

Università degli Studi Roma Tre

Scuola di Economia e Studi Aziendali

Dipartimento di Economia



Corso di Laurea Triennale in Economia e Big Data

## ANALISI VALUTATIVA DEL CORSO DI ECONOMIA E BIG DATA

NOME COGNOME

Matteo Lancellotti (574016)

Flavio Mangione (575356)

Leonardo Ruzzante (574183)

Federico Vannozzi (575663)

Gabriella Vespa (576251)

# INDICE

<b>INTRODUZIONE.....</b>	<b>1</b>
<b>ANALISI ESPLORATIVA DEI DATI.....</b>	<b>1</b>
Indirizzi scolastici di provenienza .....	1
Valutazione degli esami superati.....	3
Confronto studenti lavoratori e non.....	4
Analisi per genere .....	6
<b>MODELLI DI REGRESSIONE.....</b>	<b>7</b>
Modello di regressione lineare multipla .....	7
Coefficienti .....	7
Equazione .....	7
Residui medie aritmetiche.....	7
Modello di regressione logistica.....	9
Coefficienti .....	9
Probabilità.....	9
Numero di iterazioni del metodo di Fisher scoring .....	9
<b>TECNICHE MULTIVARIATE .....</b>	<b>10</b>
Analisi fattoriale .....	10
Analisi dei cluster .....	11
<b>CONCLUSIONE.....</b>	<b>12</b>

## INTRODUZIONE

Il seguente studio si basa su un dataset che raccoglie informazioni riguardanti gli esami sostenuti dagli studenti, suddivisi per genere, nei primi due anni del corso di Economia e Big Data. Questi dati sono stati raccolti tramite un questionario online e fornisce, oltre alle valutazioni in trentesimi e numero di esami superati, informazioni inerenti età, indirizzo scolastico di provenienza, voto della maturità, numero di fratelli e sorelle, presenza o meno di studenti lavoratori e fuori sede, presenza di fumatori e infine il soddisfacimento degli studenti per il corso. Con l'analisi dei dati raccolti su 68 studenti e 42 studentesse frequentanti il corso di Economia e Big Data di Roma Tre nato nel 2021, intendiamo sfruttare i dati campionari disponibili, al fine di fornire un punto di riferimento didattico e valutativo per l'intero corso, utilizzando gli strumenti statistici a nostra disposizione. L'analisi descrittiva ed esplorativa e i modelli di regressione sono condotti sull'intero dataset, mentre l'analisi di clustering è applicata a un sottoinsieme di 37 studenti che hanno superato tutti e 9 gli esami, garantendo così un campione completo. Inoltre, è importante sottolineare che due studenti su 37 non hanno fornito i voti di tutti gli esami (l'osservazione 27 presenta due dati mancanti, mentre l'osservazione 47 presenta un dato mancante). Pertanto, abbiamo deciso di calcolare la media aritmetica al posto dei voti mancanti ai due studenti. Questa procedura si è resa necessaria al fine di ottenere un set completo di dati su cui lavorare. Occorre tuttavia considerare che l'analisi condotta su un campione ristretto di 37 studenti, che hanno superato tutti e 9 gli esami potrebbe comportare un bias selettivo. Infatti, questo gruppo potrebbe essere composto da studenti particolarmente motivati e capaci, i quali potrebbero non rappresentare l'intera popolazione degli studenti del corso di laurea in Economia e Big Data. Tale circostanza potrebbe avere ripercussioni sui risultati e influire sugli output dei modelli e sulle deduzioni dell'analisi, generando risposte fuorvianti e distorte.

## ANALISI ESPLORATIVA DEI DATI

### Indirizzi scolastici di provenienza

Iniziamo la nostra analisi esplorativa dei dati osservando le percentuali degli indirizzi scolastici di provenienza degli studenti del corso attraverso l'utilizzo di un grafico a torta.

Dal grafico a torta della Figura 1 possiamo chiaramente osservare che il 67.3% degli studenti proviene da un indirizzo scientifico seguito dal 15.5% di studenti provenienti da un indirizzo tecnico e infine l'8.2% di studenti provenienti da un indirizzo classico e il 9.1% di studenti provenienti da altri indirizzi. La leggenda in alto a destra del grafico fornisce una descrizione per ciascuna fetta del grafico, indicando il nome dell'indirizzo e il numero di studenti corrispondente per indirizzo. Considerando la somma delle percentuali degli studenti provenienti da indirizzi scientifici e tecnici, si può notare che l'82.8% di essi mostra un interesse per le discipline STEM (Scienze, Tecnologia, Ingegneria e Matematica). Questo fattore potrebbe aver influenzato la loro decisione di iscriversi a questo nuovo corso di laurea triennale.

Grafico a torta per gli indirizzi scolastici di provenienza degli studenti

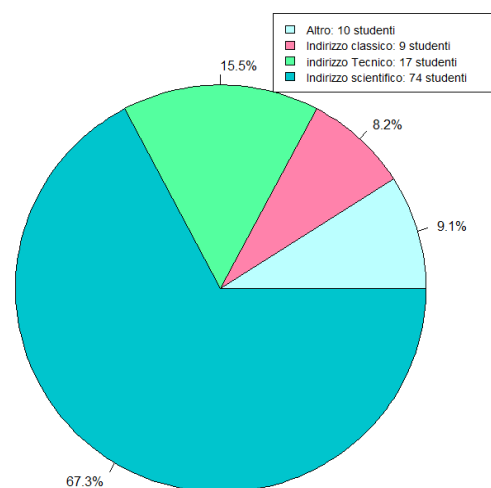


Figura 1 Grafico a torta percentuale indirizzi scolastici

Continuiamo la nostra analisi attraverso una rappresentazione grafica del voto di maturità degli studenti suddivisi per indirizzi scolastici di provenienza attraverso dei box plot.

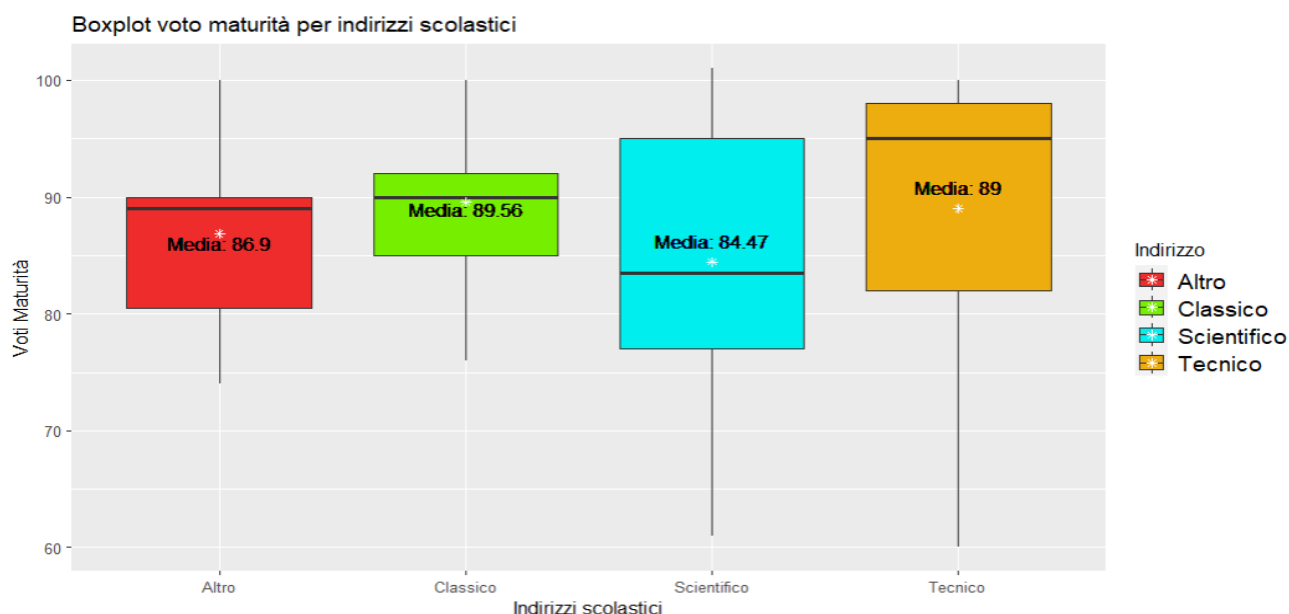


Figura 2 Box plot che descrive il voto di maturità per gli indirizzi scolastici di provenienza degli studenti

	ALTRO	CLASSICO	SCIENTIFICO	TECNICO
<b>Numero di osservazioni</b>	10	9	74	17
<b>Media</b>	86.90	89.56	84.47	89
<b>Deviazione standard</b>	9.15	7.81	10.82	11.15
<b>Minimo</b>	74	76	61	60
<b>Massimo</b>	100	100	101	100
<b>Primo quartile</b>	80	85	77	82
<b>Mediana</b>	89	90	83.50	95
<b>Terzo quartile</b>	90	92	95	98
<b>Distanza interquartile</b>	10	7	18	16
<b>Asimmetrie</b>	Positiva	Leggera asimmetria positiva	Negativa	Positiva

Tabella 1 Analisi per indirizzi scolastici dal box plot in Figura 2

In nessun box plot della Figura 2 vi è la presenza di valori anomali. Le medie dei voti di maturità fra i quattro gruppi non si discostano troppo tra di loro con un range tra 80 e 90. Tutti e quattro gli indirizzi hanno una deviazione standard molto alta soprattutto gli indirizzi scientifici e tecnici, rispettivamente 10.82 e 11.15, con un intervallo di valori che va da 61 a 101 per l'indirizzo scientifico e da 60 a 100 per l'indirizzo tecnico. Gli studenti provenienti dall'indirizzo classico hanno invece un box plot con valori della media dei voti che intervallano da 76 a 100. La deviazione standard in questo caso è minore di 7.81 e con una leggera asimmetria positiva poiché  $(Q3-Q2) < (Q2-Q1)$ . Probabilmente ciò è dovuto alle poche osservazioni appartenenti a questo gruppo di dati.

Da questa prima analisi emergerebbe che gli studenti provenienti dall'indirizzo classico siano quelli con una media dei voti più alta ma bisogna anche considerare che gli studenti provenienti dall'indirizzo scientifico hanno un numero di osservazioni nettamente maggiore che potrebbe aver influenzato la deviazione standard e la media di quel gruppo di dati. Il voto massimo 100 e lode (101) è stato ottenuto da una studentessa proveniente dall'indirizzo scientifico. Il voto 100 è stato ottenuto da 15 studenti suddivisi equamente tra il genere maschile e femminile. Il voto minimo di 60 è stato ottenuto da uno studente proveniente dall'indirizzo tecnico. Per l'indirizzo scientifico la media del voto di maturità per le 25 studentesse è di 89.16 rispetto alla media dei 49 studenti di 82.8. Le studentesse risultano quindi avere una media dei voti più alta degli studenti.

## Valutazione degli esami superati

Possiamo ora analizzare il numero di studenti che hanno passato ogni singolo esame attraverso un grafico a barre.

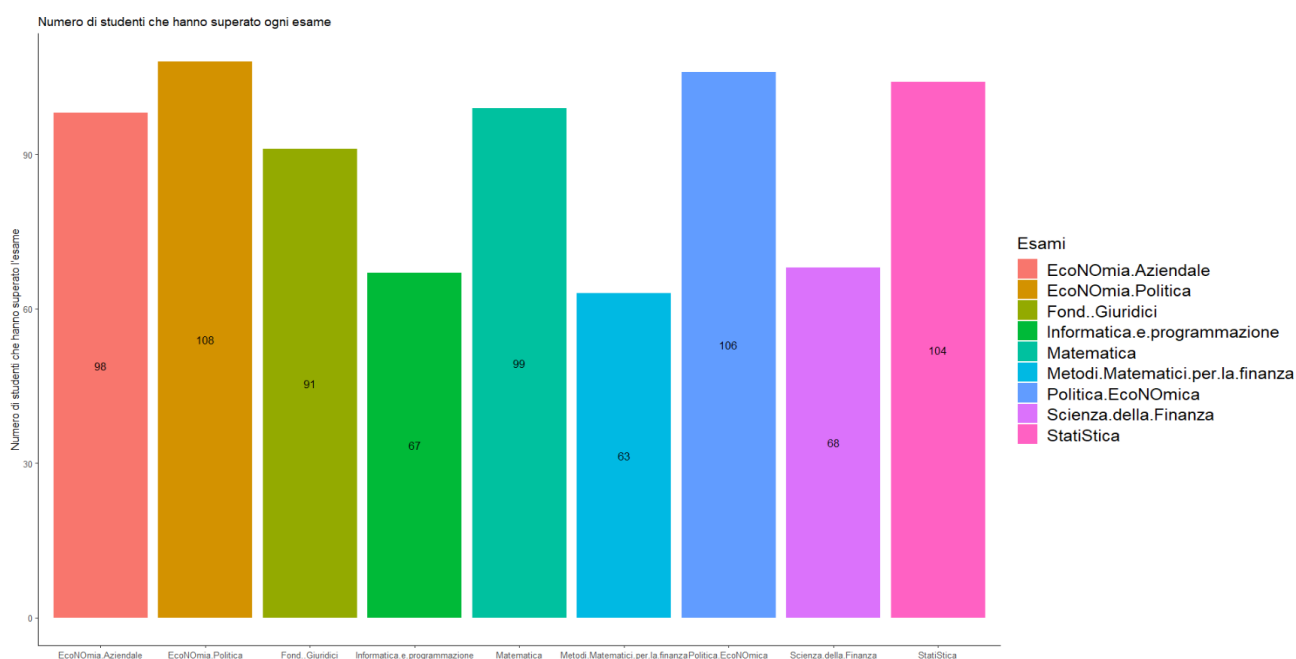
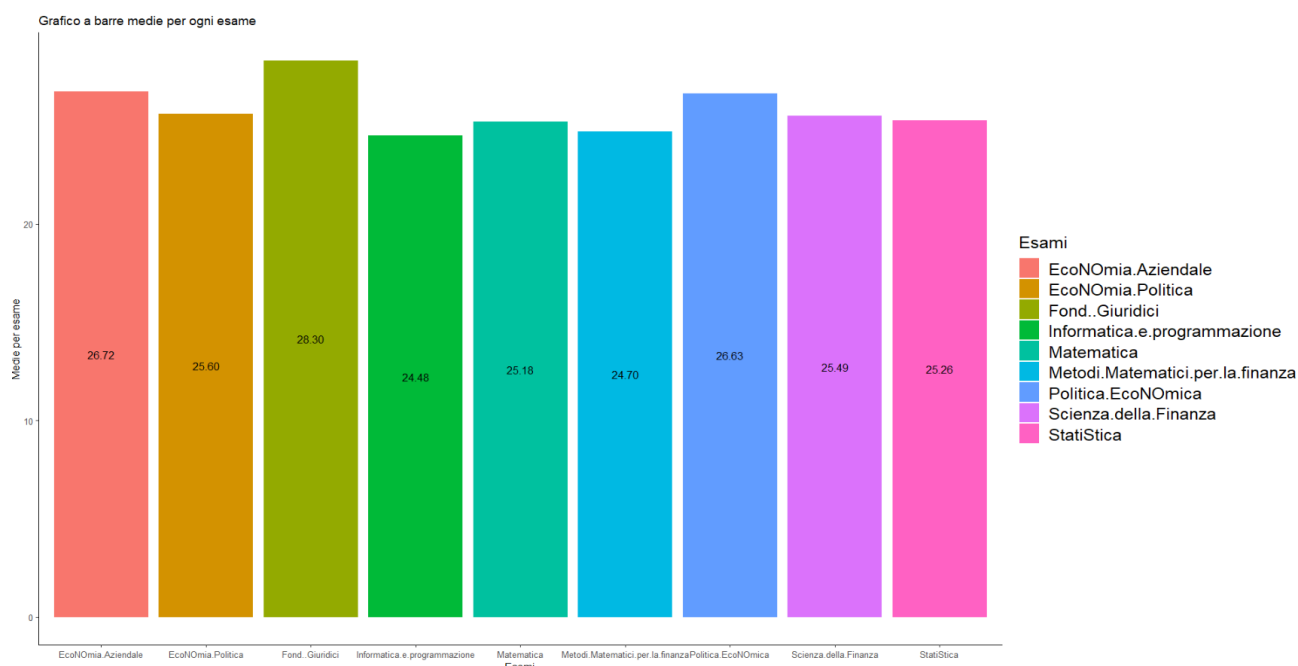


Figura 3 Grafico a barre superamento esami

Alcuni studenti non hanno inserito correttamente tutti i voti degli esami superati e quindi i risultati del grafico a barre in Figura 3 potrebbero risultare poco accurati. Gli esami di Statistica, Politica Economica, Economia Politica, e Matematica sono quelli che sono stati passati dal numero più alto di studenti. Per gli esami del primo anno del corso di Economia e Big Data (Matematica generale, Economia Politica, Economia Aziendale, Statistica, Fondamenti giuridici della digitalizzazione, Informatica e Programmazione) c'è un elevato numero di studenti che li hanno superati ad eccezione di Informatica e Programmazione. Il basso numero di studenti che ha conseguito con successo il seguente esame potrebbe farci ipotizzare che gli studenti hanno riscontrato maggiore difficoltà rispetto ad altri esami. Nel primo semestre del secondo anno il maggior numero di studenti che hanno superato un esame si riscontrano per Politica Economica con 106 mentre Metodi Matematici per la Finanza e Scienze delle Finanze presentano un numero minore però comunque maggiori del 50%. Possiamo quindi evincere che questi due esami sono quelli in cui gli studenti hanno probabilmente riscontrato maggiori difficoltà, specialmente Metodi Matematici per la Finanza.

Dopo aver osservato il numero di studenti che hanno superato ogni esame, possiamo analizzare le medie aritmetiche attraverso un grafico a barre.

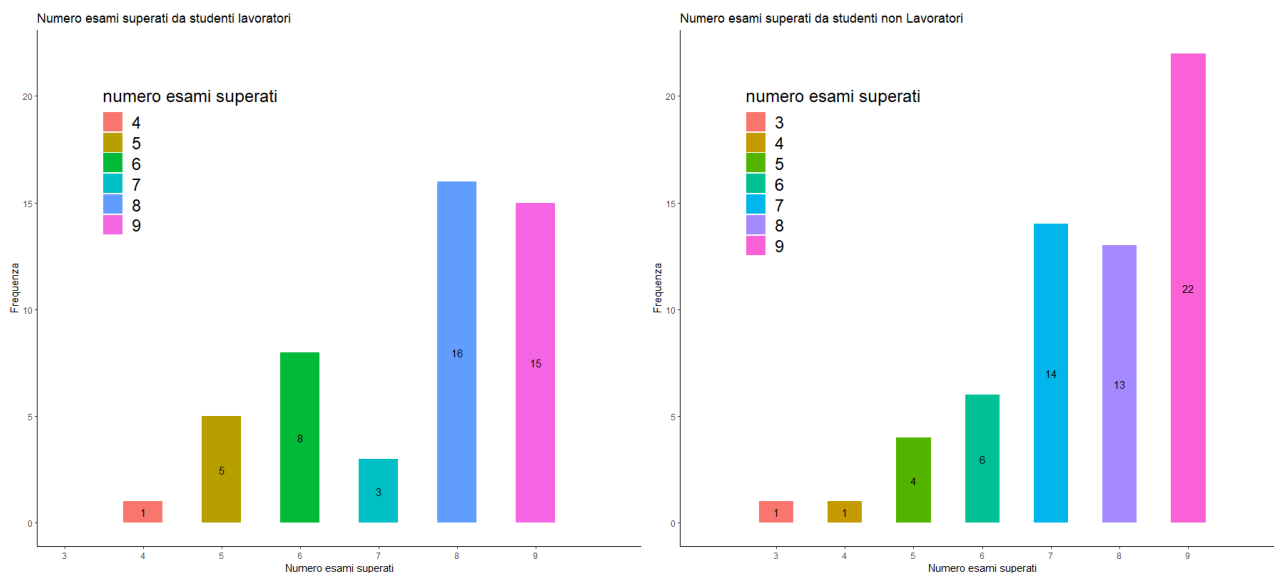


*Figura 4 Grafico a barre che descrivere le medie aritmetiche per ogni esame*

Nella Figura 4 viene rappresentata su ogni barra del grafico il valore della media aritmetica associata a ogni esame. Si può osservare che l'esame con il valore della media aritmetica più elevato è Fondamenti Giuridici del primo anno con 28.30 mentre l'esame con il valore della media aritmetica più basso è Informatica e Programmazione con 24.48. Dai valori delle medie aritmetiche per gli esami del primo semestre del secondo anno si evidenzia un valore elevato per l'esame di Politica Economica (26.63) e un bassissimo valore della media aritmetica per l'esame di Metodi Matematici per la Finanza di 24.70. L'analisi appena fatta potrebbe quindi confermare l'ipotesi evidenziata nel paragrafo precedente. Per il corso di Scienze della Finanze la media aritmetica dei voti è di 25.49 con 68 studenti che lo hanno superato quindi gli studenti potrebbero non aver riscontrato difficoltà nell'esame diversamente da quanto riscontrato nel paragrafo precedente. Osservando la media aritmetica dei risultati dell'esame di Statistica del primo anno possiamo usare un intervallo di confidenza al 95% per stimare dove dovrebbe cadere la media aritmetica dei risultati del primo esonero dell'esame del secondo anno di Statistica per Big Data. Secondo l'intervallo di confidenza la media aritmetica calcolata per gli studenti che supereranno l'esame di Statistica per Big Data con una valutazione di almeno 18 rientra in un intervallo da 24.45 a 26.06 ed effettivamente il valore di 25.14 (calcolato sulla base delle valutazioni pubblicate recentemente) rientra in questo range. Questo risultato è positivo in quanto l'esonero rientra in un semestre decisamente più complesso rispetto all'esame del primo anno per la presenza di 4 esami.

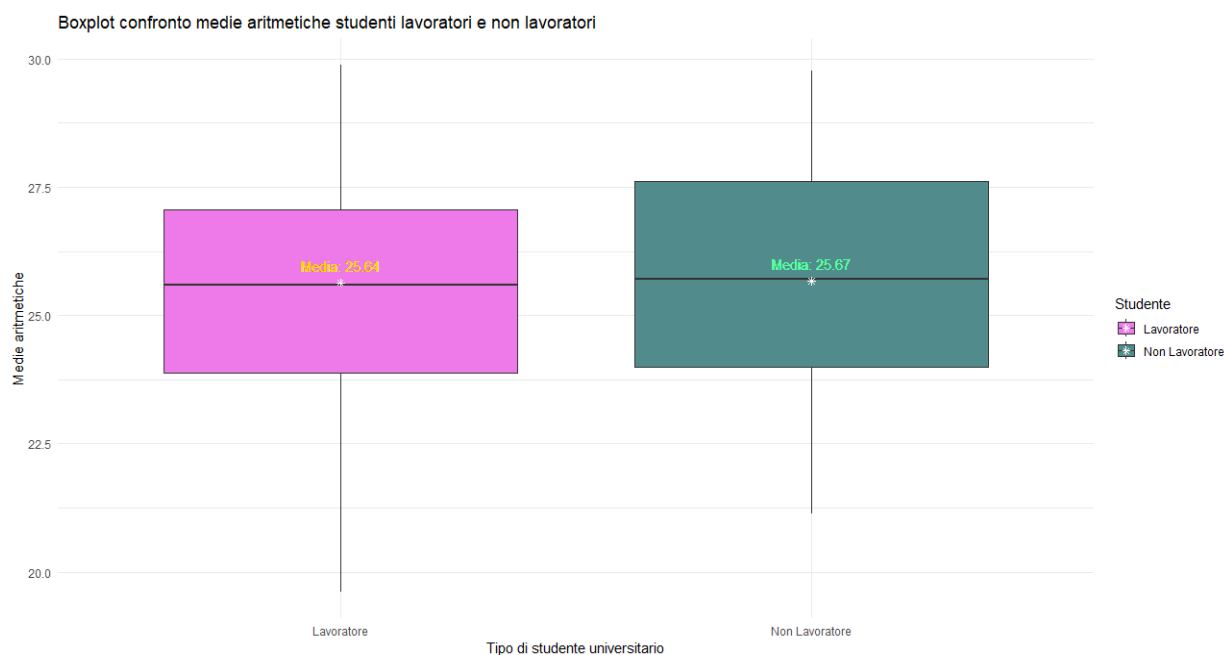
## Confronto studenti lavoratori e non

A questo punto possiamo analizzare se sussistono disparità tra i 48 studenti lavoratori e i 61 studenti non lavoratori rispetto al numero di esami superati. Facciamo questa analisi attraverso il confronto di due grafici a barre in cui vengono riportati su ogni barra il numero di esami superati.



*Figura 5 Confronto tra il numero di esami superati da studenti lavoratori e non lavoratori*

Nella Figura 5 vengono rappresentati due grafici a barre messi a confronto aventi sull'asse delle ascisse il numero di esami superati e sull'asse delle ordinate le frequenze. Si può notare che tra gli studenti lavoratori vi è uno studente che ha superato solamente 4 esami mentre tra gli studenti non lavoratori vi sono due studenti che hanno superato ciascuno 3 e 4 esami. Per un numero di esami superati pari a 7 possiamo notare solamente 3 studenti lavoratori rispetto ai 14 per gli studenti non lavoratori. Per 8 esami superati non vi è una differenza marcata e addirittura ci sono 16 studenti lavoratori rispetto a 13 studenti non lavoratori. Per un numero di esami superati pari a 9 ci sono 15 studenti lavoratori e 22 studenti non lavoratori. Questi risultati ci fanno ipotizzare che il materiale pubblicato online dai professori e la didattica a distanza siano stati d'aiuto agli studenti lavoratori per riuscire a sostenere tra gli 8 e 9 esami. Dopo aver confrontato il numero di esami superati tra studenti lavoratori e non lavoratori possiamo confrontare con due box plot le medie aritmetiche dei due gruppi dopo averle calcolate per ogni singolo studente.



*Figura 6 Confronto delle medie aritmetiche tra studenti lavoratori e non lavoratori*

Nella Figura 6 si può osservare che non c'è nessuna differenza marcata tra le medie aritmetiche degli studenti lavoratori e degli studenti non lavoratori quindi questo fattore non sembra avere nessun impatto sulle medie aritmetiche degli studenti. Entrambi i box plot risultano simmetrici con una media e una mediana che coincidono per entrambi. Possiamo osservare dai baffi del box plot dei lavoratori che ci sono più studenti con medie aritmetiche basse con un valore minimo di 19.60 rispetto al box plot dei non lavoratori con un valore minimo di 21.14. È da notare che tra gli studenti lavoratori vi è un valore massimo di 29.89 rispetto ai non lavoratori con un valore massimo di 29.79. Attraverso un t-test con 106 gradi di libertà per il confronto delle medie tra le due tipologie di studenti possiamo dire che non si può rifiutare l'ipotesi nulla con un valore del P-Value di 0.4698 e quindi c'è un'uguaglianza tra le due medie.

## Analisi per genere

Confrontiamo ora le medie aritmetiche per genere tra i 37 studenti del corso che hanno superato tutti e 9 gli esami e i 26 studenti che hanno superato tra i 3 e 6 esami.

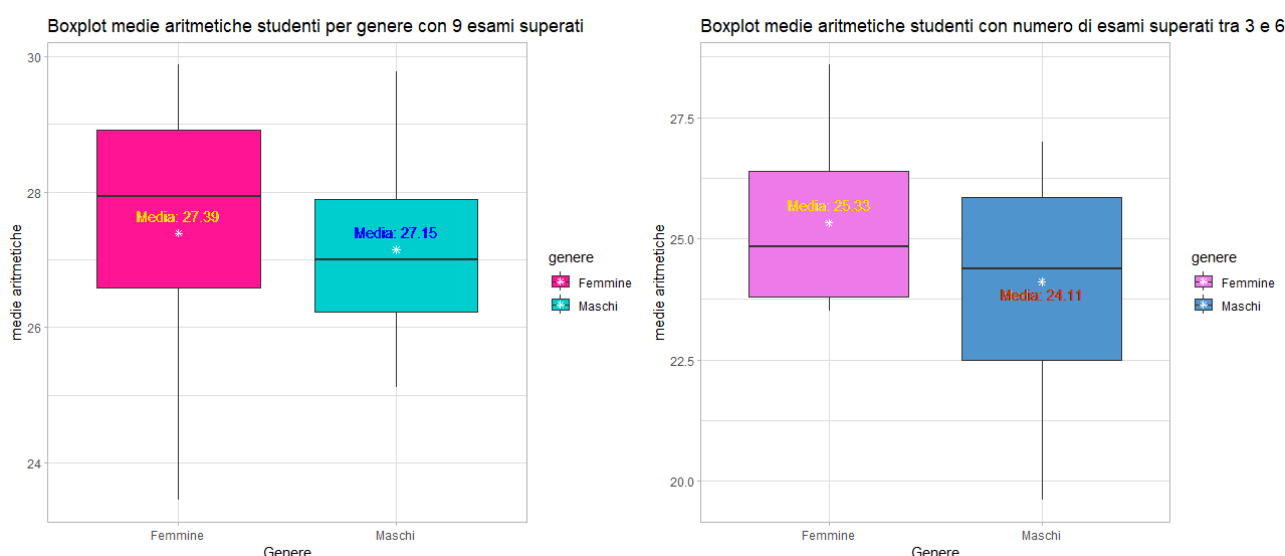


Figura 7 Boxplot confronto delle medie aritmetiche per genere e superamento esami

Dal primo box plot della Figura 7 si evince che la media aritmetica delle 16 studentesse è superiore alla media dei 21 studenti anche se lievemente. Inoltre, il box plot delle studentesse ha un prolungamento verso il basso per via dell'elevato numero di medie aritmetiche inferiori al primo quartile di 26.58 con un valore minimo di 23.44. Per il box plot degli studenti invece c'è un prolungamento verso l'alto per l'elevato numero di valori che superano la media aritmetica di 27.89 con un valore massimo di 29.78. Caso opposto per il box plot che tiene conto delle medie aritmetiche delle 10 studentesse con un numero di esami superati tra 3 e 6: si ha il massimo di 28.60 per una studentessa fuori sede con 5 esami superati mentre il box plot dei 16 studenti vi è un valore minimo di 19.60 per uno studente lavoratore con 6 esami superati. Entrambi i box plot degli studenti risultano simmetrici con un valore della media vicino a quello della mediana. Nel box plot delle studentesse con 9 esami superati vi è una leggera asimmetria positiva al contrario del box plot delle studentesse con un numero di esami superati tra i 3 e 6 esami che presenta una leggera asimmetria negativa. Nel box plot degli studenti con 9 esami superati è da evidenziare come nella Figura 5 che ci sono 15 studenti lavoratori e 22 studenti non lavoratori e anche 4 studenti fuori sede. Questi dati potrebbero confermare la nostra ipotesi riportata nei paragrafi precedenti sul funzionamento della didattica a distanza e del materiale online caricato dai docenti.



## MODELLI DI REGRESSIONE

### Modello di regressione lineare multipla

Dopo aver effettuato un'analisi descrittiva dei dati, possiamo usare dei metodi di analisi di regressione lineare multipla per individuare dipendenza tra le variabili. Analizziamo un modello di regressione lineare della variabile risposta 'medie aritmetiche' in funzione delle variabili esplicative 'voto preso alla maturità' e 'numero di esami superati dagli studenti'.

#### Coefficienti

Term	Coef	SE Coef	T-Value	P-Value	VIF
Intercept	15,90	1,52	10,48	0,000	
Voto maturità	0,0691	0,0176	3,93	0.000183	1,12
Numero esami Superati	0,509	0,128	3,97	6.32e-05	1,12

#### Equazione

**Media aritmetica = 15,90 + 0,509 Numero esami Superati + 0,0691 Voto maturità**

Il modello di regressione lineare multipla preso in analisi utilizza le variabili esplicative 'Voto preso alla maturità', 'Numero esami Superati' per prevedere il valore della variabile 'Medie aritmetiche'. Il valore di 15.90 rappresenta la media aritmetica prevista quando il coefficiente di tutte le variabili esplicative sono uguali a zero. Il coefficiente associato alla variabile esplicativa 'Voto Maturità' indica che un aumento di uno nel voto preso alla maturità è associato ad un aumento di 0,0691 nel valore della media aritmetica stimata dal modello lineare (15.90), mantenendo costanti le altre variabili esplicative. Il coefficiente associato alla variabile 'Numero esami superati' indica che, mantenendo costanti le altre variabili esplicative, un esame superato dallo studente del corso di Economia e Big Data fa aumentare il valore della media aritmetica stimata dal modello lineare (15.90). Per ogni variabile esplicativa il valore del P-Value con un  $\alpha=0,1$  è significativo e quindi rifiutiamo l'ipotesi nulla che il coefficiente associato a ogni variabile esplicativa sia nullo. Dal modello si evidenzia che la variabile esplicativa 'Numero esami superati' è quella che incide di più sulla variabile risposta 'Media aritmetica'. Tutte le variabili esplicative hanno valori VIF inferiori a 2, e questo indica che la correlazione tra le variabili esplicative è relativamente bassa e che il modello è stabile.

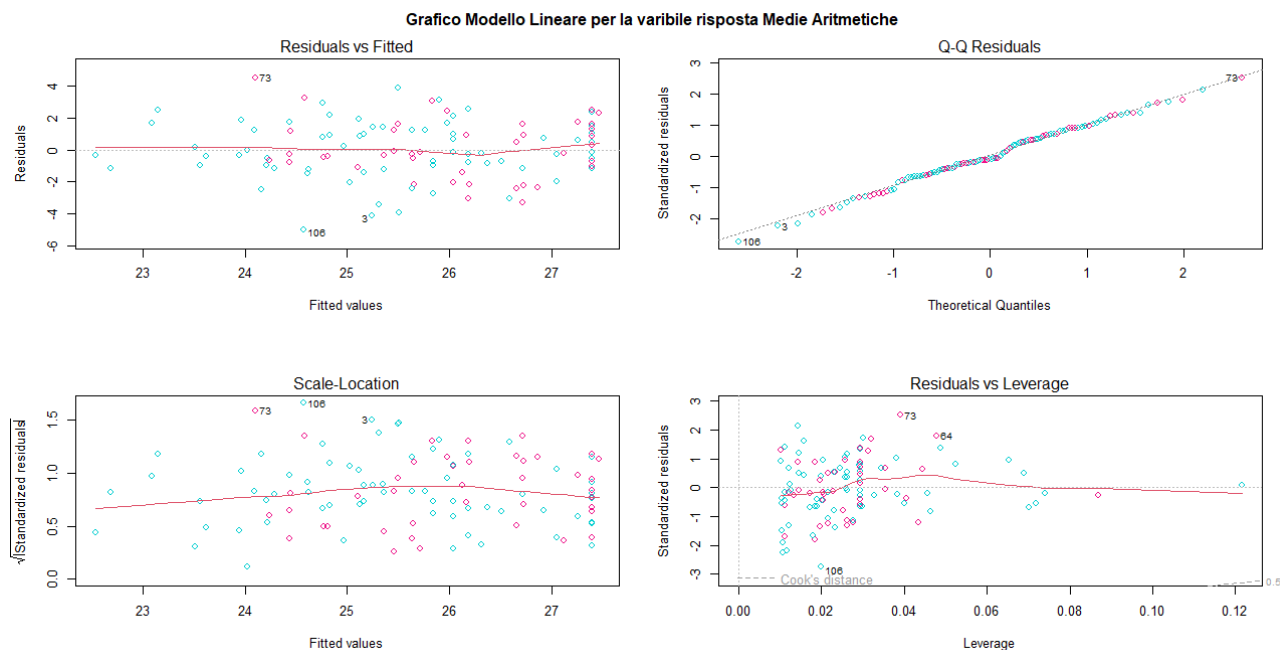
S	R-sq	R-sq(adj)	R-sq(pred)
1,82817	30,49%	29,16%	27,02%

Il valore del  $R^2$  aggiustato (R-sq(adj)) non risulta molto elevato per questo modello e indica che il 29,16% della varianza dei dati può essere spiegata dalle variabili del modello.

#### Residui medie aritmetiche

Min	Q1	Mediana	Q3	Max
-4.97	-1.1124	-0.1836	1.25	4.5

Analizzando la descrizione dei residui del modello, possiamo notare che il valore minimo e massimo dei residui sono abbastanza bassi, rispettivamente -4.97 e 4.5. Ci potrebbero essere pochi valori anomali o outliers presenti nei dati. Inoltre, il valore della mediana è vicino a zero, indicando che la maggior parte dei residui si concentra attorno ad esso rispettando ipotesi di omoschedasticità.



*Figura 8 Grafici dei residui del modello di regressione lineare multipla*

La Figura 8 mostra 4 tipi di grafici dei residui del modello di regressione lineare multipla analizzato precedentemente. Il primo grafico in alto a sinistra serve a verificare ipotesi di omoschedasticità dei residui e a mostrare una dispersione dei punti intorno allo zero e alcuni punti anomali distanti da esso. In particolare, il valore anomalo di 4.45 dello studente avente come identificativo 73, il valore di -4.096 dello studente avente come identificativo 3 e il valore di -4.97 dello studente avente come identificativo 106. Il secondo grafico in alto a destra ci mostra che la distribuzione dei residui è normale con l'eccezione di alcuni valori, tra cui le osservazioni 3, 73, 106. Il terzo grafico in basso a destra ci mostra che la radice quadratica dei residui standardizzati dovrebbero variare tra 0 e  $\sqrt{2}$  con l'eccezione delle osservazioni 3, 73, 106 che sono oltre questo range. Il quarto grafico in basso a destra ci mostra sull'asse delle ordinate i valori standardizzati dei residui e sull'asse delle ascisse il valore del leverage ovvero una misura che fornisce l'informazione su quali sono i valori influenti. In questo grafico non c'è nessuna osservazione che superi la distanza di Cook di  $D_i > 0,5$  e quindi non c'è nessuna osservazione che sia troppo influente. Va evidenziato che l'osservazione 106 è quella che si avvicina di più alla soglia critica di 0.5. In particolare gli studenti con identificativo 3, 73, 106 provenienti da indirizzo tecnico pur avendo conseguito buoni voti di maturità rispettivamente 84, 83, 82 hanno un andamento universitario diversificato: gli studenti aventi come identificativo 3, 106 hanno una media aritmetica molto bassa rispettivamente di 21.142, 19.60 mentre la studentessa di identificativo 73 ha una media aritmetica di 28.60 e questo ci fa ancora capire che l'indirizzo scolastico di provenienza degli studenti non è un fattore determinante per avere ottimi risultati nella carriera universitaria. L'istituto di provenienza può essere una base su cui partire, ma occorre metodologia e determinazione.

## Modello di regressione logistica

Analizziamo un modello logistico per calcolare la probabilità di superare l'esame di Metodi matematici della finanza in funzione del voto ottenuto dagli studenti all'esame di Matematica Generale preso in precedenza.

### Coefficienti

Term	Coef	SE Coef	Z-Value	P-Value
Intercept	-4.54394	1.379	-3.3	0.000979
Matematica	0.20608	0.0556	3.7	0.000211

### Probabilità

$$p = \frac{e^{(-4.54394 + 0.20608 \times \text{voto preso a matematica})}}{1 + e^{(-4.54394 + 0.20608 \times \text{voto preso a matematica})}}$$

La tabella dei coefficienti riporta le stime e gli errori standard per i coefficienti del modello. Nel nostro caso, abbiamo un termine per l'intercetta (Intercept) e un termine per il voto in Matematica Generale. Entrambi i coefficienti hanno valori stimati positivi, indicando una relazione positiva tra il voto di Matematica Generale e la possibilità di superare l'esame di Metodi Matematici per la Finanza. La variabile Matematica risulta essere altamente significativa, come indicato dal valore del P-Value molto basso con un  $\alpha=0,1$ , indicando che il voto di Matematica Generale ha un impatto significativo sulla possibilità di superare l'esame di Metodi Matematici per la Finanza.

### Numero di iterazioni del metodo di Fisher scoring: 4

Il numero di iterazioni del metodo di Fisher Scoring indica il numero di iterazioni necessarie per stimare i coefficienti del modello. In questo caso specifico, si hanno solo quattro iterazioni necessarie per stimare i coefficienti del modello, possiamo dunque ritenere che il processo di stima sia avvenuto in modo rapido e stabile.

Ciò suggerisce che i coefficienti stimati nel modello di regressione potrebbero essere considerati affidabili e rappresentare adeguatamente la relazione tra le variabili nel contesto dello studio. Tuttavia, è importante notare che il numero di iterazioni da solo potrebbe non essere sufficiente per trarre conclusioni definitive sulla bontà di adattamento del modello.

$$p = \frac{e^{(-4.54394 + 0.20608 \times 31)}}{1 + e^{(-4.54394 + 0.20608 \times 31)}} = 0.8635$$

Infine, viene calcolata una probabilità  $p$  basata sui coefficienti del modello. Nel nostro caso, viene calcolata la probabilità che uno studente con un voto di Matematica Generale pari a 31 (30 e lode) superi l'esame di Metodi Matematici per la Finanza. La probabilità calcolata è di 0.8635, indicando che c'è una probabilità del 86.35% che uno studente con un voto di Matematica Generale pari a 31 (30 e lode) superi l'esame di Metodi Matematici per la Finanza.

L'analisi dei dati ha evidenziato che un ottimo voto all'esame di Matematica Generale è associato a una maggiore probabilità di successo nell'esame di Metodi Matematici per la Finanza.

Questo suggerisce che consolidare solide basi e competenze attraverso lo studio di Matematica Generale può fornire agli studenti strumenti di analisi più robusti per affrontare l'esame di Metodi Matematici.

Alla luce di tali risultati, ricordando il nostro obiettivo di valutazione e feedback a livello didattico del corso in Economia e Big Data, è possibile suggerire alcune strategie didattiche per favorire il successo degli studenti. Ad esempio, potrebbe essere utile introdurre tutorati di matematica e offrire sportelli di sostegno nel semestre successivo a quello in cui gli studenti partecipano al corso di Matematica Generale. Queste iniziative mirano a mantenere gli studenti coinvolti nel processo di apprendimento della matematica e a consolidare ulteriormente le conoscenze acquisite durante il corso. Inoltre, si potrebbe considerare l'organizzazione di un corso estivo di ripasso e consolidamento di Matematica Generale da tenere nel mese di settembre. Questo corso fornirebbe agli studenti un'opportunità aggiuntiva per rafforzare le loro competenze matematiche prima di affrontare l'esame di Metodi Matematici. L'obiettivo di questa proposta è quello di aumentare ulteriormente le probabilità di eccellere da parte degli studenti nel superare l'esame di Metodi Matematici.

## TECNICHE MULTIVARIATE

### Analisi fattoriale

Svolgiamo un'analisi fattoriale con l'utilizzo di tre fattori sui 9 esami superati da 37 studenti usando come tecnica di rotazione dei fattori la "varimax". L'obiettivo della rotazione "varimax" è di ottenere una soluzione dei fattori più interpretabile e significativa, in cui le variabili hanno carichi fattoriali elevati su un singolo fattore e carichi bassi o nulli sugli altri fattori.

ESAMI	FATTORE 1	FATTORE 2	FATTORE 3
Matematica generale	0.511	0.160	0.492
Economia politica	0.282	0.953	
Economia Aziendale	0.678	0.255	
Fondamenti giuridici	0.179		0.981
Informatica e programmazione	0.125	0.387	0.395
Statistica	0.428	0.444	0.165
Metodi Matematici per la finanza	0.724	0.160	0.241
Politica Economica	0.149	0.351	0.163
Scienza della Finanza	0.607	0.258	0.205

Tabella 2 Analisi fattoriale

Analizzando la tabella 2 composta da 9 variabili, possiamo osservare che i dati sono spiegati da 3 fattori. Tuttavia, i risultati ottenuti non sono molto concreti contro ogni previsione. Sono però scaturiti alcuni risultati interessanti. In particolare, le valutazioni dell'esame di Fondamenti Giuridici mostrano un andamento anomalo e viene meglio rappresentato dal terzo fattore. Il primo fattore evidenzia l'importanza delle valutazioni di Metodi Matematici per la Finanza, Economia aziendale, Scienza della Finanza e Matematica generale. Dal secondo fattore emerge che solo le valutazioni di Economia Politica vengono spiegate.

Anche se si tentasse di aggiungere ulteriori fattori per spiegare meglio il modello non si riuscirebbe a ottenere risultati più concreti. Possiamo concludere che il primo fattore tende a spiegare meglio le discipline scientifiche.

Tra le variabili si riscontrano correlazioni significative e inaspettate come quella tra i risultati di Matematica generale e Fondamenti Giuridici (0.579), Metodi Matematici ed Economia Aziendale (0.594), Economia Aziendale e Scienze delle Finanze (0.594) e tra i risultati di Scienze delle Finanze e Matematica generale (0.541). Non si osservano correlazioni significative tra le discipline scientifiche.

## Analisi dei cluster

Come ultima analisi, osserviamo tramite due cluster la distribuzione dei 37 studenti che hanno superato tutti e 9 gli esami del corso di Economia e Big Data

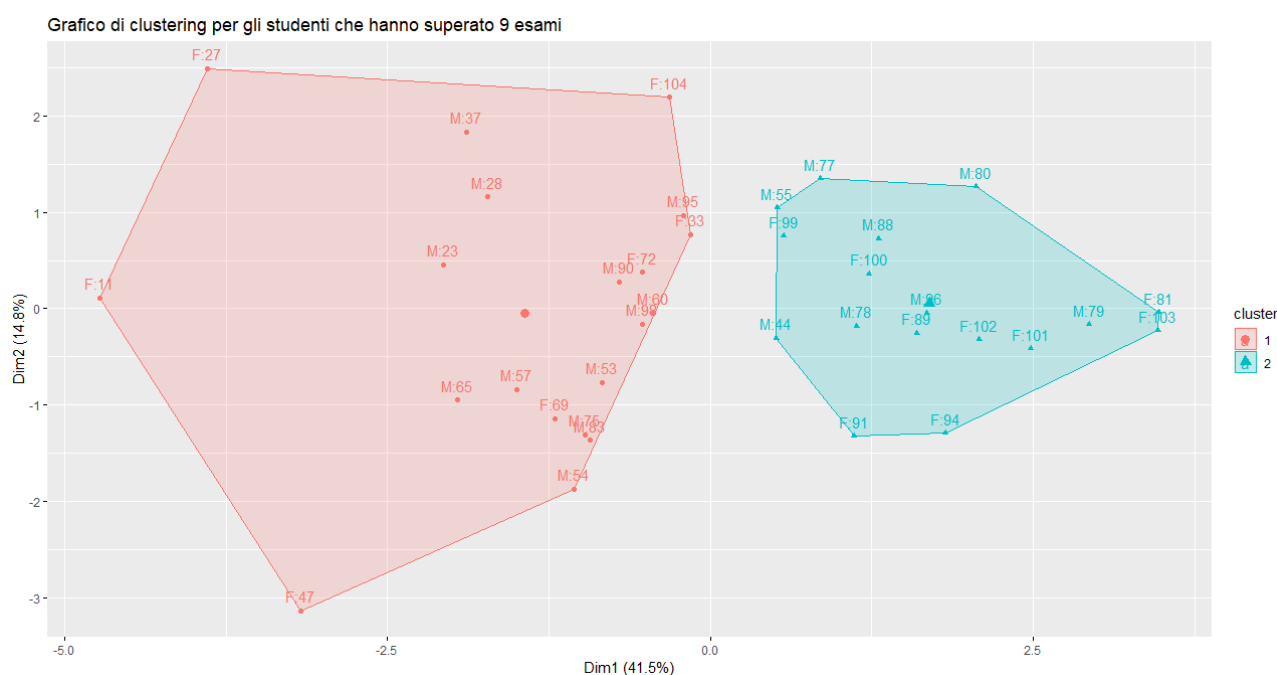


Figura 9 Cluster degli studenti che hanno superato tutti e 9 gli esami

Nella Figura 9 possiamo osservare la distribuzione dei due cluster aventi su ogni punto il genere e l'identificativo dello studente. Nel primo cluster, troviamo gli studenti aventi come valore delle medie aritmetiche inferiori a 27.89 e che hanno superato 9 esami, mentre nell'altro cluster si trovano gli studenti con medie aritmetiche maggiori o uguali di 27.89. Osservando la distribuzione all'interno di ciascun cluster, notiamo che gli studenti del cluster 1 presentano valori con più variabili, con minimi e massimi molto distanti tra di loro. D'altra parte, il secondo cluster sembra essere più omogeneo, con una distanza più ridotta tra i punti e il centroide. Inoltre, possiamo notare che nel secondo cluster ci sono 9 studentesse e 8 studenti rispetto al primo cluster con 7 studentesse e 13 studenti. Da questa analisi si nota che in questo corso, le studentesse ottengono risultati migliori rispetto agli studenti. Nel primo cluster, le medie aritmetiche degli studenti si attestano intorno a 23.44 con un massimo di 27.22. Per il secondo cluster, medie aritmetiche da 27.89 a 29.89 di una studentessa. Le dimensioni 1 e 2 spiegano complessivamente il 55.93% della variazione presente nei dati mentre il restante 44.07% della varianza rimanente non viene spiegata dalle dimensioni principali utilizzate nel modello.

## CONCLUSIONE

Con l'analisi appena svolta sul campione di 110 studenti del corso di Economia e Big Data, abbiamo osservato le difficoltà riscontrate durante alcuni esami e le differenze tra le valutazioni di studenti e studentesse. Quasi il 34% degli studenti ha superato tutti i 9 esami con medie aritmetiche molto elevate, il che suggerisce che una porzione di questo campione potrebbe rappresentare i migliori studenti del corso, che probabilmente saranno i primi laureati del prossimo anno. In questo campione di 37 studenti, abbiamo identificato 15 studenti lavoratori (8 studentesse e 7 studenti) e 22 studenti non lavoratori, oltre a 4 studenti fuori sede. Probabilmente, la didattica a distanza e il materiale fornito dai docenti hanno contribuito agli ottimi risultati di studenti lavoratori e fuori sede. Da notare che il divario di genere si è ridotto: il numero di studenti di ciascun sesso tende ad equivalersi, merito anche della borsa di studio Girls4Innovation dedicata a studentesse meritevoli iscritte al primo anno. Si spera che le competenze acquisite si traducano in tassi di occupazione più elevati e redditi superiori per le studentesse. Attraverso un modello di regressione lineare multipla, abbiamo osservato una correlazione significativa tra le medie aritmetiche e il numero di esami superati. Ciò suggerisce che una parte degli studenti di Economia e Big Data sono propensi a mantenere una media aritmetica stabile e tentare di migliorarla anche ripetendo gli esami. Per un'analisi più precisa della capacità di studio degli studenti, sarebbe stato utile avere dati sul numero di appelli che sono stati necessari a ciascun studente per superare ogni esame. L'esame di Informatica e Programmazione del primo anno ha prodotto le maggiori difficoltà, con i valori più bassi di superamenti e medie aritmetiche rispetto agli altri esami. Pertanto, sarebbe utile rafforzare le competenze degli studenti in questo campo, molto richiesto nel mondo del lavoro attuale, ad esempio, rendendo disponibile una parte delle lezioni in laboratorio. Abbiamo scelto di non considerare le variabili età e numero di fratelli e sorelle, non ritenendole determinanti per i risultati ottenuti. Tuttavia, i dati disponibili mostrano che l'età delle studentesse è per lo più di 20 anni (66%), seguite da 13 di 21 anni e una di 23 anni. Per gli studenti, l'età è più varia, con un picco di ventenni al 53%, seguiti da 21 studenti di 21 anni, 5 di 22, 3 di 23, e 2 ciascuno di 24 e 26 anni. Da notare la presenza di uno studente di 19 anni. Il 65% degli studenti ha un fratello o una sorella, mentre il 16.3% sono figli unici. Solo 3 studenti hanno 4 fratelli. Questo risulta conforme ai dati ISTAT "Popolazione e famiglie" del 2022. La variabile fumatori non è stata considerata in quanto non rilevanti alla fine dell'indagine. Si segnala che 31 studenti fumano (20 maschi e 11 femmine), anche se l'osservazione non è completa poiché non si conosce la frequenza del fumo e potrebbe essere occasionale. Infine, 85 studenti si ritengono soddisfatti di questo nuovo corso di laurea triennale, mentre 10 non lo sono, tra cui 4 che hanno superato tutti i 9 esami con valutazioni eccellenti. Nonostante i dati limitati, l'analisi riflette pienamente l'epoca che stiamo vivendo: l'età anagrafica degli studenti, la composizione media delle famiglie, la tenacia e la determinazione delle studentesse che si avvicinano sempre più alle lauree STEM, e lo sforzo delle università per incentivare le immatricolazioni femminili e colmare il divario esistente. È positiva l'opportunità data agli studenti di avere 3 appelli per le sessioni invernali e estive, oltre a un appello a settembre, e di avere tutor per aiutare a superare le difficoltà e le frustrazioni legate agli esami più ostici.