che il voto di maturità non è una variabile che nel nostro campione influenza il reddito netto mensile in un contratto di lavoro. Ciò viene confermato da un ulteriore parametro, l'indice di determinazione  $R^2$  il quale può variare da 0 a 1 e fornisce una misura della varianza spiegata, cioè la quota di variabilità della variabile dipendente che viene spiegata dalla variabile indipendente. Quest'ultimo parametro assume valore pari a 0.023 pertanto, solo il 2.3% dei redditi può essere spiegato dal voto di maturità.

Infine, per avere un'ulteriore conferma dell'assenza di relazione abbiamo calcolato l'indice di correlazione fra il reddito e il voto di diploma che risulta pari a -0.0489, che prova le conclusioni sopra. Ipotizzando poi che il calcolo della correlazione fra il reddito e il voto di diploma possa essere influenzata dai soggetti che decidono di proseguire gli studi iscrivendosi in università e che pertanto hanno reddito molto basso o pari a zero, abbiamo ritenuto più opportuno valutare la correlazione fra il reddito e il voto di diploma dei soli studenti che hanno deciso di non proseguire gli studi universitari. Ciò che è emerso

nuovamente è assenza di correlazione, con un valore dell'indice più alto rispetto al primo e pari a 0.038.

### 3.3 Analisi di regressione multipla sui redditi

Date le conclusioni sopra abbiamo voluto analizzare il reddito in funzione non solo del voto di diploma ma anche della dimestichezza in internet di ogni studente. Quest'ultima variabile quantitativa assume valori da 0 a 10. Anche in questo caso riportiamo i risultati ottenuti:

	Estimate	Std. Error	t-value	P >   t
Intercept	1105.517	38.850	28.456	< 2e-16
Slope1	-2.013	0.410	-4.909	9.28e-07
Slope2	-7.732	3.310	-2.336	0.0195

Multiple R-squared	0.002893
Adjusted R-squared	0.0027090

Come possiamo notare ancora una volta il valore dell'intercetta pari a 1105.517 rimane molto alto mentre i valori dei coefficienti angolari delle due variabili indipendenti (voto di diploma e dimestichezza in internet) rimangono bassi anche se il secondo coefficiente angolare sembra avere più effetto sui redditi rispetto al voto di diploma. Guardando i valori dell'errore standard, del t-value e del p-value che è inferiore a 0.05 possiamo concludere che anche in questo caso le stime sembrano essere corrette. Tutto ciò viene ulteriormente confermato dall'indice  $R^2$  che è molto basso e pari a 0.002893. Pertanto nè il voto di diploma nè la dimestichezza in internet influenzano il livello di reddito che le imprese offrono agli studenti neo-diplomati.

# Chapter 4

## Conclusioni

I risultati emersi dalla prima analisi concernente il voto di maturità e la cittadinanza sembrano confermare i risultati ufficiali dell'Istat, del Miur e della fondazione Ismu. Gli studenti italiani e gli studenti stranieri del nostro campione sono caratterizzati da prestazioni scolastiche molto simili, pertanto se nel passato le discrepanze sembravano più accentuate ad oggi possiamo dire che si stanno affievolendo.

Andando a svolgere un'analisi bivariata, notiamo che vi è assenza di relazione fra il voto di diploma e la cittadinanza. Passando dalla cittadinanza italiana a quella europea i voti del diploma non variano in modo consistente. Ulteriore conferma del fatto che la cittadinanza non può e non deve essere utilizzata per fare discriminazione tra gli studenti; le capacità cognitive e pertanto la bravura scolastica dipendono da moltissimi altri fattori che andrebbero presi come riferimento se si vuole analizzare questa variabile.

I risultati dell'ANOVA si mostrano in linea con quanto sopra. In definitiva possiamo concludere che non è possibile mostrare una chiara relazione fra le due variabili.

Anche gli esiti del secondo studio sulla distribuzione degli studenti stranieri nelle varie scuole di tutta Italia avvalorano il primato delle regioni del nord-Italia nell'ospitare gli studenti stranieri. In Lombardia infatti troviamo il 15% degli studenti stranieri, in Emilia Romagna il 10%, in Friuli-Venezia Giulia il 9%, in Liguria l'8.9% seguiti da Veneto e Piemonte. Scarsa è invece la presenza degli stessi nelle regioni del sud-Italia dove la percentuale risulta inferiore all'1.5%.

Dall'analisi tra il tipo di scuola (pubblica e privata) e iscrizione universitaria viene messa in evidenza una maggiore tendenza degli studenti che frequentano scuole pubbliche a frequentare l'università rispetto agli studenti che provengono da una scuola privata, tuttavia questa differenza non è così significativa.

Infine dallo studio svolto sul reddito e sulle variabili che potrebbero essere legate ad esso emergono alcune importanti informazioni.

La maggioranza degli studenti stranieri a quattro anni dal diploma tende a entrare nel mondo del lavoro prima rispetto agli studenti italiani e questo spiegherebbe anche perché le iscrizioni universitarie degli studenti italiani siano maggiori rispetto agli studenti stranieri.

Inoltre il reddito netto prodotto dagli stranieri a 4 anni dal conseguimento del diploma è mediamente superiore a quello prodotto dagli italiani, ciò probabilmente è dovuto al fatto che all'interno dell'analisi sono stati inclusi gli studenti che si iscrivono in università a tempo pieno e che pertanto non percepiscono reddito.

Mentre dalla regressione lineare e dalla regressione multipla emerge che né il voto di diploma né la dimestichezza in internet possono essere considerati fattori appropriati a spiegare le variazioni di reddito.

Variabili che possono essere prese in considerazione in questo caso sono molteplici, pensiamo ad esempio alla performance del lavoratore, al valore aggiunto che viene fornito alla società o azienda, capacità del soggetto nel lavorare in squadra e così via.

Per concludere possiamo dire che il dataset utilizzato ci ha consentito di confermare alcune delle conclusioni a cui sono giunti istituti nazionali e pubblici tramite le loro ricerche, così come ci ha consentito di contraddire alcune opinioni diffuse nella società.