

Indagine sulle intenzioni di mobilità degli studenti campani

Stefania Cardone, Marta Cicco, Antonio Cola, Maristella Simonetti, Rosario Urso

1 Descrizione dell'indagine

Da marzo a giugno 2022, alcuni studenti delle magistrali di Scienze Statistiche per le Decisioni e Relazioni Internazionali hanno collaborato, insieme al **Professore Giancarlo Ragozini**, alla realizzazione di una ricerca dal titolo *“Indagini sulle intenzioni di mobilità degli studenti campani”* che fa parte del progetto PRIN dal titolo *“Student mobility at University level: Statistical models and Network approaches”*.

2 Obiettivo e disegno dell'indagine

L'obiettivo dell'indagine è di analizzare le intenzioni dei ragazzi che hanno raggiunto la fine del percorso di studi secondari di secondo grado dal punto di vista formativo e migratorio verso istituti universitari fuori dalla Campania. Lo scopo è avere risposte sul perché gli studenti campani preferiscono continuare altrove i propri studi.

La nostra indagine è stata condotta su un collettivo iniziale di 28540 studenti. È stato deciso di applicare due filtri: le scuole che hanno un numero di iscritti all'università ≥ 20 e almeno 5 movers. Questi filtri ci hanno consentito di ridurre la numerosità fino a raggiungere un collettivo finale di 25289 studenti. Questo collettivo corrisponde ad un totale selezionato di 309 scuole di cui 221 licei, 9 professionali e 79 tecnici.

Le strategie di campionamento inizialmente proposte sono state due: Campionamento stratificato autoponderante è stata fissata la numerosità delle scuole che si volevano campionare a 40; tale numero è stato moltiplicato per la percentuale di incidenza di ciascuna scuola sul totale. Questa strategia è corretta se assumiamo che l'unità campionaria è la scuola. Al termine di questo campionamento abbiamo ottenuto i seguenti risultati:

CAMPIONAMENTO 1	LICEI	PROFESSIONALI	TECNICI
Avellino	3	0	1
Benevento	2	0	1
Caserta	4	0	2
Napoli	15	1	6
Salerno	6	0	2
TOTALI	29	1	10

Campionamento casuale con estrazione basata sulla numerosità degli studenti si stabiliscono le quote delle scuole da campionare sulla base del numero medio di studenti iscritti all'anno scolastico. Ciascuna delle percentuali ottenute è stata moltiplicata per 2500, ovvero la numerosità di studenti che si voleva considerare nel campione; i dati ottenuti sono stati divisi per la numerosità media degli iscritti all'anno scolastico che risultano proseguire gli studi dopo il diploma.

Tale procedimento ha consentito di ottenere i seguenti risultati:

CAMPIONAMENTO 2	LICEI	PROFESSIONALI	TECNICI	DA CAMPIONARE
Avellino	2	0	0	2
Benevento	1	0	0	2
Caserta	5	0	1	6
Napoli	17	0	3	20
Salerno	6	0	1	7
TOTALI	32	0	5	37

La nostra scelta è ricaduta sul secondo tipo di campionamento, al quale abbiamo apportato delle modifiche affinché il campione includesse almeno un istituto professionale e uno tecnico per provincia. Si è raggiunto così ad un totale complessivo di 44 scuole secondarie di secondo grado.

CAMPIONE SCUOLE	LICEI	PROFESSIONALI	TECNICI
Avellino	2	1	1
Benevento	1	1	1
Caserta	5	1	1
Napoli	17	1	1
Salerno	6	1	1
TOTALI	32	5	7

Al termine della raccolta dei dati sono state applicate delle correzioni mediante un sistema di pesi costruiti facendo riferimento ai dati della popolazione, per correggere eventuali sovra-risposte o sotto-risposte.

3 Analisi esplorativa

Dopo aver caricato il dataset su R, abbiamo avuto accesso alle prime informazioni. Tra queste:

- Dimensione: 641 righe, 122 colonne, 78202 dati totali;
- Dati mancanti: 49253 (ca. il 63% dei dati totali).

Dato l'elevato numero di variabili (numero di colonne) e di dati mancanti, abbiamo deciso di ridurre il dataset alle sole variabili con meno del 70% di dati mancanti. In questo modo è stato più semplice individuare le variabili dipendenti ed esplicative da utilizzare.

Di seguito le covariate di nostro interesse: provincia, comune, sesso, età, istituto, classe, voto, studio madre, studio padre, lavoro madre, lavoro padre, orientamento, continuo studi, iscrizione, corso.

Per alcune di queste abbiamo deciso di eseguire una dicotomizzazione per una più facile rappresentazione ed interpretazione. Successivamente ne abbiamo analizzato la distribuzione:

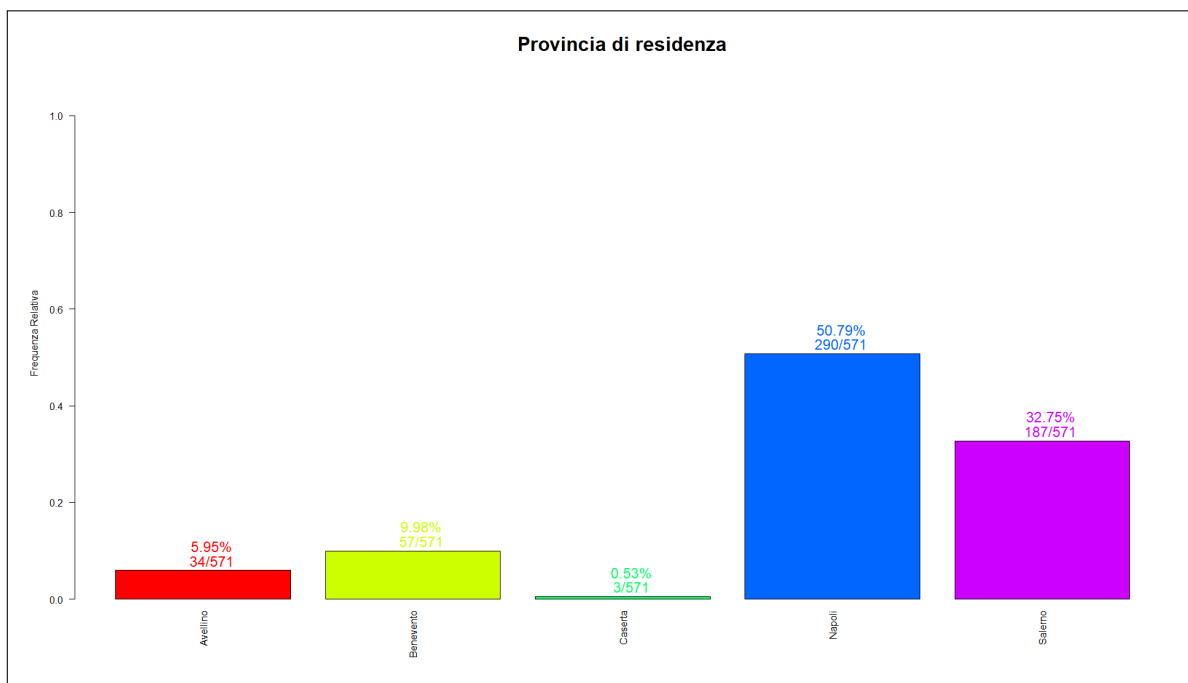


Figura 1: Distribuzione di frequenze provincia

Com'è possibile evincere dal grafico il 50% degli studenti che sono stati intervistati risiede nella provincia di Napoli, mentre il restante 50% risiede nelle altre 4 province, con una maggiore concentrazione a Salerno. Caserta risulta essere la provincia con il minore numero di studenti intervistati.

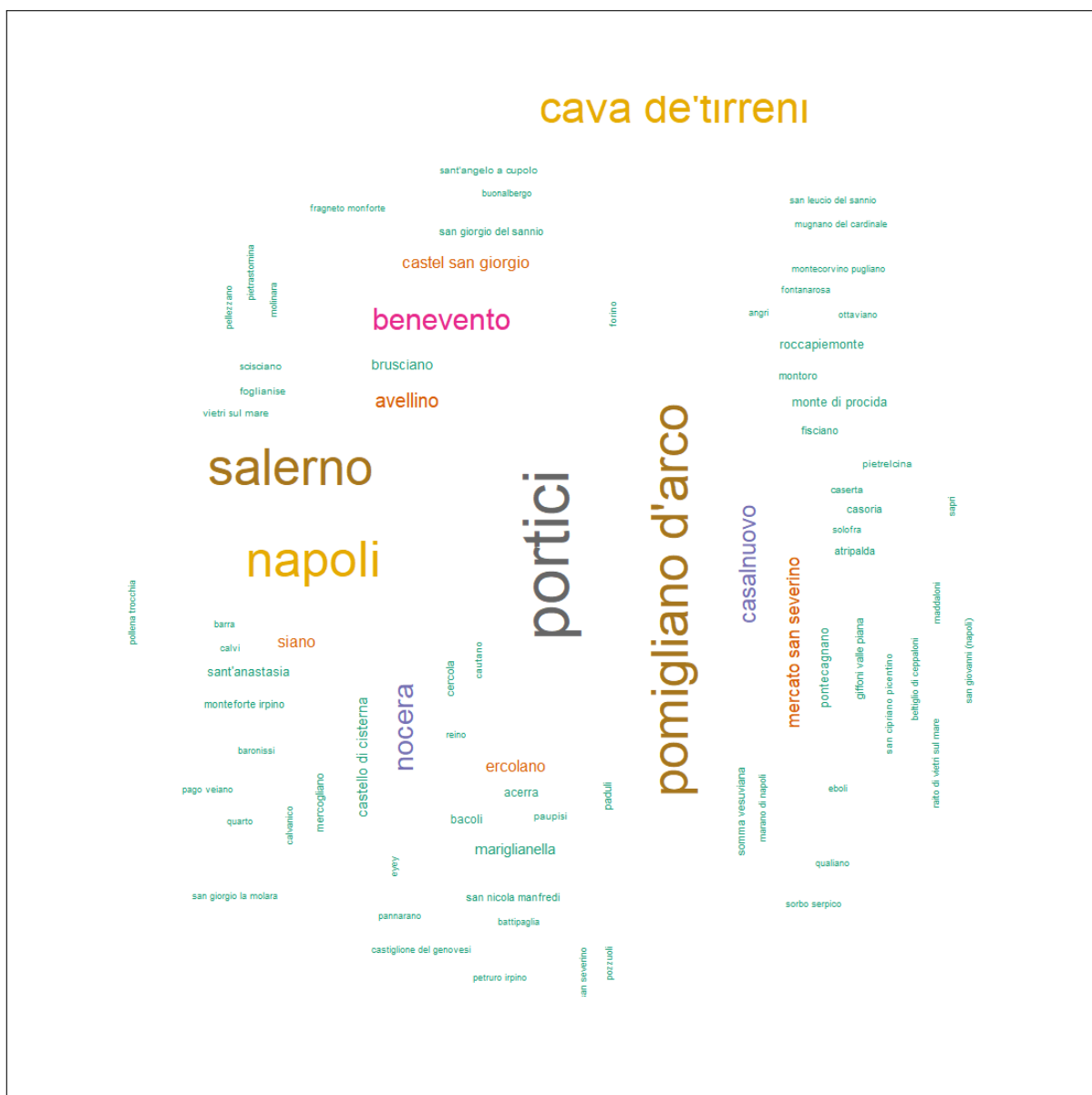


Figura 2: Wordcloud comune

Dato l'elevato numero di diversi comuni presenti nel dataset (82), abbiamo deciso di rappresentarli tramite wordcloud (rappresentazione visiva di etichette o parole-chiave). Successivamente abbiamo cercato i 10 comuni di residenza più frequenti e sono stati raffigurati con un barplot. Da quest'ultimo grafico è possibile affermare che la frequenza cumulata dei primi 6 comuni costituisce il 50% del totale. Il restante 50% è ripartito tra i restanti 74 comuni.

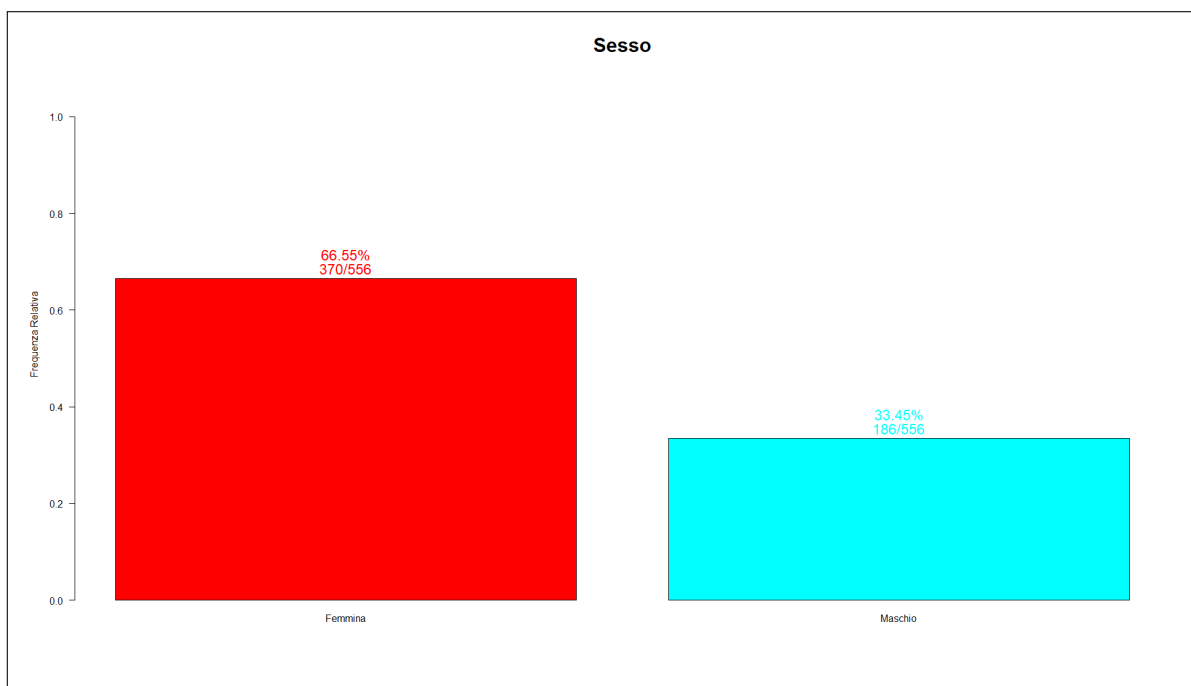


Figura 3: Distribuzione di frequenze sesso

La variabile sesso presentava 4 modalità: Femmina, Maschio, Altro, Preferisco non dichiarare. Abbiamo quindi deciso di dicotomizzarla (Femmine e Maschi), escludendo le seguenti modalità: Altro, Preferisco non dichiarare. Nel dataset ci sono 50 uomini ogni 100 donne, un dato lontano da quello rilevato del nostro Paese (95).

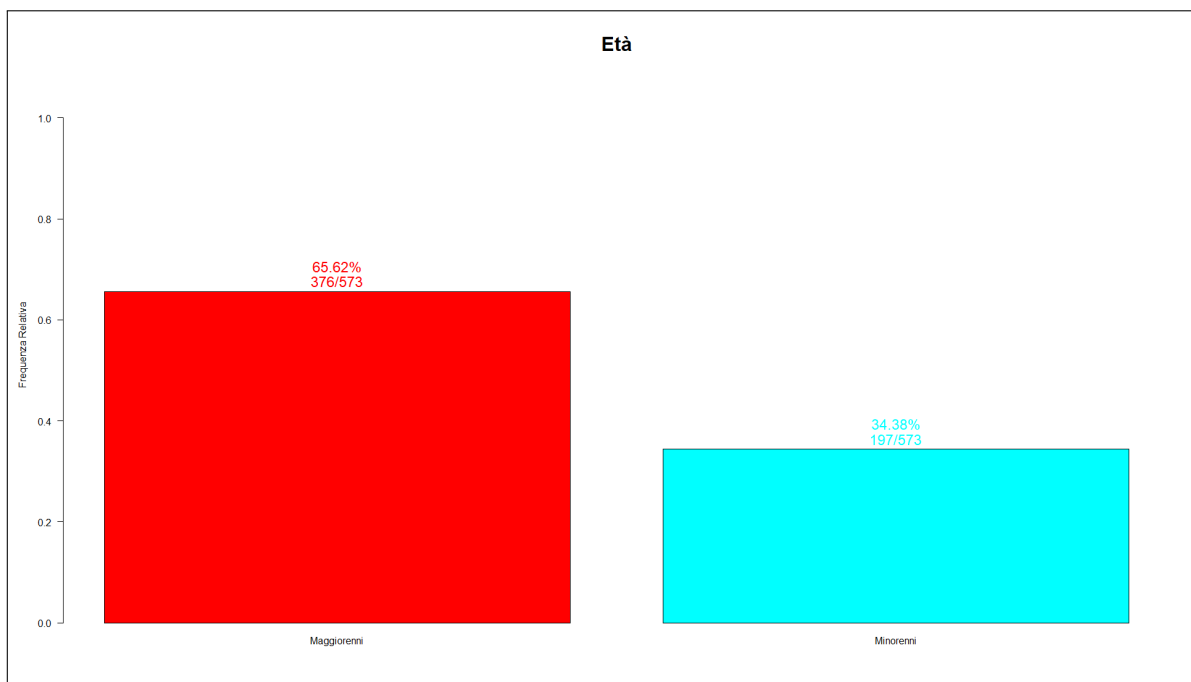


Figura 4: Distribuzione di frequenze età

La variabile età riportava diverse modalità numeriche comprese nell'intervallo $[16,29]$, per questo abbiamo ragionevolmente deciso di dicotomizzare in Maggiorenni e Minorenni. Gli studenti intervistati

sono prevalentemente maggiorenni (ca. 66%), dato plausibile, dal momento che gli studenti intervistati erano iscritti al IV o al V anno della scuola secondaria di secondo grado.

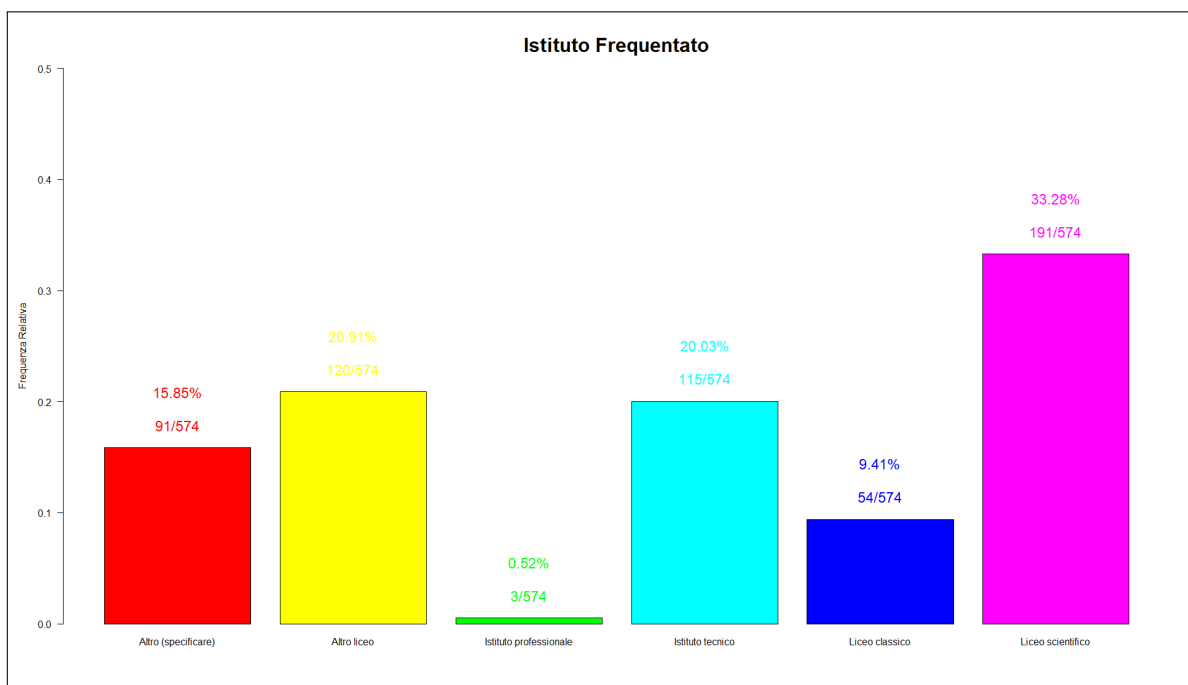


Figura 5: Distribuzione di frequenze istituto

Per quanto concerne la variabile istituto, abbiamo constatato 7 diverse modalità. Una di queste, Altro (specificare), ha creato un legame con la variabile (colonna) successiva, dove è avvenuta la specificazione. Dopo la dicotomizzazione le modalità sono: Altro istituto, Liceo. I liceali sono più del doppio degli studenti di altri istituti. Questo dato era prevedibile per via della selezione del campione.

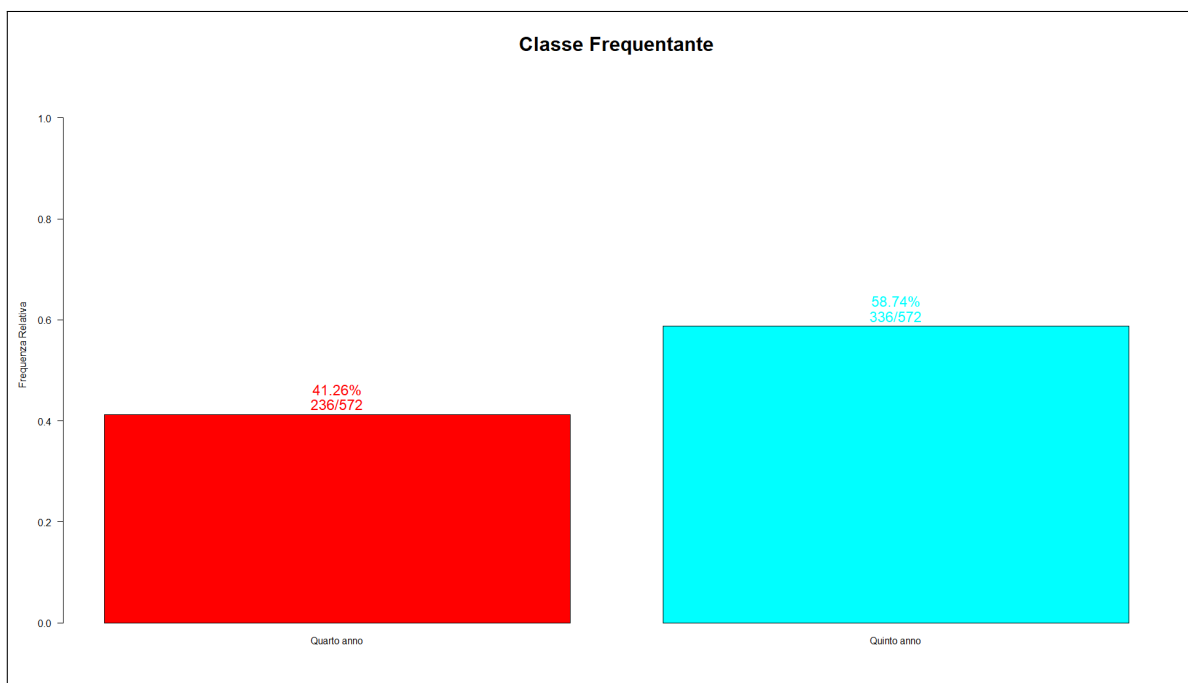


Figura 6: Distribuzione di frequenze istituto

La distribuzione di questa variabile conferma, circa, i dati ottenuti dalla precedente distribuzione di frequenze dell'età. Gli studenti delle classi V sono in superiorità numerica rispetto agli studenti delle classi IV.

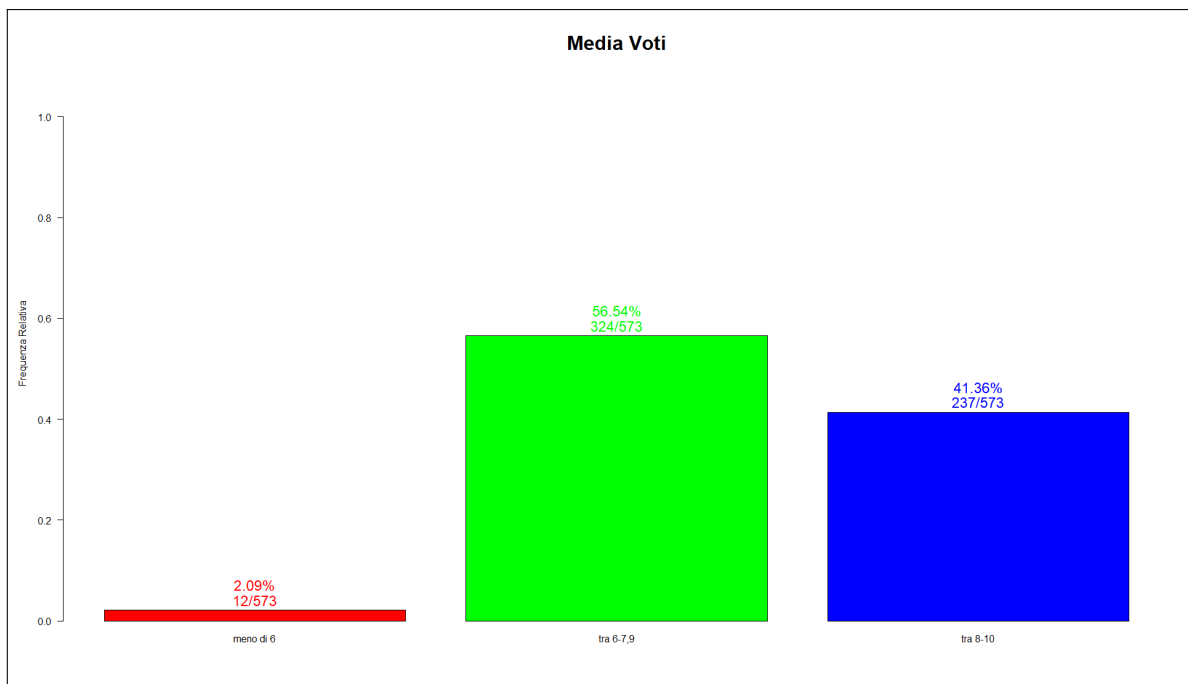


Figura 7: Distribuzione di frequenze istituto

Più del 97% degli studenti intervistati ha una media di voti compresa nell'intervallo [6,10].

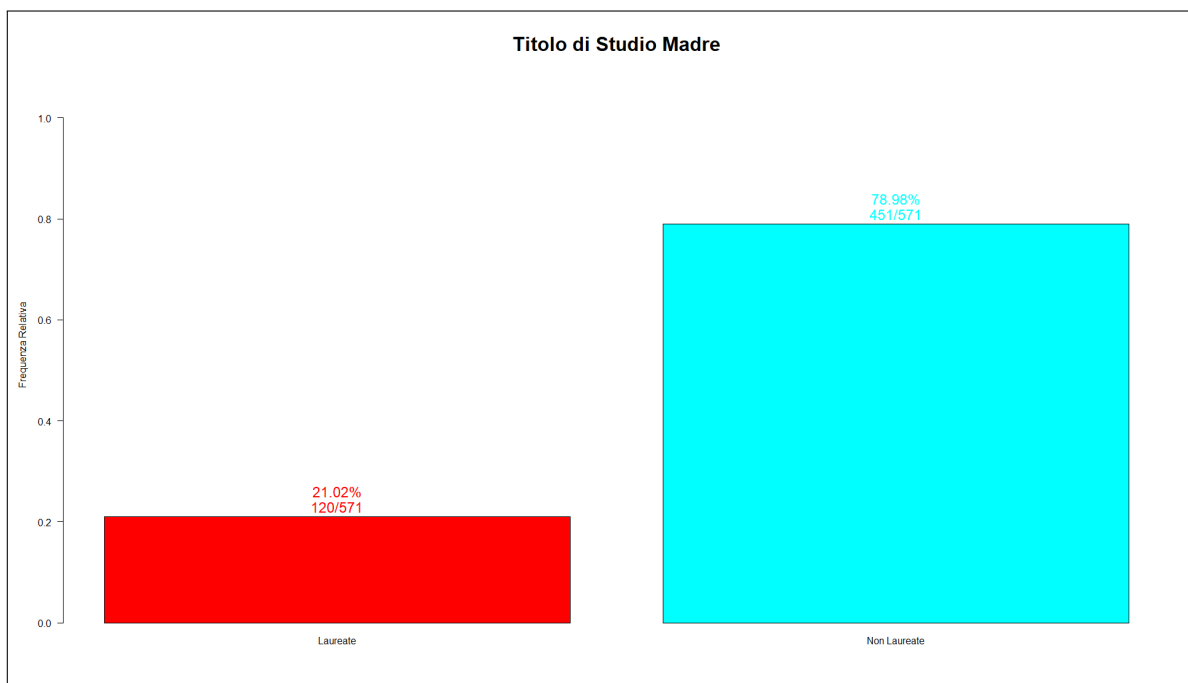


Figura 8: Distribuzione di frequenze titolo madre

In seguito alla dicotomizzazione di questa variabile possiamo affermare che solo il 21% delle madri sono laureate.

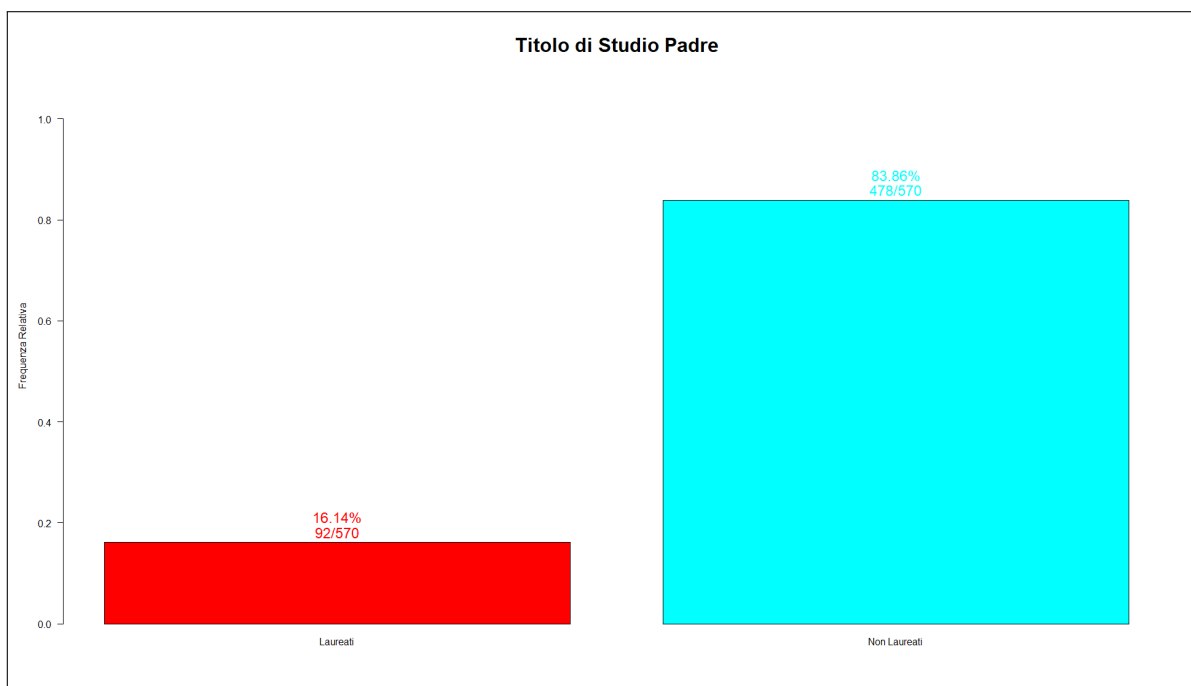


Figura 9: Distribuzione di frequenze titolo padre

Come per la variabile precedente, anche in questo caso il numero di laureati è inferiore ai non laureati. La percentuale di padri laureati (16%) è addirittura più bassa di quella delle madri.

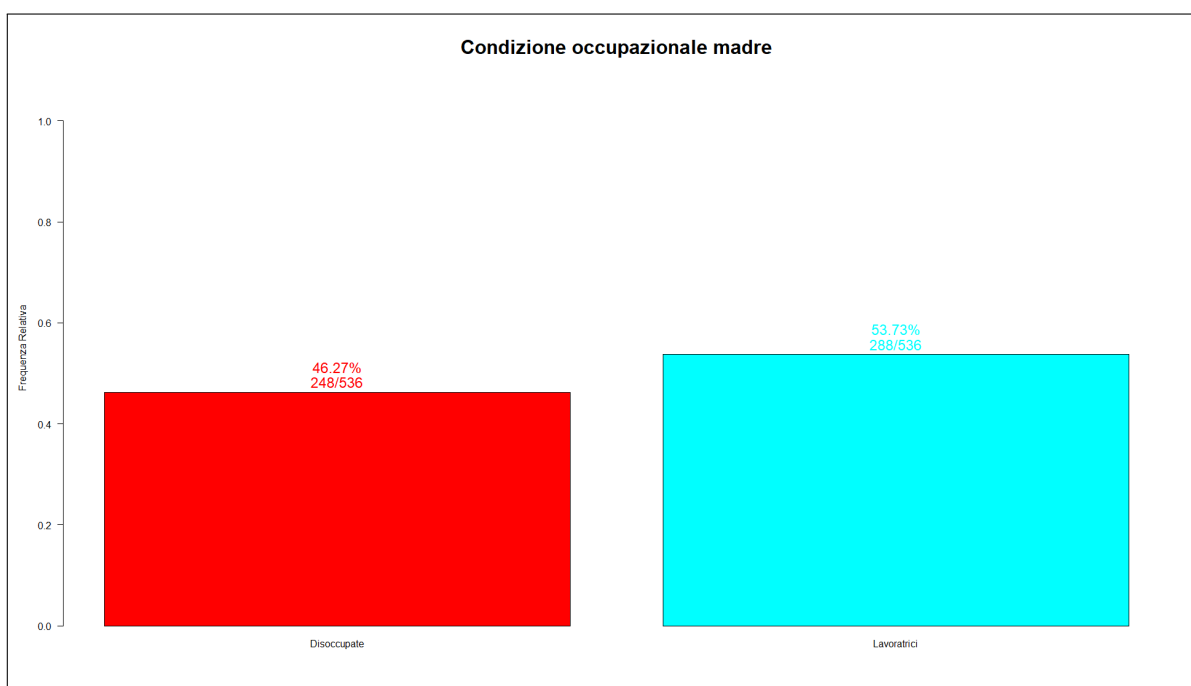


Figura 10: Distribuzione di frequenze occupazione madre

Grazie alla dicotomizzazione effettuata è possibile cogliere un quasi equilibrio tra le madri disoccupate e lavoratrici. La differenza è minima (ca. il 7%).

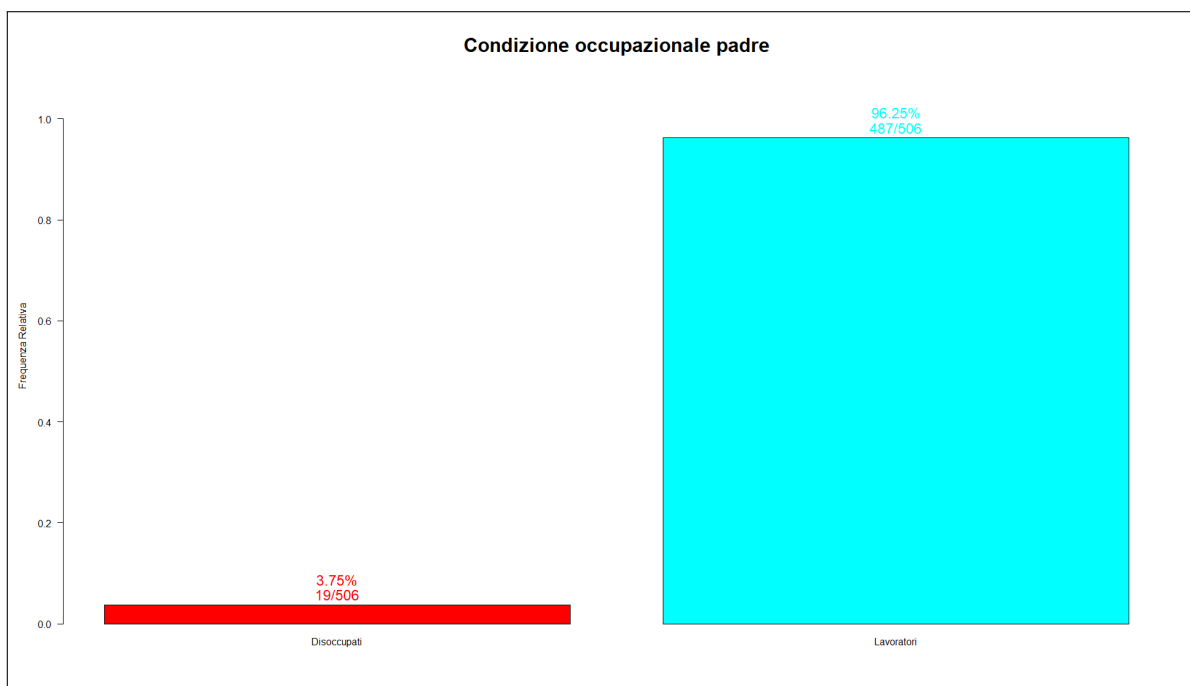


Figura 11: Distribuzione di frequenze occupazione padre

Confrontando questo grafico con il precedente possiamo notare come i padri lavoratori siano, in percentuale, molti di più rispetto alle madri lavoratrici, nonostante i padri laureati siano, in percentuale, di meno delle madri laureate.

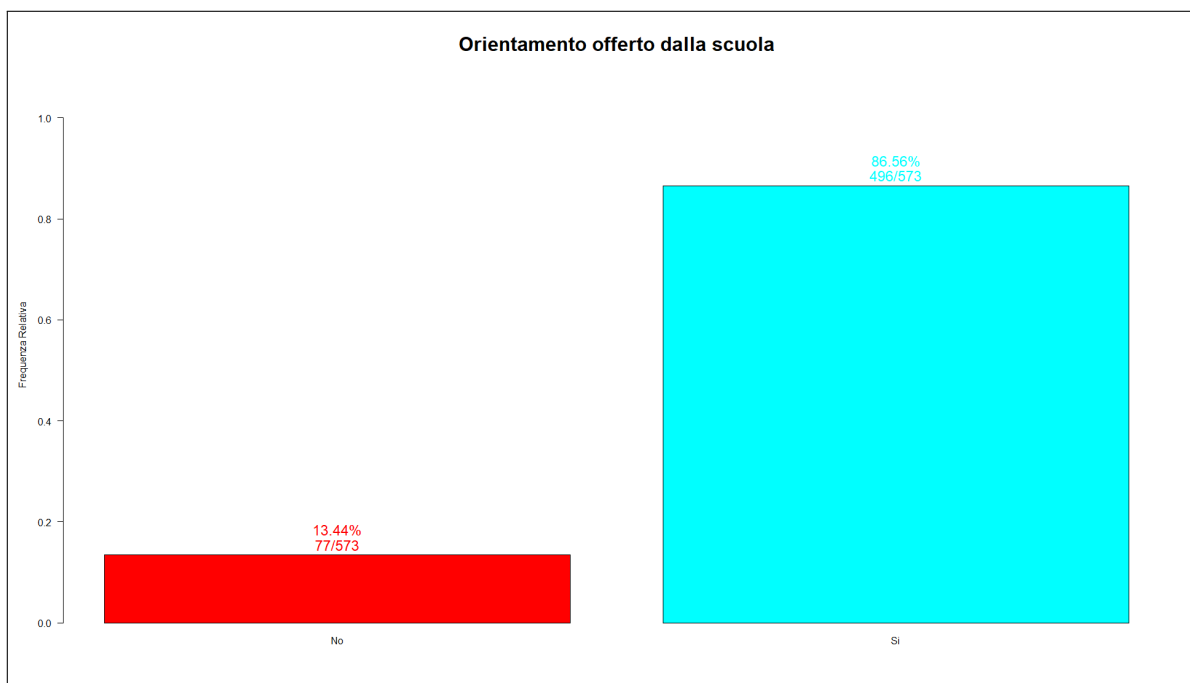


Figura 12: Distribuzione di frequenze orientamento

La maggior parte degli studenti intervistati è iscritto ad una scuola che ha offerto un servizio di orientamento universitario.

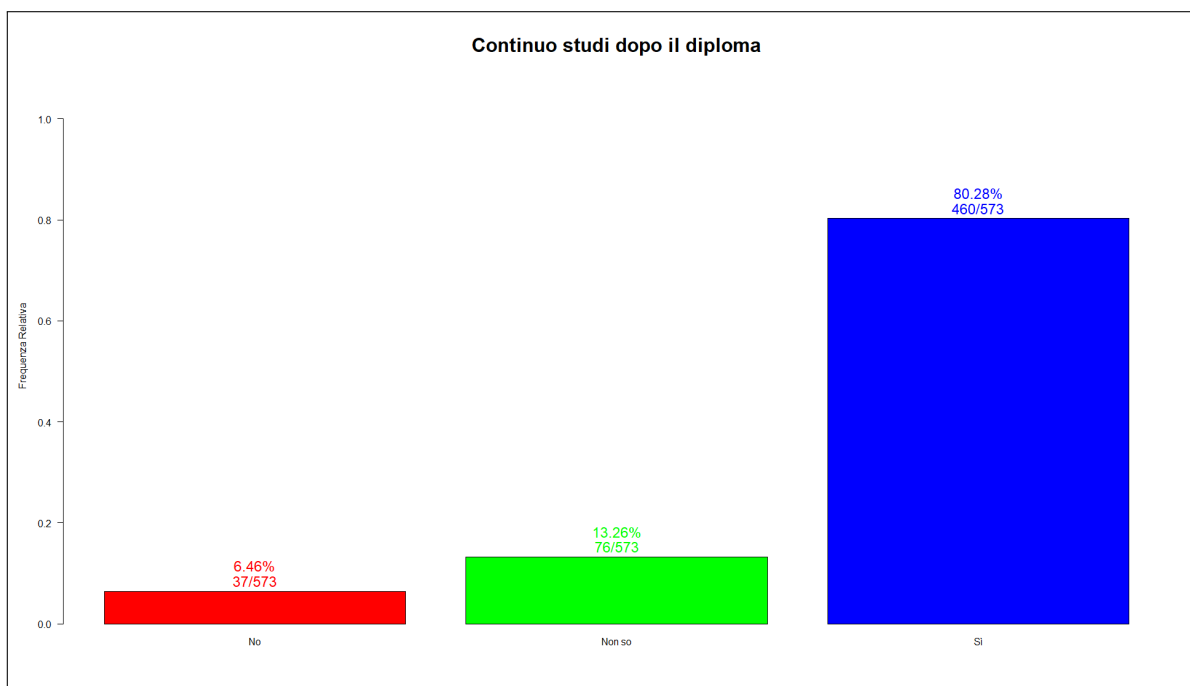


Figura 13: Distribuzione di frequenze continuo gli studi

La percentuale di studenti che ha intenzione di continuare gli studi è di poco superiore all'80%. Gli indecisi sono il 13% e solo il 6% ca. non ha intenzione di continuare gli studi.

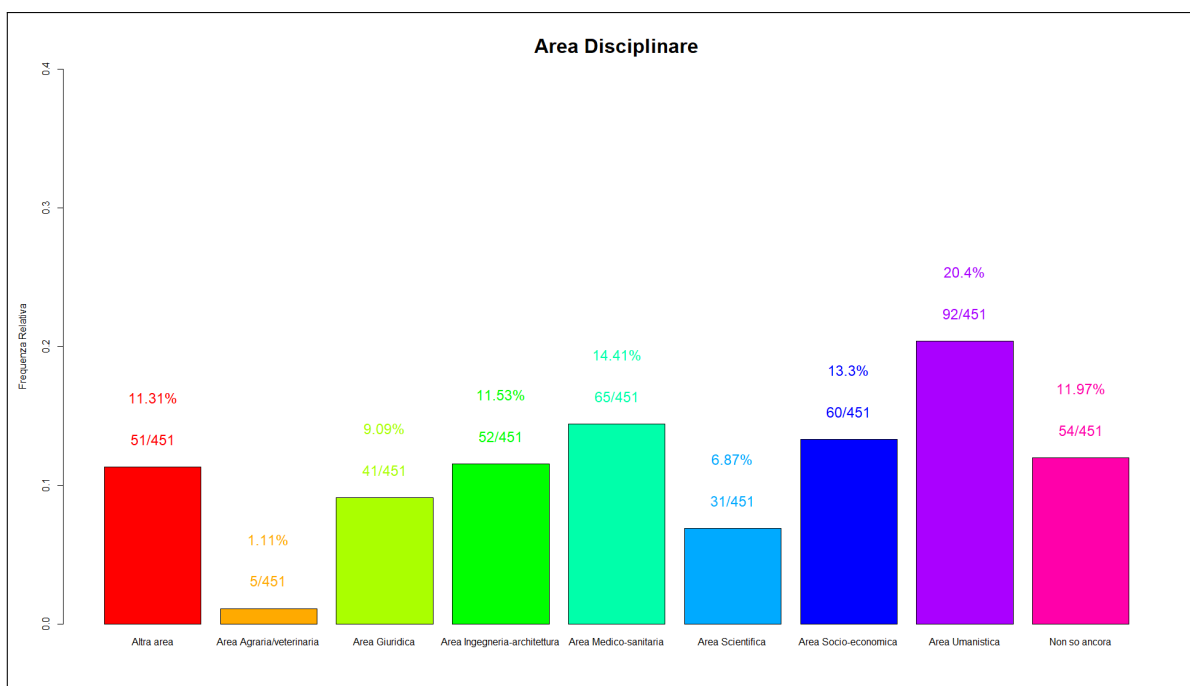


Figura 14: Distribuzione di frequenze area disciplinare

Il 50% ca. degli studenti intervistati ha espresso una preferenza per una delle seguenti aree disciplinari: Umanistica, Medico-sanitaria, Socio-economica. Molti sono gli indecisi (12% ca.).

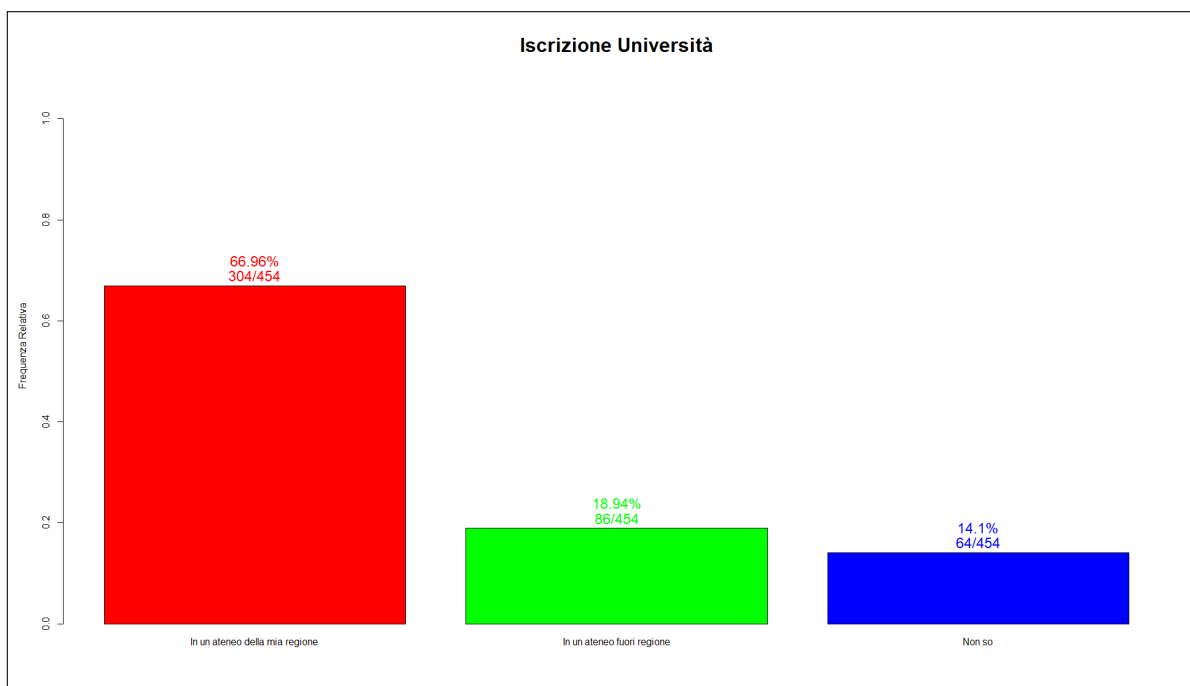


Figura 15: Distribuzione di frequenze iscrizione università

Questo è uno dei grafici più significativi per l'obiettivo dell'indagine. La maggior parte dei rispondenti ha intenzione di rimanere nella propria regione dopo il diploma, mentre il 19% ha intenzione di spostarsi altrove. Gli indecisi ricoprono una fetta non troppo grande (14%).

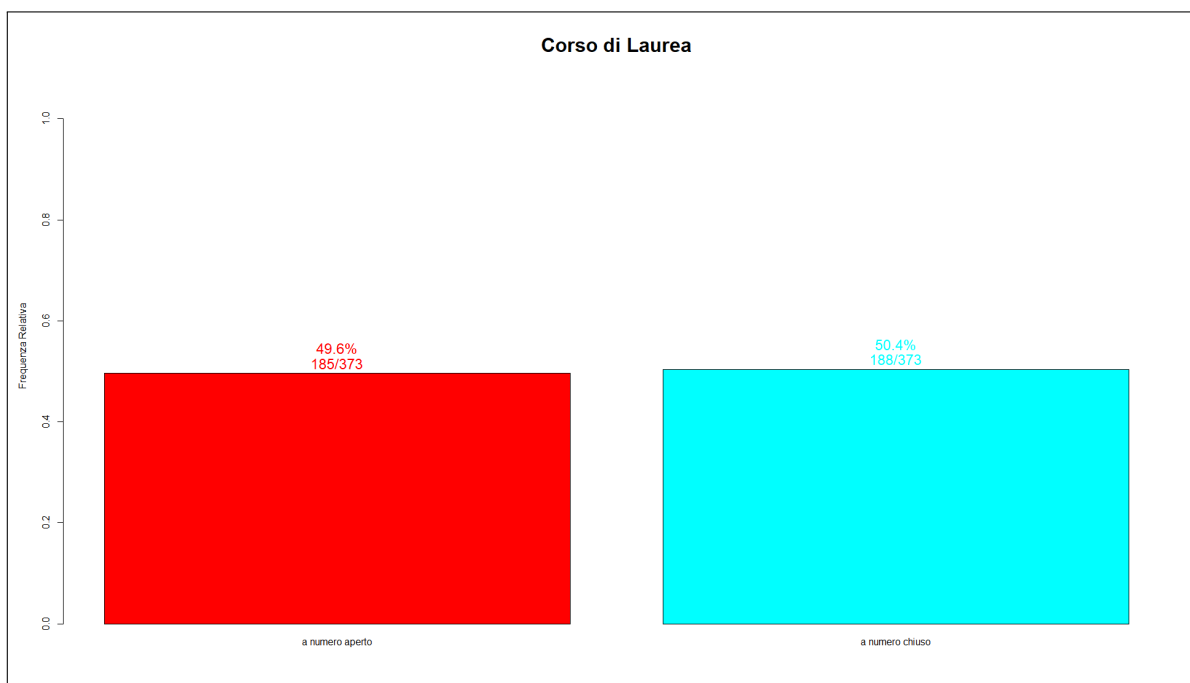


Figura 16: Distribuzione di frequenze corso di laurea

C'è un equilibrio tra gli studenti che intendono iscriversi ad un corso di laurea a numero programmato e gli studenti che intendono iscriversi ad un corso di laurea a numero aperto.

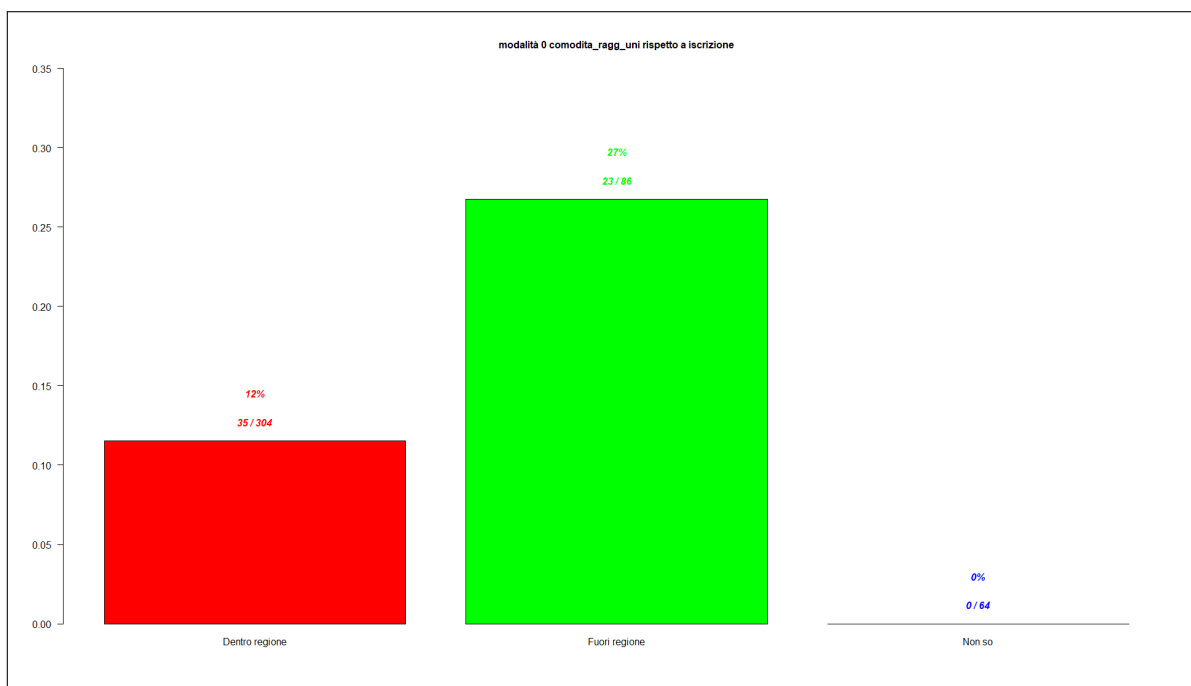


Figura 17: Modalità 0 comodità raggiungere università rispetto a iscrizione

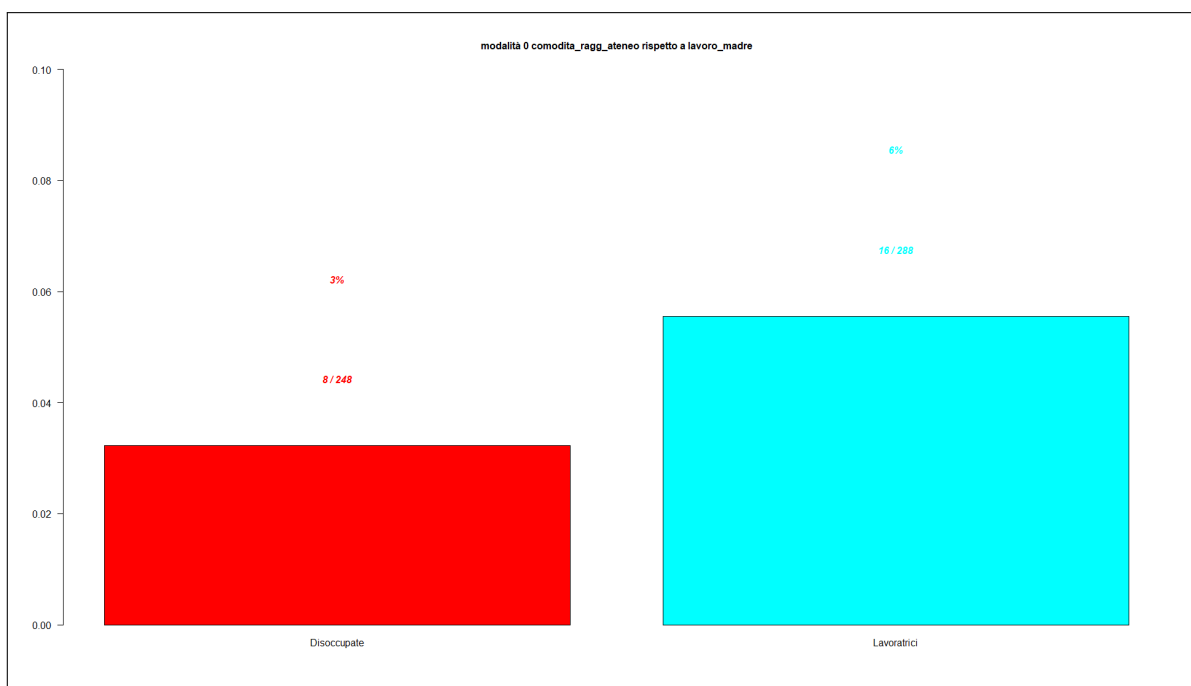


Figura 18: Modalità 0 comodità raggiungere ateneo rispetto a lavoro madre

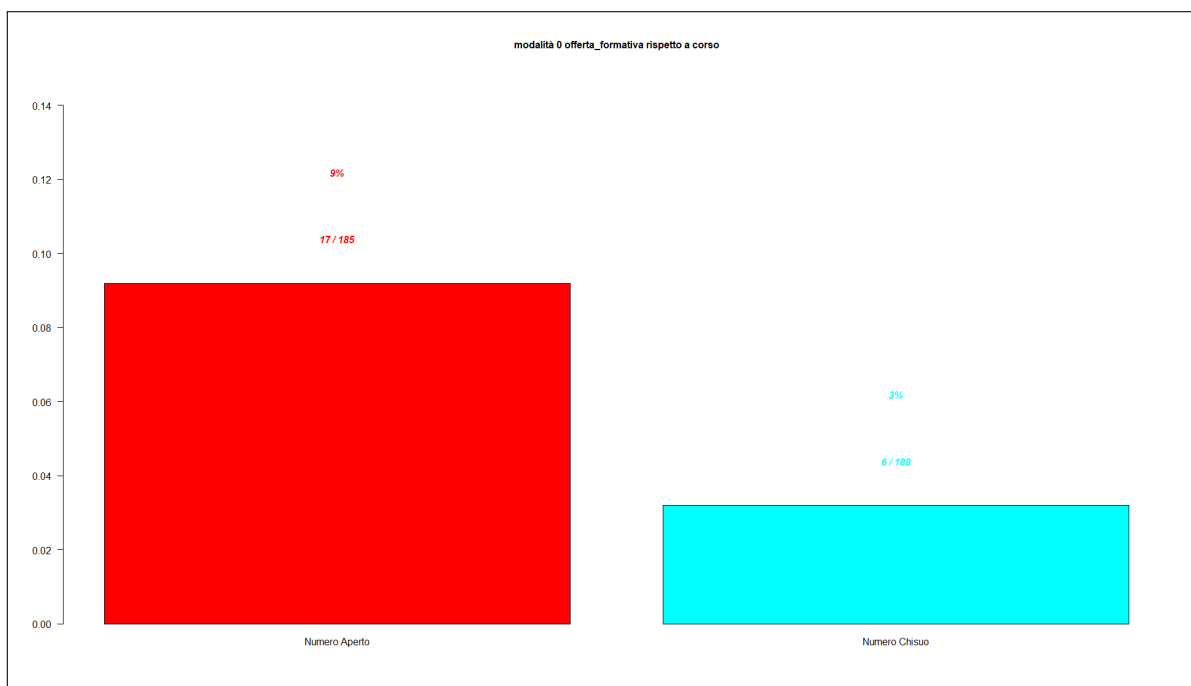


Figura 19: Modalità 0 offerta formativa rispetto a corso

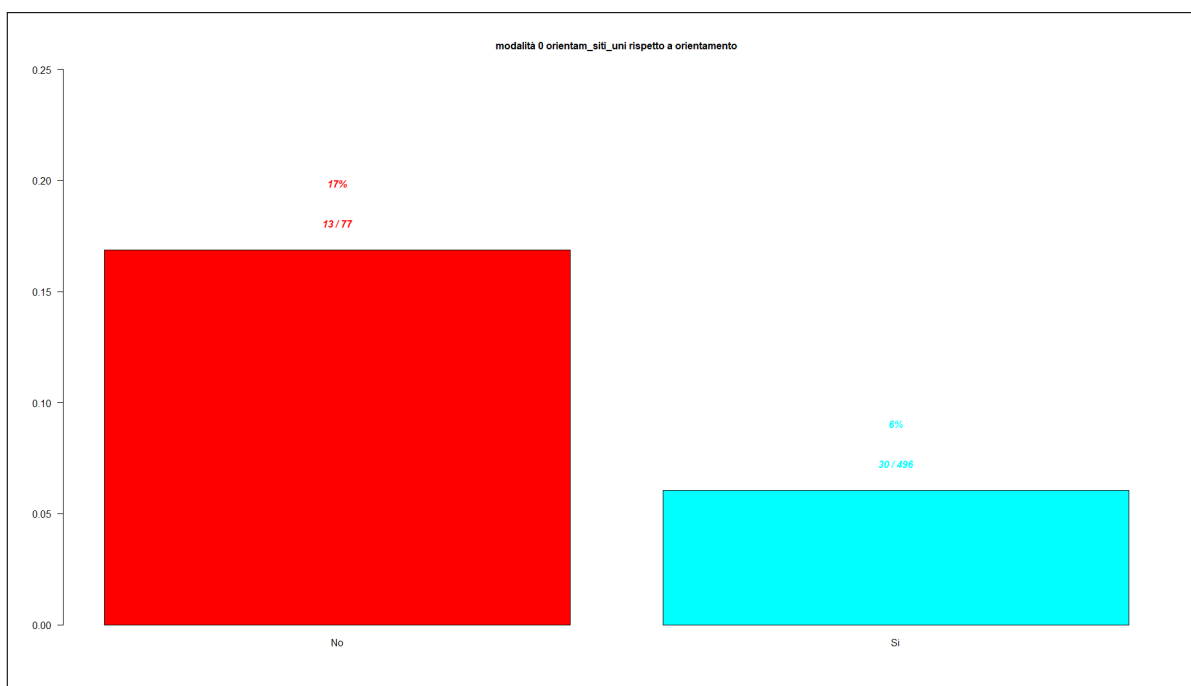


Figura 20: Modalità 0 orientamento siti uni rispetto a orientamento

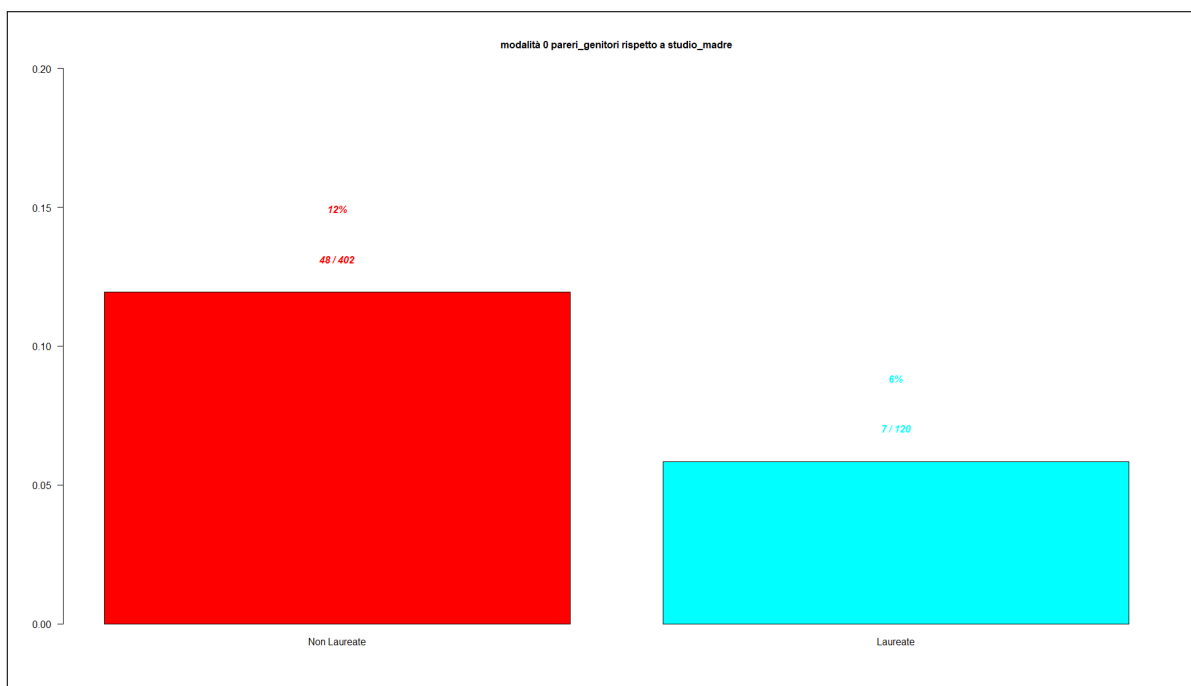


Figura 21: Modalità 0 pareri genitori rispetto a studio madre

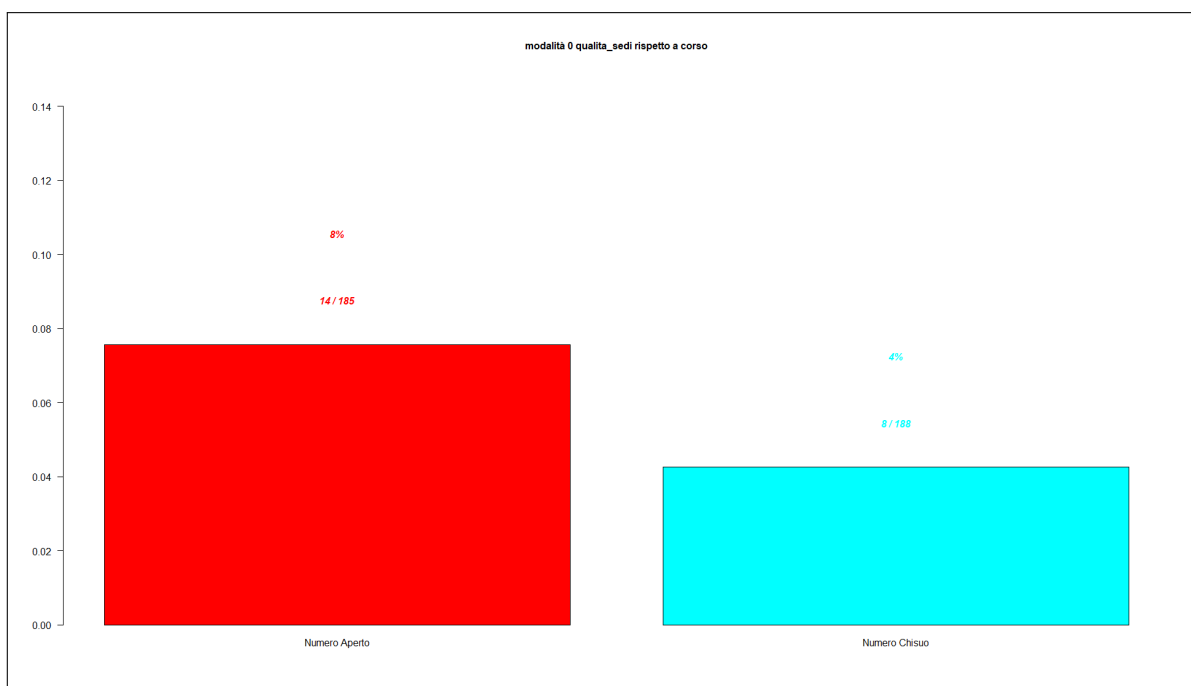


Figura 22: Modalità 0 qualita sedi rispetto a corso

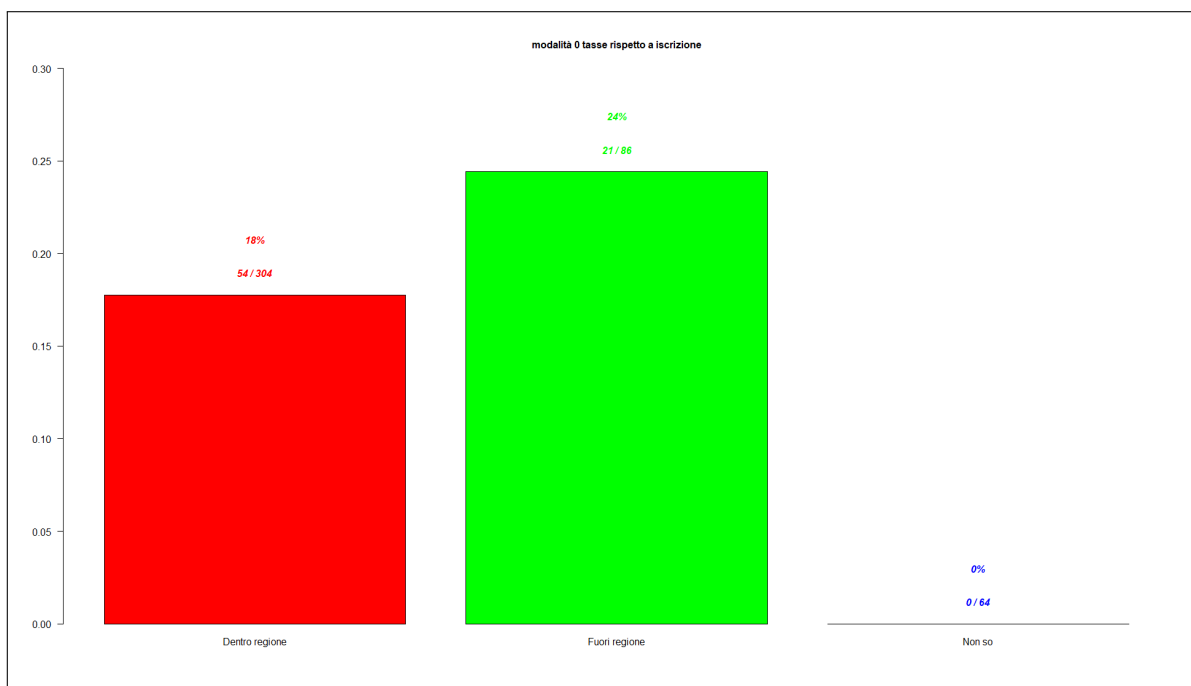


Figura 23: Modalità 0 tasse rispetto a iscrizione

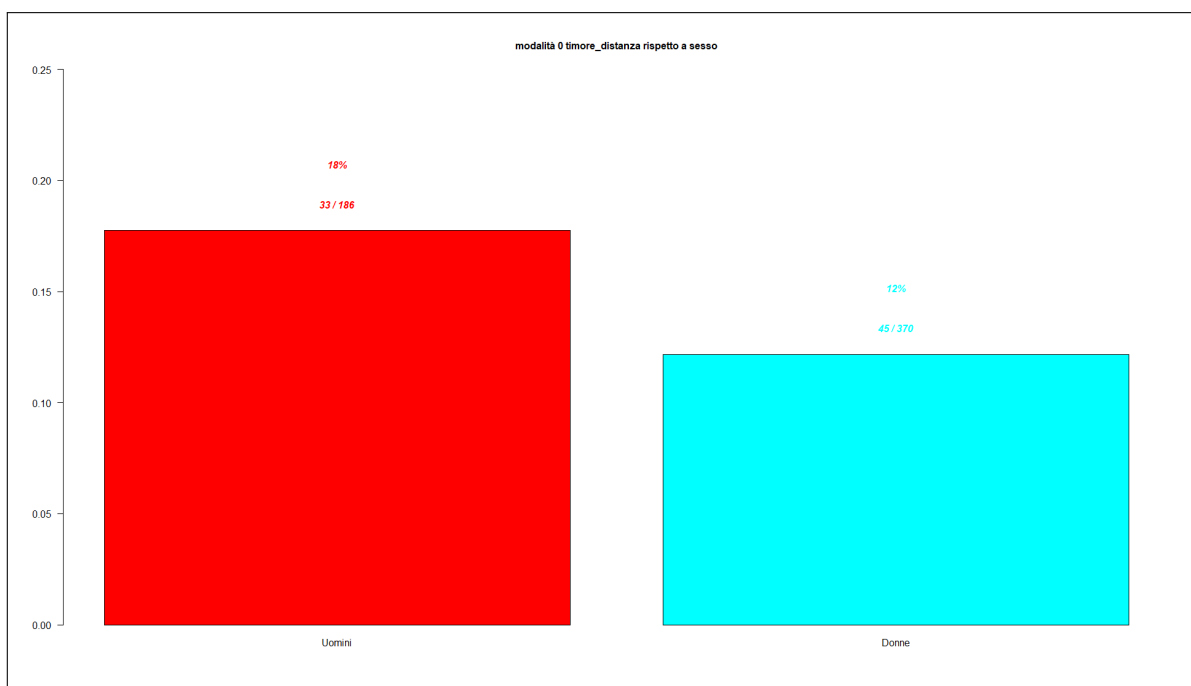


Figura 24: Modalità 0 timore distanza rispetto a sesso

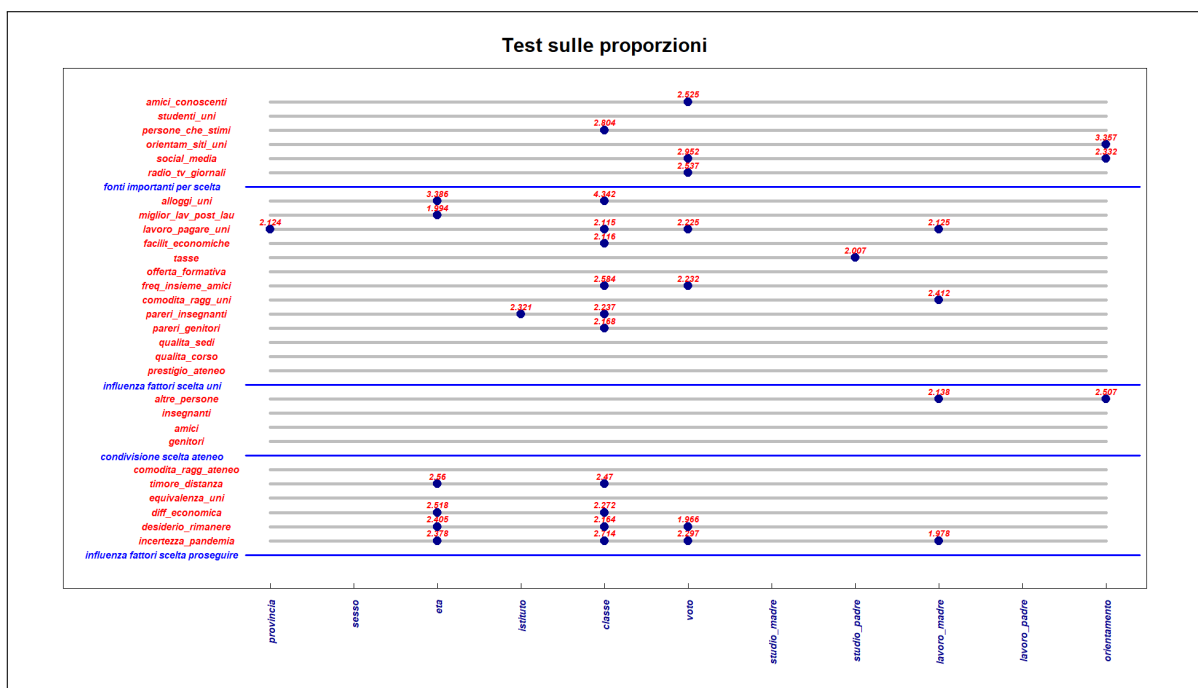


Figura 25: Griglia test sulle proporzioni

Il test sulle proporzioni è un test statistico per verificare se la differenza tra le frequenze relative delle modalità di una variabile dicotomica è significativa o meno. I test che risultano più spesso significativi sono quelli relativi alle variabili classe e voto.

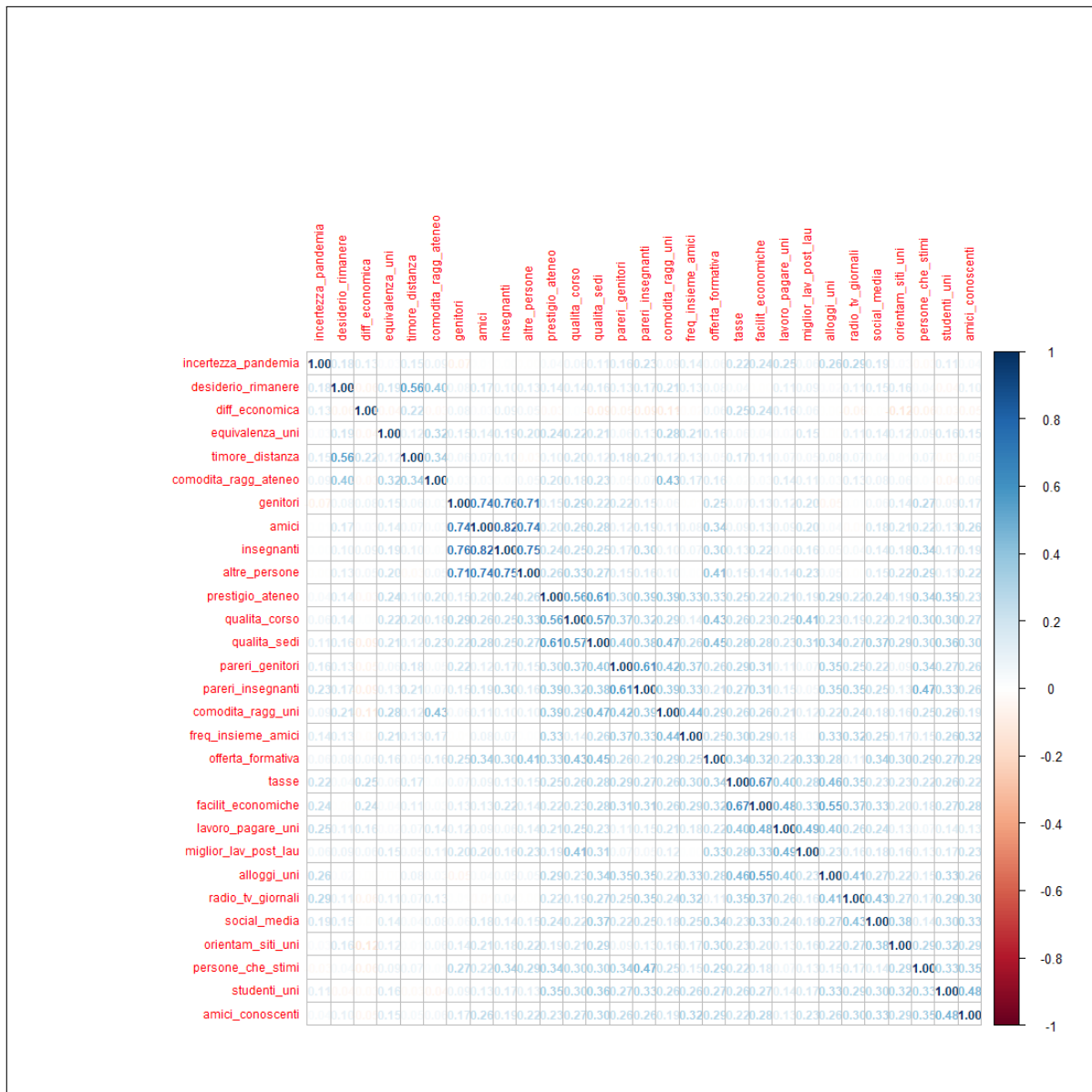


Figura 26: Matrice di correlazione

Grazie all'ausilio del package *Corrplot*, abbiamo ricavato la matrice di correlazione. Dalla stessa, che riporta al suo interno i coefficienti di correlazione tra le variabili, si evince che, per quanto riguarda la scelta dell'università a cui iscriversi, i pareri delle figure di riferimento per l'alunno (famiglia, amici, insegnanti) sono tutti fortemente correlati positivamente tra di loro. Ad esempio, un coefficiente di correlazione tra parere di amici ed insegnanti pari a 0.82 rivela che se lo studente reputa importante il parere degli amici, riterrà importante anche il parere degli insegnanti, così come quello dei genitori (0.74) e quello di altre persone di riferimento (0.71).

Correlazione positiva forte (0.67) anche tra l'ammontare delle tasse universitarie e le facilitazioni economiche, quali borse di studio, alloggi per studenti, etc. Ciò significa che se lo studente ha un impatto l'ammontare delle tasse, avrà un impatto anche il ricevere o meno una borsa di studio.

Medesima situazione (0.61) tra il prestigio dell'ateneo e la qualità delle sedi e delle strutture scolastiche, evidenziando il fatto che se lo studente dovesse dare peso alla qualità delle sedi, darà altrettanto peso al prestigio dell'ateneo, denotando così una forte correlazione tra i due fattori.

4 Analisi in Componenti Principali

L'Analisi in Componenti Principali (**ACP**) è uno strumento di **Data Reduction** ed ha lo scopo di sintetizzare il più possibile i dati con la minima perdita di informazione. Tale obiettivo è perseguito sostituendo alle p variabili originarie rilevate un numero di combinazioni lineari. Le variabili ottenute (definite Variabili Latenti) devono essere ricercate in modo che riescano a sintetizzare quanta più inerzia possibile rispetto alle p variabili di partenza. Quindi, lo scopo dell'ACP è quello di sintetizzare e descrivere in maniera ottimale il dataset con criteri di ottimalità, ovvero massima variabilità, mediante le cd Componenti Principali.

Criteri per la scelta di Componenti Principali da ritenere:

- 1) **Tasso di Inerzia**, in virtù del quale teoricamente ci fermiamo quando tale tasso è circa uguale al 75%. Si ritiene che si debba considerare il 75% della variabilità di un fenomeno poichè tale quota corrisponde ad un reale contenuto informativo, mentre il restante 25% non ha contenuto informativo reale, è rumore.
- 2) **Eigen-Value One**: ogni p variabile originaria (essendo stata precedentemente standardizzata) ha varianza pari ad 1. Le Componenti principali essendo combinazioni lineari delle p variabili originarie dovrebbero avere una varianza maggiore di 1. Se ciò non dovesse accadere significa che le Componenti Principali non sono in grado di catturare variabilità, non converrebbe quindi considerare CP associate ad un autovalore minore di 1, ma dovremmo prendere in considerazioni le variabili originarie.
- 3) **Scree-Plot**, criterio in base al quale si plottano sul piano cartesiano gli autovalori in ordine decrescente e si va ad individuare il cosiddetto punto di gomito (*elbow point*). Solitamente, le Componenti Principali più importanti (che spiegano la maggiore porzione di inerzia) sono quelle che precedono il punto di gomito.

Per quanto riguarda l'analisi effettuata (condotta grazie all'ausilio di *FactoMineR*) i fattori che influenzano sulla scelta di proseguire gli studi nella propria regione sono incertezza dovuta alla situazione pandemica, desiderio di rimanere nel proprio ambiente, mancanza di disponibilità economica, convinzione che le università campane siano equivalenti alle altre, timore di allontanarsi dalla propria casa ed infine la comodità di raggiungere l'ateneo.

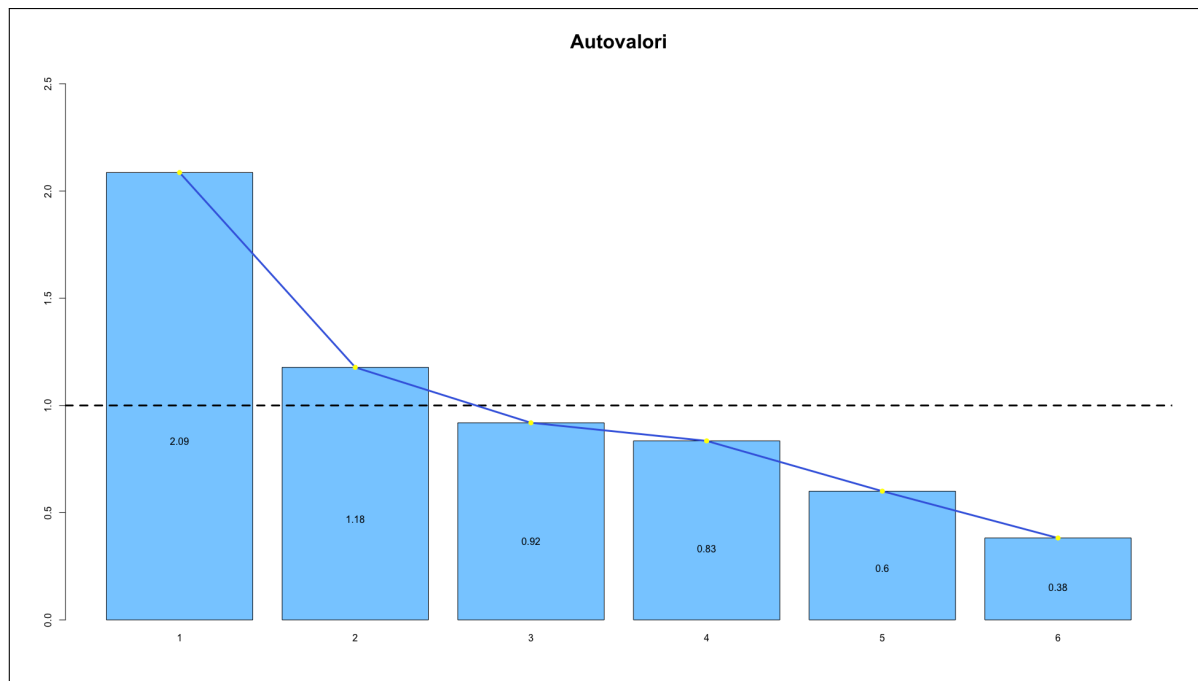


Figura 27: Plot Autovalori sulla scelta di proseguire gli studi nella propria regione

Secondo il criterio dell'Eigen-Value One andiamo a considerare le prime 2 Componenti Principali, che riportano un autovalore maggiore di 1. Non consideriamo la terza CP in quanto presenta autovalore

minore di 1 (0.92).

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	2.0861994	34.769990	34.76999
comp 2	1.1776840	19.628066	54.39806
comp 3	0.9191989	15.319981	69.71804
comp 4	0.8347786	13.912977	83.63101
comp 5	0.6002333	10.003888	93.63490
comp 6	0.3819059	6.365098	100.00000

Figura 28: Output autovalori e percentuale di varianza associata al corrispondente autovalore

Le prime due Componenti Principali spiegano circa il 55% dell'inerzia totale (si noti che la prima CP spiega la maggiore porzione di inerzia rispetto alle altre componenti principali). Anche se non raggiungiamo la cifra teorica del 75%, possiamo comunque ritenerci soddisfatti in quanto si è raggiunta una porzione di inerzia rilevante con solo 2 CP.

Variables											
	Dim.1	ctr	cos2	Dim.2	ctr	cos2	Dim.3	ctr	cos2		
incertezza_pandemia	0.318	4.849	0.101	0.608	31.425	0.370	-0.128	1.794	0.016		
desiderio_rimanere	0.809	31.394	0.655	-0.070	0.418	0.005	-0.329	11.799	0.108		
diff_economica	0.111	0.589	0.012	0.741	46.661	0.550	0.539	31.617	0.291		
equivalenza_uni	0.452	9.811	0.205	-0.414	14.574	0.172	0.660	47.454	0.436		
timore_distanza	0.774	28.744	0.600	0.171	2.483	0.029	-0.209	4.746	0.044		
comodita_ragg_ateneo	0.717	24.613	0.513	-0.229	4.439	0.052	0.154	2.590	0.024		

Figura 29: Output del contributo delle variabili sulle Componenti Principali

La prima CP (che rappresenta l'impatto sulla scelta di proseguire gli studi nella propria regione) è spiegata principalmente dal desiderio di rimanere nel proprio ambiente (0.809), dal timore della distanza (0.774) e comodità di raggiungere l'ateneo (0.717). Di fatti, il loro contributo sulla relativa componente principale è particolarmente rilevante: rispettivamente circa il 31%, il 29% ed il 25%. Possiamo anche considerare i cos2, che rappresentano il contributo relativo della variabile j-esima sull'asse, e che dal punto di vista algebrico sono uguali alle coordinate elevate al quadrato. La seconda CP riflette in che modo l'ambito socio-economico e lavorativo impatta su quelli che sono i fattori sulla scelta di continuare gli studi nella propria regione.

Dall'output in figura 34 si può infatti evincere che l'impatto maggiore (in termini di contributi) è riferito alla mancanza di disponibilità economica (per circa il 50%) e all'incertezza dovuta alla pandemia (per più del 30%).

Ciò si può dedurre anche dal seguente grafico:

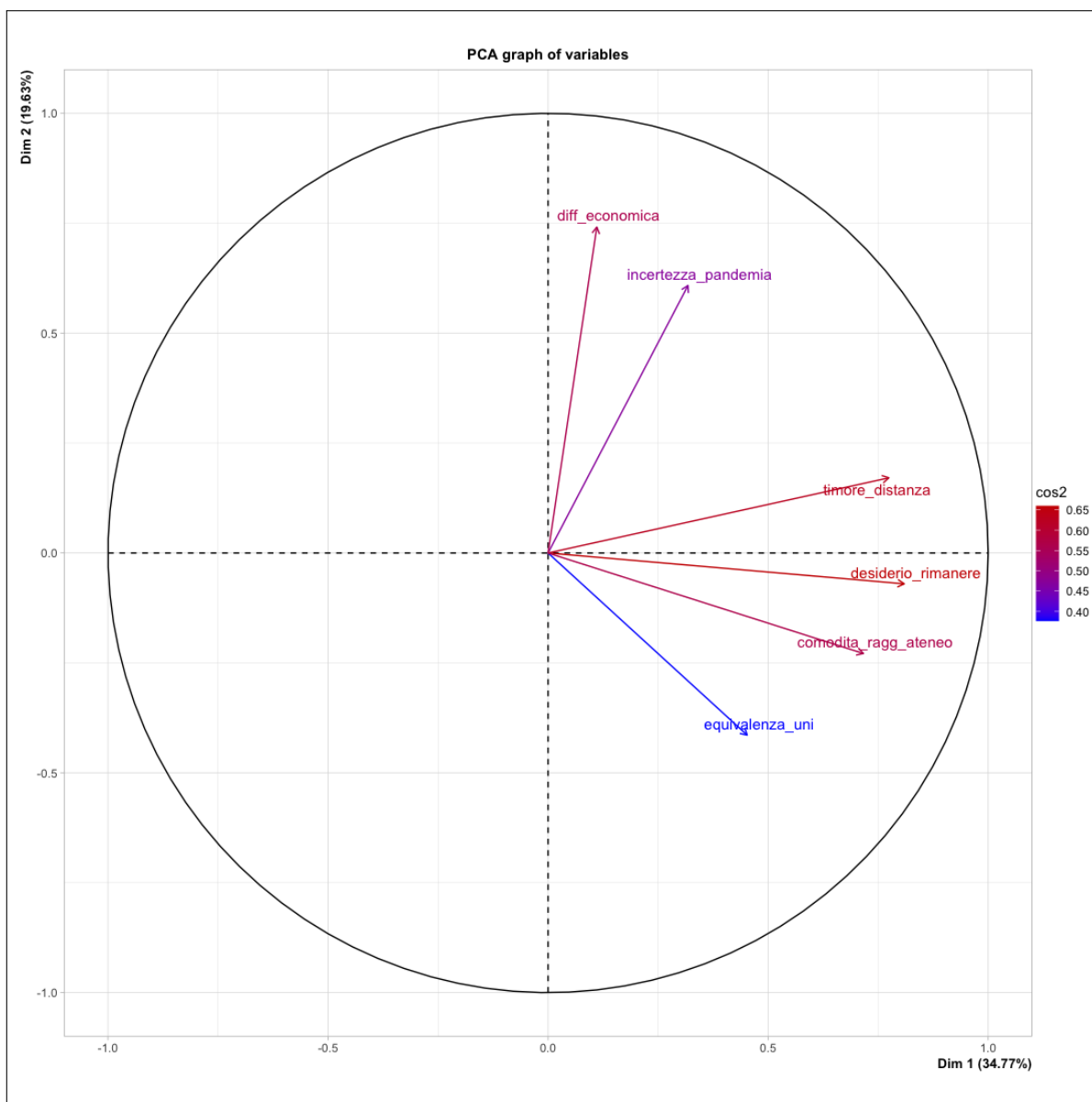


Figura 30: Grafico ACP delle variabili

Come individuato già in precedenza, la Dim 1 (ovvero la prima CP) spiega circa il 35% dell'inerzia totale, mentre la Dim 2 (ovvero la seconda CP) spiega circa il 20% dell'inerzia totale. Si arriva alla medesima conclusione (qui abbiamo considerato come riferimento i \cos^2) proiettando i vettori sulle due componenti principali, dove i vettori proiettati in blu rappresentano le variabili che hanno impattato di meno mentre quelli proiettati in rosso descrivono le cause che hanno avuto un impatto significativo.

Invece, per quanto riguarda l'analisi condotta, i fattori che influiscono sulla scelta dell'università a cui un individuo intende iscriversi sono il prestigio dell'ateneo, la qualità del corso di laurea scelto, la qualità delle sedi e delle strutture scolastiche, il parere dei propri genitori e degli insegnanti, la comodità di raggiungere l'università, la possibilità di frequentare corsi insieme ai propri amici, l'offerta formativa dei corsi disponibili, l'ammontare delle tasse universitarie, le facilitazioni economiche (borse di studio, alloggi per studenti, etc), la possibilità di trovare un lavoro per mantenersi gli studi, un contesto più favorevole dal punto di vista lavorativo post laurea e la presenza di alloggi università disponibili.

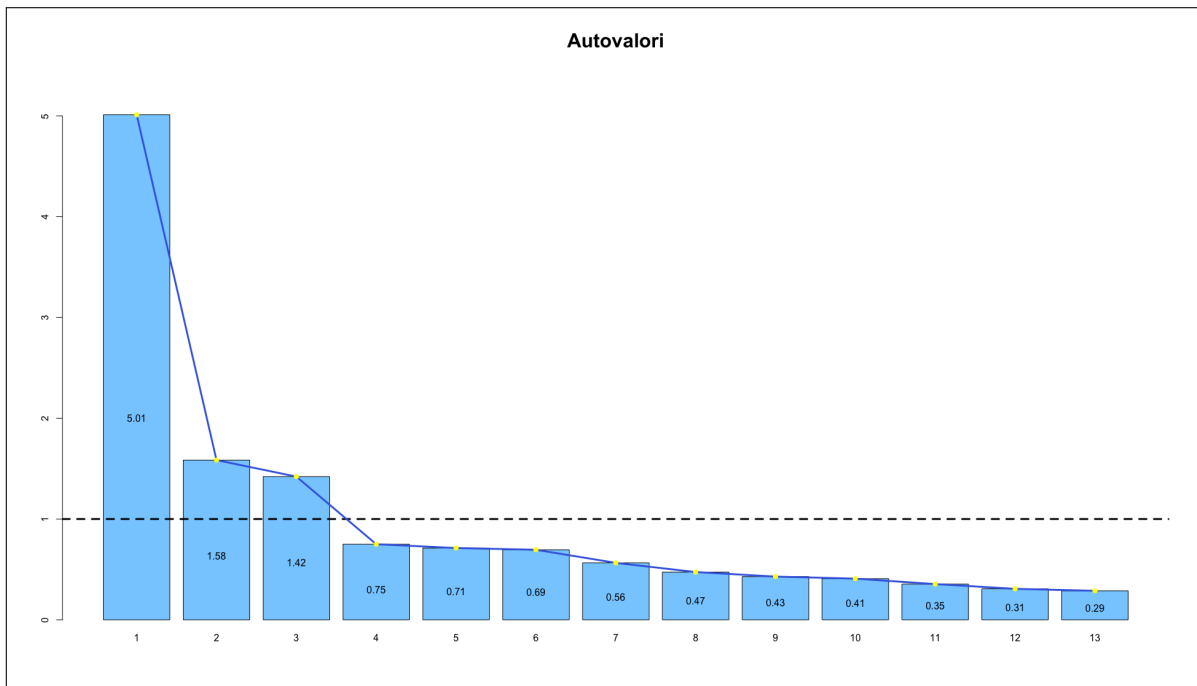


Figura 31: Plot Autovalori sulla scelta dell'università a cui un individuo intende iscriversi

In questo caso (secondo il criterio dell'eigen-value one) andiamo a considerare 3 Componenti Principali, che riportano un autovalore maggiore di 1. Per questa analisi è evidente il cosiddetto salto tra il primo autovalore (5.01) e il secondo autovalore (1.58).

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	5.0115817	38.550629	38.55063
comp 2	1.5848380	12.191061	50.74169
comp 3	1.4209013	10.930010	61.67170
comp 4	0.7500939	5.769953	67.44165
comp 5	0.7121801	5.478309	72.91996
comp 6	0.6949340	5.345646	78.26561
comp 7	0.5649828	4.346021	82.61163
comp 8	0.4736407	3.643390	86.25502
comp 9	0.4289152	3.299347	89.55437
comp 10	0.4084416	3.141858	92.69623
comp 11	0.3544582	2.726601	95.42283
comp 12	0.3074129	2.364715	97.78754
comp 13	0.2876196	2.212458	100.00000

Figura 32: Output autovalori e percentuale di varianza associata al corrispondente autovalore

Le prime tre Componenti Principali spiegano circa il 62% dell'inerzia totale (la prima CP, come si evince anche dal salto dal primo al secondo autovalore, spiega la maggiore porzione di inerzia rispetto alle altre componenti principali). Anche qui, pur non raggiungendo il 75% di inerzia spiegata, consideriamo le sole 3 CP con autovalore maggiore di 1.

Variables (the 10 first)									
	Dim.1	ctr	cos2	Dim.2	ctr	cos2	Dim.3	ctr	cos2
prestigio_ateneo	0.622	7.732	0.387	0.343	7.440	0.118	-0.354	8.816	0.125
qualita_corso	0.635	8.047	0.403	0.512	16.563	0.263	-0.225	3.561	0.051
qualita_sedi	0.705	9.914	0.497	0.396	9.910	0.157	-0.262	4.829	0.069
pareri_genitori	0.653	8.515	0.427	-0.266	4.455	0.071	-0.342	8.231	0.117
pareri_insegnanti	0.611	7.441	0.373	-0.376	8.910	0.141	-0.331	7.729	0.110
comodita_ragg_uni	0.593	7.011	0.351	-0.311	6.090	0.097	-0.297	6.217	0.088
freq_insieme_amici	0.525	5.494	0.275	-0.515	16.704	0.265	-0.115	0.932	0.013
offerta_formativa	0.559	6.228	0.312	0.375	8.892	0.141	-0.031	0.069	0.001
tasse	0.665	8.833	0.443	-0.319	6.411	0.102	0.306	6.583	0.094
facilit_economiche	0.708	9.993	0.501	-0.159	1.594	0.025	0.433	13.208	0.188

Figura 33: Output del contributo delle prime 10 variabili sulle Componenti Principali

La prima CP (che rappresenta l'impatto sulla scelta dell'università a cui un individuo intende iscriversi) è spiegata principalmente (in modo pressochè uguale) da tutti i fattori, rilevando un contributo che varia dal 7% al 9% circa per i fattori considerati. Arriviamo alla medesima conclusione con i cos2. La seconda CP riflette ciò che spinge lo studente ad iscriversi ad un'università piuttosto che un'altra. Su questa componente principale impatta particolarmente la qualità del corso di laurea e la possibilità di frequentare corsi insieme ai propri amici (con un contributo pari al 16% circa per entrambe). Tale CP riguarda quindi la spinta motivazionale. La terza CP, in questo caso, riflette in che modo l'ambito socio-economico e lavorativo impatta su quelli che sono i fattori sulla scelta dell'università a cui un individuo intende iscriversi. L'impatto maggiore (in termini di contributi) è riferito alle facilitazioni economiche, quali borsa di studio, agevolazioni per studenti, etc per circa il 13%) e alla possibilità di trovare un lavoro per mantenersi gli studi.

Dall'output in figura 33 si può infatti evincere che:

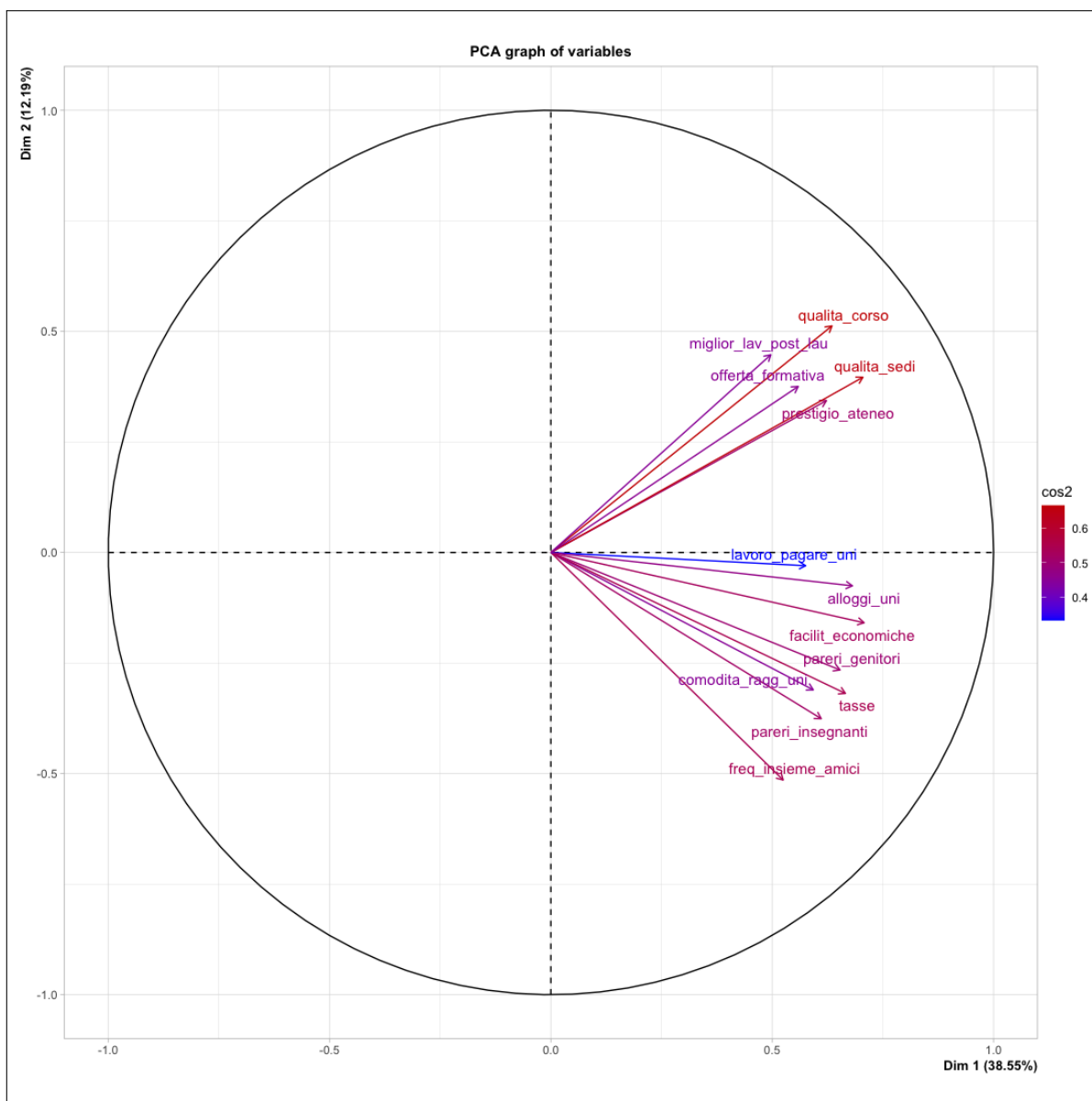


Figura 34: Grafico ACP delle variabili

Come individuato già in precedenza, la Dim 1 (ovvero la prima CP) spiega circa il 39% dell'inerzia totale, mentre la Dim 2 (ovvero la seconda CP) spiega circa il 12% dell'inerzia totale. Non è presente la terza componente principale sul grafico in quanto è possibile rappresentare solo due dimensioni sul piano cartesiano.

Infine, così come effettuato per l'analisi precedente, si arriva alla medesima conclusione (abbiamo considerato anche qui come riferimento i \cos^2) proiettando i vettori sulle due componenti principali, dove i vettori proiettati in blu rappresentano le variabili che hanno impattato di meno mentre quelli proiettati in rosso descrivono le cause che hanno avuto un impatto significativo.

5 Modelli

Dopo aver analizzato il dataset, abbiamo deciso di costruire dei modelli per analizzare come variano, al variare delle covariate di interesse, le risposte ad alcune variabili ordinali. Queste ultime riguardano l'impatto di certi fattori sulla scelta dell'università alla quale le unità statistiche hanno deciso di iscriversi.

5.1 Modelli di regressione logistica

Dall'analisi esplorativa sulle variabili ordinali, abbiamo evinto una sostanziale differenza tra la frequenza di coloro sui quali il fattore in esame non ha avuto alcun impatto e la frequenza di coloro sui quali, invece, il fattore ha avuto un certo impatto. Per questa ragione, abbiamo dicotomizzato l'ordinale, che rappresenta la nostra variabile di risposta, ponendo $Y_i = 1$ se il fattore non ha avuto alcun impatto e $Y_i = 0$ altrimenti. Volendo spiegare una risposta dicotomica Y_i in funzione di p esplicative X_1, X_2, \dots, X_p , abbiamo utilizzato un modello di regressione logistica, specificato come:

$$\text{logit}(\Pr(Y_i = 1)) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}, \quad i = 1, \dots, n, \quad (1)$$

dove con x_1, x_2, \dots, x_p indichiamo i valori delle p variabili esplicative rilevate sulle n unità statistiche.

Nei modelli costruiti, tutte le variabili esplicative sono variabili dummy, ovvero variabili indicatrici volte a segnalare la presenza o l'assenza di una certa modalità di un fattore. Per le covariate di natura dicotomica e per quelle rese tali è stata inserita nel modello un'unica variabile dummy. Per la provincia di residenza e l'area disciplinare scelta, sono state inserite rispettivamente 4 e 5 dummy.

Con il primo modello costruito vogliamo spiegare l'impatto delle tasse universitarie sulla scelta dell'università a cui iscriversi.

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -0.8897     0.2802  -3.175   0.0015 **
fuori regione   -0.7471     0.3343  -2.235   0.0254 *
areaIngegneria  0.6850     0.3795   1.805   0.0710 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 277.85  on 268  degrees of freedom
Residual deviance: 270.71  on 266  degrees of freedom
(371 observations deleted due to missingness)
AIC: 276.71
```

Figura 35: Impatto ammontare tasse universitarie sulla scelta dell'università.

Dall'output in figura 35 deduciamo che l'unica variabile significativa ($p - \text{value} < 0.05$) è fuori regione, con segno negativo del parametro. Questo significa che l'ammontare delle tasse universitarie è stato un fattore influente sulla scelta dell'università a cui iscriversi, soprattutto per coloro che hanno deciso di studiare in un ateneo campano.

Con il secondo modello costruito vogliamo spiegare l'impatto della comodità di raggiungere l'ateneo sulla scelta dell'università a cui iscriversi.


```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.7765     0.2519  -3.082 0.002056 **
fuori regione -1.1571     0.3101  -3.731 0.000191 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 314.32  on 349  degrees of freedom
Residual deviance: 301.16  on 348  degrees of freedom
(290 observations deleted due to missingness)
AIC: 305.16

```

Figura 36: Impatto ammontare tasse universitarie sulla scelta dell'università.

Dall'output in figura 36 deduciamo che l'unica variabile significativa è fuori regione, con segno negativo del parametro. Questo significa che la comodità nel raggiungere l'ateneo è stato un fattore molto influente sulla scelta dell'università a cui iscriversi, soprattutto per coloro che hanno deciso di restare a studiare in Campania.

Con il terzo modello costruito vogliamo spiegare l'impatto della comodità di raggiungere l'ateneo sulla scelta di proseguire gli studi nella propria regione.

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.8622     0.3635  -7.874 3.44e-15 ***
lavoro_madre  0.8899     0.4509   1.973  0.0484 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 163.63  on 278  degrees of freedom
Residual deviance: 159.49  on 277  degrees of freedom
(361 observations deleted due to missingness)
AIC: 163.49

```

Figura 37: Impatto comodità di raggiungere l'ateneo sulla scelta di studiare in Campania.

Dall'output in figura 37 deduciamo che l'unica variabile significativa è lavoromadre, con segno positivo del parametro. Questo significa che la comodità di raggiungere l'ateneo è stato un fattore influente sulla scelta di studiare in Campania, soprattutto per le unità la cui madre non lavora.

Con il quarto modello costruito vogliamo spiegare l'importanza di incontri di orientamento e siti delle università per la scelta dell'ateneo dove proseguire gli studi.

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.4796    0.3529  -1.359    0.174
orientamento -1.7540    0.4018  -4.365 1.27e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 259.22  on 343  degrees of freedom
Residual deviance: 242.35  on 342  degrees of freedom
(296 observations deleted due to missingness)
AIC: 246.35

```

Figura 38: Importanza servizio di orientamento per la scelta dell'ateneo a cui iscriversi.

Dall'output in figura 38 deduciamo che l'unica variabile significativa è orientamento, con segno negativo del parametro. Questo significa che gli incontri di orientamento e i siti delle università sono stati fattori importanti sulla scelta dell'ateneo dove proseguire gli studi, soprattutto per le unità le cui scuole hanno fornito un servizio di orientamento.

Con il quinto modello costruito vogliamo spiegare l'impatto delle offerte formative dei corsi sulla scelta dell'ateneo dove proseguire gli studi.

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.2037    0.2556  -8.623  <2e-16 ***
corsolaurea  -1.0921    0.4880  -2.238   0.0252 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 168.17  on 338  degrees of freedom
Residual deviance: 162.51  on 337  degrees of freedom
(301 observations deleted due to missingness)
AIC: 166.51

```

Figura 39: Impatto offerte formative sulla scelta dell'ateneo dove proseguire gli studi.

Dall'output in figura 39 deduciamo che l'unica variabile significativa è corsolaurea, con segno negativo del parametro. Questo significa che l'offerta formativa dei corsi è stata un fattore importante sulla scelta dell'ateneo dove proseguire gli studi, soprattutto per chi ha scelto corsi a numero chiuso.

Con il sesto modello costruito vogliamo spiegare l'impatto del parere dei genitori sulla scelta dell'ateneo dove proseguire gli studi.

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.3967      0.1612  -8.664  <2e-16 ***
studio_madre -0.8471      0.4289  -1.975   0.0483 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 291.76  on 314  degrees of freedom
Residual deviance: 287.21  on 313  degrees of freedom
(325 observations deleted due to missingness)
AIC: 291.21

```

Figura 40: Impatto parere genitori sulla scelta dell'ateneo dove proseguire gli studi.

Dall'output in figura 40 deduciamo che l'unica variabile significativa è studiomadre, con segno negativo del parametro. Questo significa che il parere dei genitori è stato un fattore influente sulla scelta dell'ateneo dove proseguire gli studi, soprattutto per le unità le cui madri hanno un titolo universitario.

Con il settimo modello costruito vogliamo spiegare l'impatto del timore di allontanarsi dalla propria casa sulla scelta di proseguire gli studi nella propria regione.

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.5978      0.2167  -2.759   0.00581 **
sesso        -0.6389      0.2750  -2.323   0.02017 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 339.56  on 292  degrees of freedom
Residual deviance: 334.24  on 291  degrees of freedom
(347 observations deleted due to missingness)
AIC: 338.24

```

Figura 41: Impatto timore di allontanarsi sulla scelta di proseguire gli studi nella propria regione.

Dall'output in figura 41 deduciamo che l'unica variabile significativa è sesso, con segno negativo del parametro. Questo significa che il timore di allontanarsi da casa è stato un fattore influente sulla scelta di proseguire gli studi nella propria regione, soprattutto per le donne.

Con il ottavo modello costruito vogliamo spiegare l'impatto della qualità delle sedi e delle strutture universitarie sulla scelta dell'ateneo dove proseguire gli studi.

Coefficients:					
	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.7813	0.3621	-7.682	1.57e-14	***
corsolaurea	-1.4419	0.7230	-1.995	0.0461	*
areaMedicoSanitaria	1.6057	0.7407	2.168	0.0302	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
(Dispersion parameter for binomial family taken to be 1)					
Null deviance: 104.431 on 271 degrees of freedom					
Residual deviance: 98.066 on 269 degrees of freedom					
(368 observations deleted due to missingness)					
AIC: 104.07					

Figura 42: Impatto qualità di sedi e strutture sulla scelta dell'ateneo dove proseguire gli studi.

Dall'output in figura 42 deduciamo che le variabili significative sono corsolaurea (parametro negativo) e areaMedicoSanitaria(parametro positivo). Questo significa che la qualità di sedi e strutture universitarie è stato un fattore influente sulla scelta dell'ateneo dove proseguire gli studi, soprattutto per chi ha scelto di iscriversi ad un corso a numero chiuso e per chi non ha scelto un corso in area medico sanitaria.