

# UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Scienze Statistiche

Corso di Laurea Triennale in  
Statistica per l'Economia e l'Impresa



## **DONNE NELLE DISCIPLINE STEM: UNA VALUTAZIONE DELL'EQUITÀ DI GENERE**

***Relatrice:*** Prof.ssa Francesca Bassi  
Dipartimento di Scienze Statistiche

***Laureanda:*** Rui Giulia  
***Matricola N°:*** 2001327

Anno Accademico 2022/2023



*A mia nonna Lola*



# INDICE

<b><i>Premessa</i></b> .....	<b>7</b>
Il divario di genere nello scenario STEM.....	7
<b><i>Introduzione</i></b> .....	<b>9</b>
<b><i>Materiali e metodi</i></b> .....	<b>11</b>
Raccolta e struttura dati .....	11
Popolazione di riferimento.....	12
Strumenti per l'analisi dei dati.....	12
<b><i>Analisi dei dati</i></b> .....	<b>17</b>
Iscritti diciannovenni .....	17
Iscritti per gruppo e sesso .....	18
<b><i>Conclusioni</i></b> .....	<b>43</b>
<b><i>Appendice A</i></b> .....	<b>45</b>
Formule Excel.....	45
Comandi RStudio.....	46
<b><i>Bibliografia e sitografia</i></b> .....	<b>47</b>



# Premessa

## Il divario di genere nello scenario STEM

“L'elenco dei grandi della matematica e della scienza nel corso della storia è un elenco di uomini: Pitagora, Ippocrate, Aristotele, Euclide, Copernico, Galileo, Keplero, Fermat, Newton, Gauss, Faraday, Darwin, Mendel, Einstein, con alcune donne che entrano in questo pantheon a partire dalla fine del XIX e l'inizio del XX secolo, come Curie e Mead. Anche i campi più nuovi, come l'informatica e la tecnologia dell'informazione, nonché la genomica, sono dominati dagli uomini: da Jobs e Gates a Sanger, Watson e Collins. Pertanto, non sorprende che le persone associno scienza, tecnologia, ingegneria e matematica (STEM) agli uomini e concludano che, poiché gli scienziati di maggior successo sono uomini, forse solo gli uomini possono perseguire con successo la scienza.” (Kahn S. e Ginther D., 2017).

Il gender gap, il divario di genere tra uomo e donna, si colloca in tanti ambiti diversi che impattano profondamente sulla vita quotidiana e sul suo svolgimento, come per esempio la salute, l'educazione, il lavoro e l'accesso alle attività economiche.

Si tratta di uno scompenso che si cerca di colmare da ormai molto tempo, in particolare, gli ultimi anni sono stati segnati da importanti battute d'arresto per la parità di genere a livello globale, con precedenti progressi interrotti dall'impatto della pandemia di COVID-19 per quanto riguarda donne e ragazze nell'istruzione e nella forza lavoro.

Uno dei tratti caratterizzanti del gender gap è la scarsa presenza di donne nelle discipline e nelle professioni STEM, ovvero l'acronimo con cui si indicano tutte le discipline scientifico-tecnologiche (dall'inglese Science, Technology, Engineering and Mathematics) sia in ambito accademico sia in ambito lavorativo, che può avere rilevanti conseguenze nella società e nel mondo del lavoro.

La disuguaglianza di genere nelle STEM, che parte già in età infantile e cresce a partire dall'età adolescenziale, è un percorso tortuoso fatto di stereotipi e mancanza di modelli di ruolo.





# Introduzione

Il seguente elaborato si propone di analizzare la disparità di genere nelle professioni STEM (Science, Technology, Engineering and Mathematics) con l'ausilio di dati forniti dall'ente MUR.

Lo scopo è quello di analizzare nel corso dell'ultimo decennio, dall'anno accademico 2012/2013 all'anno 2021/2022, la sottorappresentanza della donna nelle discipline STEM nel percorso universitario.

Per poter svolgere tale analisi sono state prese in considerazione alcune serie storiche, ovvero insieme di dati ordinati rispetto al tempo, perché utili strumenti sia per interpretare un fenomeno, individuare componenti di trend, di ciclicità, di stagionalità e di accidentalità, e sia per prevedere il suo andamento futuro.

È stato riposto interesse in questo tema perché nonostante ci siano stati progressi nell'equità di genere e una crescita di interesse nell'ultimo decennio in informatica, ingegneria, matematica e statistica, tra uomini e donne, le donne continuano a restare una componente inferiore nei campi STEM.

Questa disparità è decisamente preoccupante perché porta a una mancanza di diversità e inclusione e, in definitiva, limita il potenziale dell'industria STEM. Affrontare la sottorappresentazione è fondamentale poiché il mondo è alle prese con i rischi economici, ambientali, geopolitici, sociali e tecnologici.

Colmare il divario di genere non solo rafforzerà la crescita e l'innovazione sostenibili abilitate alla tecnologia, ma è anche considerata una necessità economica.

Il divario di genere nelle STEM è stato attribuito a diverse realtà di lunga data e profondamente radicate nella cultura. Tra i principali fattori che influiscono si hanno gli stereotipi: molti individui associano sin da piccoli i campi STEM alle qualità maschili mentre, ad esempio, discipline letterarie alle qualità femminili portando a scoraggiare le ragazze e le donne dal perseguire l'istruzione e le carriere STEM, e la mancanza di modelli di ruolo: le donne rimangono sottorappresentate nelle posizioni di leadership STEM, il che rende più difficile per le future e attuali donne trovare modelli di ruolo e mentori sul campo.

Questo elaborato, perciò, ha lo scopo di analizzare nel corso del tempo se ci sono stati dei progressi per quanto riguarda la componente femminile nelle discipline STEM e prevedere se negli anni avvenire ci sarà un possibile miglioramento o se la situazione rimarrà stazionaria.

In conclusione, verranno fatte considerazioni generali sui risultati ottenuti dalle analisi svolte e verranno supposte varie strategie che potrebbero essere implementate per colmare il divario di genere nelle STEM.



# Materiali e metodi

## Raccolta e struttura dati

I dati sono stati raccolti dal sito <https://ustat.mur.gov.it/>, specificamente dalla sezione "Open Data", selezionando la categoria "Studenti" e accedendo ai dati disponibili per il "Bilancio di Genere".

Il MIUR “Ministero dell’Istruzione, dell’Università e della Ricerca”, che si occupa della gestione dell’istruzione scolastica, universitaria e di alta formazione artistica, musicale e coreutica e di ricerca scientifica e tecnologica, il 9 gennaio 2020 si è scisso in due dicasteri distinti: il MI, cioè il Ministero che si occupa dell’Istruzione, e il MUR ovvero il Ministero dell’Università e Ricerca.

Il MUR ha il compito di promuovere e sviluppare l’attività di ricerca scientifica e tecnologica italiana, in collegamento con le altre amministrazioni dello stato, centrali e regionali, che hanno competenze analoghe in materia.

Tale organo è composto da ricercatori (pubblici o privati) e, per le università, da docenti impegnati in attività di ricerca.

Una delle molteplici azioni e attività svolte dal MUR, utile per reperire i dati necessari per le analisi svolte nell’elaborato, è la comunicazione delle attività di ricerca promosse dal MUR e/o svolte dai soggetti del sistema nazionale, ad esempio, attraverso il sito Ustat.

Il portale dell’Ustat è gestito dall’Ufficio Statistica e Studi che fa parte del Sistema Statistico Nazionale (Sistan) in quanto si occupa del coordinamento e della pubblicazione delle statistiche ufficiali riguardanti la formazione terziaria.

L’acquisizione, l’elaborazione e la diffusione dei dati è svolta attraverso fonti di dati come l’Anagrafe Nazionale degli Studenti Universitari e dei Laureati (ANS), dalla quale sono estratte le statistiche riguardanti immatricolati, iscritti e laureati.

La serie storica utilizzata per questa tesi e poi utilizzata per l’analisi nei programmi RStudio, ambiente di sviluppo integrato per R, ed Excel di Microsoft Office è:

- “Iscritti per sesso e gruppo”: serie storica a partire dall’anno accademico 2012/2013 degli iscritti ai corsi di laurea disaggregati per genere e campi dell’istruzione e della formazione (ISCED-F 2013); l’analisi verrà prima svolta con il genere femminile e poi con il genere maschile.

Inoltre, il sito Ustat è stato consultato per i dati riguardanti gli iscritti all’università in Italia dall’anno accademico 2012/2013 all’anno 2021/2022.

I dati sono stati reperiti dal sito <https://ustat.mur.gov.it/>, consultando la sezione "Open Data" e selezionando le categorie "Studenti" e poi "Iscritti". Nello specifico, sono stati esaminati i dati relativi agli "Iscritti per anno accademico".

Per condurre l'analisi preliminare, sono stati consultati due siti web ufficiali. Nel dettaglio, il sito dell'Istat, <http://dati.istat.it/>, è stato utilizzato per ottenere dati sulla popolazione residente in Italia. Sono state esaminate le sezioni "Popolazione e famiglie" e "Popolazione", con particolare attenzione ai dati sulla "Popolazione residente al 1° gennaio" per gli anni 2019-2022 e sulla "Popolazione residente ricostruita-Anno 2001-2019", per gli anni 2012-2018. Per ottenere informazioni specifiche sugli studenti universitari diciannovenni è stato consultato il sito dell'Ustat, <https://ustat.mur.gov.it/>. In particolare, sono stati analizzati i dati disponibili nella sezione "Open Data" sotto la categoria "Studenti", focalizzandosi sugli "Iscritti per anno di nascita".

## Popolazione di riferimento

La popolazione di riferimento considerata corrisponde agli iscritti all'università in Italia dall'anno accademico 2012/2013 all'anno 2021/2022.

La scelta di considerare un intervallo temporale di un decennio è per poter meglio evidenziare nel tempo la disuguaglianza nelle discipline STEM.

Iscritti	2012/2013	2013/2014	2014/2015	2015/2016	2016/2017
Femmine	981.449	955.162	938.170	923.873	926.994
Maschi	742.891	734.175	727.511	727.146	741.141
<b>Totale</b>	<b>1.724.340</b>	<b>1.689.337</b>	<b>1.665.681</b>	<b>1.651.019</b>	<b>1.668.135</b>

Tabella 1.1: *Iscritti all'università in Italia dall'a.a. 2012/2013 all'a.a. 2016/2017.*

Iscritti	2017/2018	2018/2019	2019/2020	2020/2021	2021/2022
Femmine	939.146	953.962	980.276	1.029.054	1.028.829
Maschi	756.615	766.798	783.248	810.792	793.312
<b>Totale</b>	<b>1.695.761</b>	<b>1.720.760</b>	<b>1.763.524</b>	<b>1.839.846</b>	<b>1.822.141</b>

Tabella 1.2: *Iscritti all'università in Italia dall'a.a. 2017/2018 all'a.a. 2021/2022.*

## Strumenti per l'analisi dei dati

La prima parte dell'analisi si concentrerà sull'uso di variazioni assolute, variazioni relative e tassi per esplorare le dinamiche di partecipazione delle donne nelle discipline STEM.

Questi strumenti consentiranno di identificare tendenze significative e di valutare la direzione e l'entità delle variazioni nel tempo.

Successivamente, si analizzerà la serie storica dei dati che permetterà di comprendere come la componente femminile nelle STEM si sia evoluta nel corso degli anni.

L'obiettivo principale è quello di garantire un'analisi dei dati rigorosa, affidabile e in grado di rispondere alle domande di ricerca poste nel corso di questo studio.

La combinazione di variazioni assolute, relative, tassi e analisi della serie storica consentirà di ottenere una comprensione completa e approfondita della dinamica della partecipazione delle donne nelle discipline STEM.

### **Variazione assoluta**

La variazione assoluta rappresenta lo scarto tra due valori consecutivi,  $Y_{t-1}$  e  $Y_t$ , in una successione.

Misura quanto il valore dell'elemento in posizione  $t$  sia cambiato rispetto al valore dell'elemento in posizione  $t-1$ , e permette di avere una visione immediata del cambiamento (crescente o decrescente) della variabile utilizzata per l'analisi.

Può essere espressa in termini di unità operando con aggregati o in termini di punti percentuali operando con tassi.

$$\text{VARIAZIONE ASSOLUTA: } \Delta Y_t = Y_t - Y_{t-1}$$

Dove:

-  $Y_t$  è il valore dell'elemento in posizione  $t$  nella successione.

-  $Y_{t-1}$  è il valore dell'elemento in posizione  $t-1$  nella stessa successione.

### **Variazione relativa**

La variazione relativa rappresenta lo scarto percentuale tra due valori consecutivi,  $Y_{t-1}$  e  $Y_t$ , in una successione, diviso il valore in posizione  $t-1$ .

Questa misura è espressa come un numero percentuale, il che la rende particolarmente utile per valutare l'andamento della variabile nel tempo.

$$\text{VARIAZIONE RELATIVA: } \frac{Y_t - Y_{t-1}}{Y_{t-1}} \times 100$$

Dove:

-  $Y_t$  è il valore dell'elemento in posizione  $t$  nella successione.

-  $Y_{t-1}$  è il valore dell'elemento in posizione  $t-1$  nella stessa successione.

### **Tasso**

Il tasso rappresenta un indicatore che misura l'intensità di un fenomeno in un certo periodo, mettendo in relazione il numero di casi osservati durante quell'intervallo temporale con la dimensione della popolazione di riferimento.

$$\text{TASSO: } \frac{\text{Casi osservati}}{\text{Popolazione di riferimento}}$$

### **Serie storica**

La serie storica è una successione di osservazioni su una variabile numerica fatte sequenzialmente nel tempo.

L'ordine dei dati non è casuale, infatti, gli elementi sono ordinati in modo naturale dal valore dell'indice temporale e le osservazioni sono dipendenti perché i dati presentano "regolarità" o persistenze legate alla posizione dell'osservazione nella sequenza.

$$\{y_t\}_{t=1}^n = y_1, \dots, y_t, \dots, y_n$$

Può essere di tre tipi: univariata quando viene osservato un solo fenomeno, multipla quando ci sono più fenomeni osservati ed equispaziata quando le rilevazioni sono fatte in intervalli di tempo regolari.

Il parametro temporale  $t$ , che definisce l'ordinamento dei dati, appartiene ad un insieme parametrico  $T$ , che può essere continuo o discreto; quindi, si parla serie storica a parametro continuo o di serie storica a parametro discreto.

In particolare, le serie storiche discrete possono essere di due tipi ovvero serie storica di dati di flusso, quando si ottiene aggregando o cumulando i valori fatti registrare da una variabile in intervalli di tempo di pari ampiezza, e serie storica di dati di stock, quando si ottiene campionando un processo continuo in prefissati istanti temporali.

La serie storica utilizzata per questo elaborato è equispaziata a parametro discreto e relativa ad una variabile di stock.

Un modello stocastico generale per descrivere il processo generatore dei dati di una serie storica  $\{y_t\}_{t=1}^n$  relativa ad una variabile  $Y$  è dato da:

$$Y_t = f(t) + u_t$$

Composizione di:  $f(t)$  componente sistematica, puramente deterministica che dipende dal tempo e definisce la legge di evoluzione temporale del processo, e  $u_t$  sequenza di variabili casuali che rappresenta la componente stocastica della serie e segue una determinata legge di probabilità.

Secondo un approccio classico, dal punto di vista statistico, la componente  $u_t$  è generata da un processo white noise (processo a componenti incorrelate), ossia da una successione di variabili casuali indipendenti e identicamente distribuite, di media nulla e varianza costante. Se si considerasse, invece, un approccio moderno la componente stocastica sarebbe considerata un processo a componenti correlate.

La componente sistematica  $f(t)$ , invece, si assume che sia la risultante dell'azione congiunta di tre componenti, non osservabili direttamente, ma definibili nel piano concettuale: trend, ciclo e stagionalità.

$$f(t) = f(T_t, C_t, S_t)$$

### **Funzione di autocorrelazione (globale e parziale)**

L'autocorrelazione globale (ACF) e l'autocorrelazione parziale (PACF) sono due misure statistiche utilizzate nell'analisi delle serie storiche per valutare le correlazioni tra le osservazioni in diverse posizioni temporali.

- Autocorrelazione Globale (ACF):

è una misura della correlazione tra le osservazioni di una serie storica a diverse distanze temporali di ritardo.

$$\rho_{t,t+k} = p_k = \frac{\gamma_k}{\gamma_0}, \quad k = 0, \pm 1, \pm 2, \dots$$

dove  $\gamma_k$  è l'autocovarianza e  $\gamma_k = \gamma_{t,t+k} = E\{[Y_t - \mu][Y_{t+k} - \mu]\}$   
 $k = 0, \pm 1, \pm 2, \dots$

- Autocorrelazione Parziale (PACF):

misura l'autocorrelazione tra  $Y_t$  e  $Y_{t-k}$  al netto delle variabili intermedie.

$$P_k = \text{Corr}(Y_t, Y_{t-k} | Y_{t-1}, \dots, Y_{t-k+1})$$

Sia l'ACF che la PACF sono utili strumenti nell'analisi delle serie storiche perché aiutano a rivelare le relazioni temporali nei dati, identificare la stagionalità e selezionare i parametri appropriati per i modelli di previsione.

### **Previsione con i modelli ARIMA**

Si definisce processo autoregressivo integrato a media mobile di ordine (p, d, q), ARIMA (p, d, q), se  $\{\varepsilon_t\}$  è un processo white noise di media 0 e varianza  $\sigma_\varepsilon^2$  e se  $X_t$ , d-esima differenza di  $Y_t$ ,  $X_t = (1 - B)^d Y_t$ , è un processo ARMA (p, q).

Per prevedere il valore di una serie storica al tempo  $n + k$  si ipotizza di avere una serie  $\{y_t\}_{t=1}^n$  e di avere identificato e stimato, mediante la procedura di Box e Jenkins, un modello ARIMA (p, d, q), utilizzando l'informazione fino all'istante n.

La previsione  $\hat{y}_{n+k}$  di  $y_{n+k}$  è il particolare valore che assume il previsore  $\hat{Y}_{n+k}$  all'istante  $n + k$  ed è data da:

$$\hat{y}_{n+k} = E[Y_{n+k} | Y_n = y_n, Y_{n-1} = y_{n-1}, \dots, Y_1 = y_1, Y_0 = y_0, \dots] = E_n[Y_{n+k}]$$

ossia dal valore atteso di  $Y_{n+k}$  condizionato all'informazione passata.

Nella pratica ciò non è possibile e il condizionamento avviene solamente rispetto all'informazione realizzatasi nel passato, questo porta a considerare l'approssimazione:

$$E[Y_{n+k} | Y_n = y_n, Y_{n-1} = y_{n-1}, \dots, Y_1 = y_1, Y_0 = y_0, \dots] \approx E[Y_{n+k} | Y_n = y_n, \dots, Y_1 = y_1].$$





# Analisi dei dati

## Iscritti diciannovenni

Tasso	2012/2013	2013/2014	2014/2015	2015/2016	2016/2017
Femmine	39,42%	39,35%	40,67%	42,14%	44,20%
Maschi	26,72%	27,18%	28,20%	29,67%	30,98%
Totale	32,88%	33,08%	34,23%	35,69%	37,34%
Tasso	2017/2018	2018/2019	2019/2020	2020/2021	2021/2022
Femmine	46,00%	47,28%	50,18%	51,40%	54,14%
Maschi	31,61%	31,50%	33,65%	34,35%	35,57%
Totale	38,51%	38,94%	41,50%	42,50%	44,49%

Tabella 2.1: Tasso di iscrizione dei diciannovenni suddivisi per genere dall'a.a. 2012/2013 all'a.a. 2021/2022.

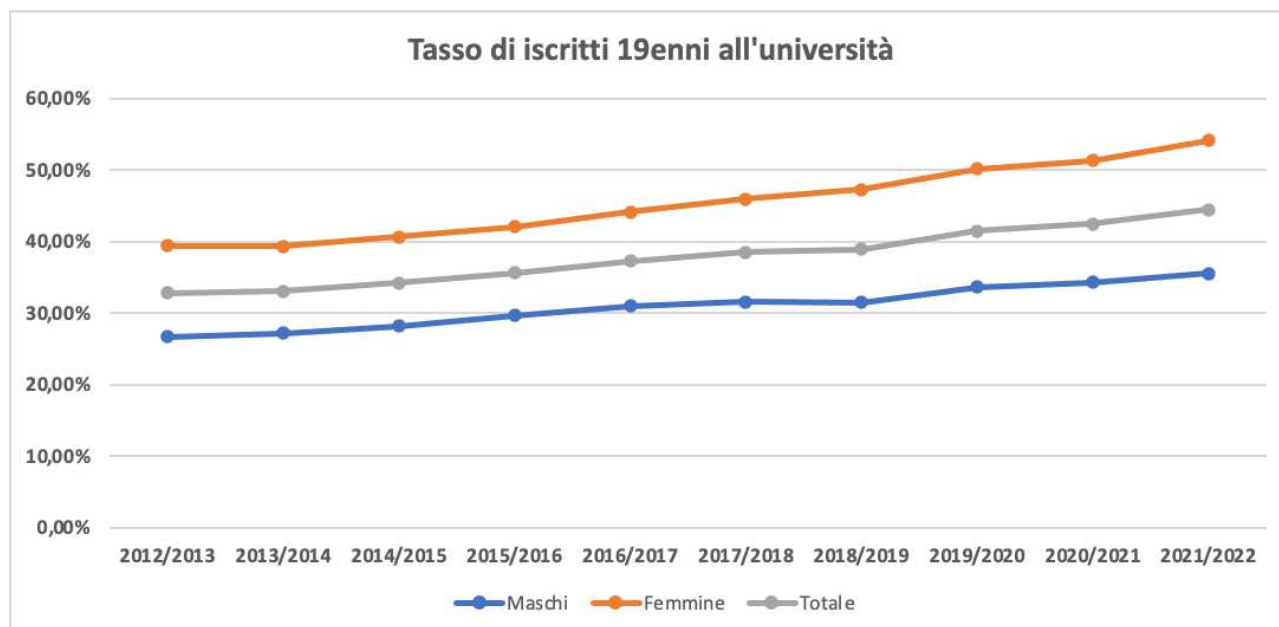


Figura 2.1: Andamento del tasso di iscritti diciannovenni suddivisi per genere dall'a.a. 2012/2013 all'a.a. 2021/2022.

Si è voluta svolgere un'analisi preliminare in cui si confronta il tasso di iscritti diciannovenni all'università in Italia suddivisi per genere dall'a.a. 2012/2013 all'a.a. 2021/2022.

Questo studio fornisce una panoramica dettagliata del cambiamento nel numero di iscrizioni universitarie tra i giovani adulti e identifica eventuali variazioni legate al genere che potrebbero essersi verificate nel corso degli anni.

Nel corso del periodo preso in considerazione, si è osservato un andamento generalmente crescente nel numero di iscritti e questo aumento può essere indicativo di diversi fattori, tra cui l'accessibilità all'istruzione superiore, le politiche educative nazionali e le opportunità di carriera correlate all'istruzione universitaria.

Tuttavia, è interessante notare una leggera diminuzione (figura 2.1), indipendentemente dal genere, nel numero di iscritti nell'anno accademico 2018/2019. Questo calo potrebbe essere dovuto a una serie di ragioni, come cambiamenti nelle politiche di ammissione o fluttuazioni demografiche, infatti si nota un calo di diciannovenni maschi a partire dall'anno 2019, o influenze socioeconomiche.

Per quanto riguarda la suddivisione per genere, si osserva che il tasso di femmine diciannovenni iscritte all'università è di circa il 20% maggiore, lungo tutto il decennio preso in considerazione per l'analisi, rispetto alla proporzione di maschi diciannovenni iscritti.

### **Iscritti per gruppo e sesso**

In questa parte dell'elaborato verrà analizzata la serie storica "Iscritti per sesso e gruppo" in cui sono registrate le iscrizioni, suddivise per genere, ai corsi di laurea relativi alle discipline STEM.

Le analisi verranno svolte con l'obiettivo di evidenziare la disuguaglianza nell'area STEM negli iscritti dell'ultimo decennio.

Guardando le tabelle di frequenza (tabella 2.3 e tabella 2.4), riportate qui di seguito, nel corso degli anni, emerge chiaramente un modello di sottorappresentazione delle studentesse nelle discipline STEM che può essere osservato in modo consistente nei diversi anni accademici.

Tuttavia, c'è un'interessante considerazione da fare in relazione alla variabile Natural sciences, Mathematics and Statistics nel corso dell'ultimo decennio che spicca nel primo grafico a barre (figura 2.2) dove si mette a confronto il genere femminile con quello maschile.

Nonostante il divario significativo delle studentesse nelle discipline STEM, per questa area di studio, si nota un aumento costante nella rappresentanza femminile. Questo indica un cambiamento positivo e incoraggiante nelle iscrizioni femminili per quello specifico settore nel corso del tempo.

Questo dato potrebbe essere il risultato di tentativi mirati per incoraggiare e supportare le ragazze nelle discipline legate alla variabile Natural sciences, Mathematics and Statistics.

Questi tentativi potrebbero includere iniziative per ridurre gli stereotipi di genere e presentazioni di corso più inclusive che attirano studentesse interessate a questo ambito.

Nei grafici a barre delle variabili Information and Communication Technologies (ICTs) ed Engineering, Manufacturing and Construction (figura 2.3 e figura 2.4), che mettono a confronto la rispettiva variabile con il genere, invece, emerge chiaramente un quadro in cui le donne iscritte sono nettamente inferiori rispetto ai maschi nel corso di tutto il decennio.

Questa discrepanza potrebbe essere dovuta ad esempio a stereotipi di genere radicati, mancanza di modelli femminili nelle discipline STEM, o barriere culturali e sociali che scoraggiano le ragazze dal perseguire queste carriere.

Per affrontare questa disparità, sono necessarie azioni concrete e mirate, come iniziative per incoraggiare le ragazze sin dalle prime fasi dell'istruzione a sviluppare un interesse per le materie STEM.

Iscritti Femmina	Natural sciences, Mathematics and Statistics	Information and Communication Technologies (ICTs)	Engineering, Manufacturing and Construction	Totale
2012/2013	80.033	3.237	86.999	170.269
2013/2014	76.873	3.218	85.436	165.527
2014/2015	74.611	3.278	84.711	162.600
2015/2016	75.456	3.376	83.363	162.195
2016/2017	77.631	3.756	83.182	164.569
2017/2018	81.094	4.047	83.024	168.165
2018/2019	83.864	4.328	82.968	171.160
2019/2020	85.981	4.776	83.826	174.583
2020/2021	90.437	5.274	84.840	180.551
2021/2022	91.795	5.910	85.169	182.874
<b>Totale</b>	<b>817.775</b>	<b>41.200</b>	<b>843.518</b>	<b>1.702.493</b>

Tabella 2.3: Iscrizioni femmine per discipline STEM dall'a.a. 2012/2013 all'a.a. 2021/2022.

Iscritti Maschi	Natural sciences, Mathematics and Statistics	Information and Communication Technologies (ICTs)	Engineering, Manufacturing and Construction	Totale
2012/2013	55.849	21.117	202.609	279.575
2013/2014	54.764	21.190	201.168	277.122
2014/2015	54.690	21.949	198.561	275.200
2015/2016	56.131	23.162	198.441	277.734
2016/2017	58.655	24.955	201.622	285.232
2017/2018	61.122	27.087	204.227	292.436
2018/2019	62.990	28.712	205.962	297.664
2019/2020	65.102	30.426	209.467	304.995
2020/2021	67.038	32.281	212.592	311.911
2021/2022	67.713	34.157	207.310	309.180
<b>Totale</b>	<b>604.054</b>	<b>265.036</b>	<b>2.041.959</b>	<b>2.911.049</b>

Tabella 2.4: Iscrizioni maschi per discipline STEM dall'a.a. 2012/2013 all'a.a. 2021/2022.

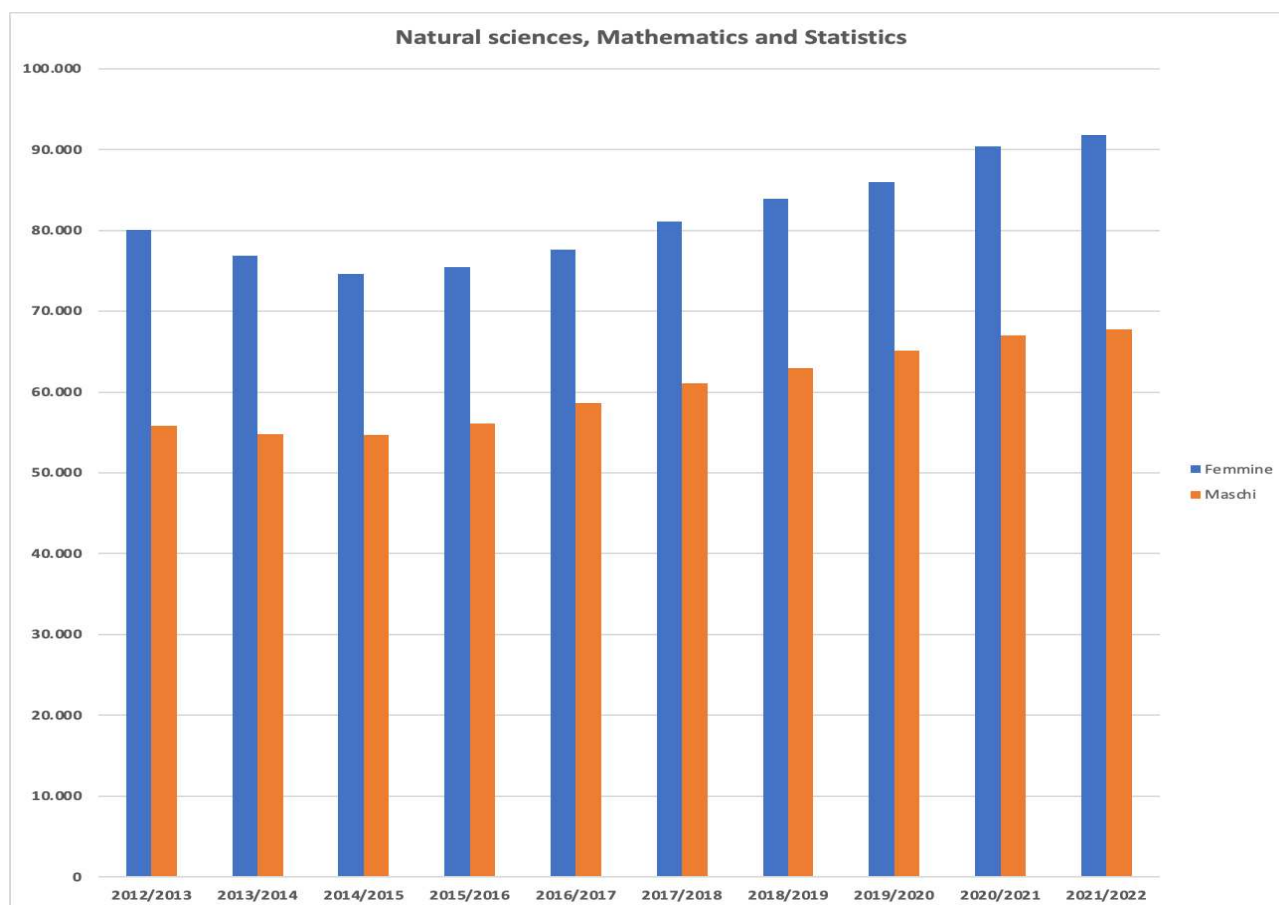


Figura 2.2: Diagramma a barre della variabile Natural Sciences, Mathematics and Statistics diviso per genere dall'a.a. 2012/2013 all'a.a. 2021/2022.

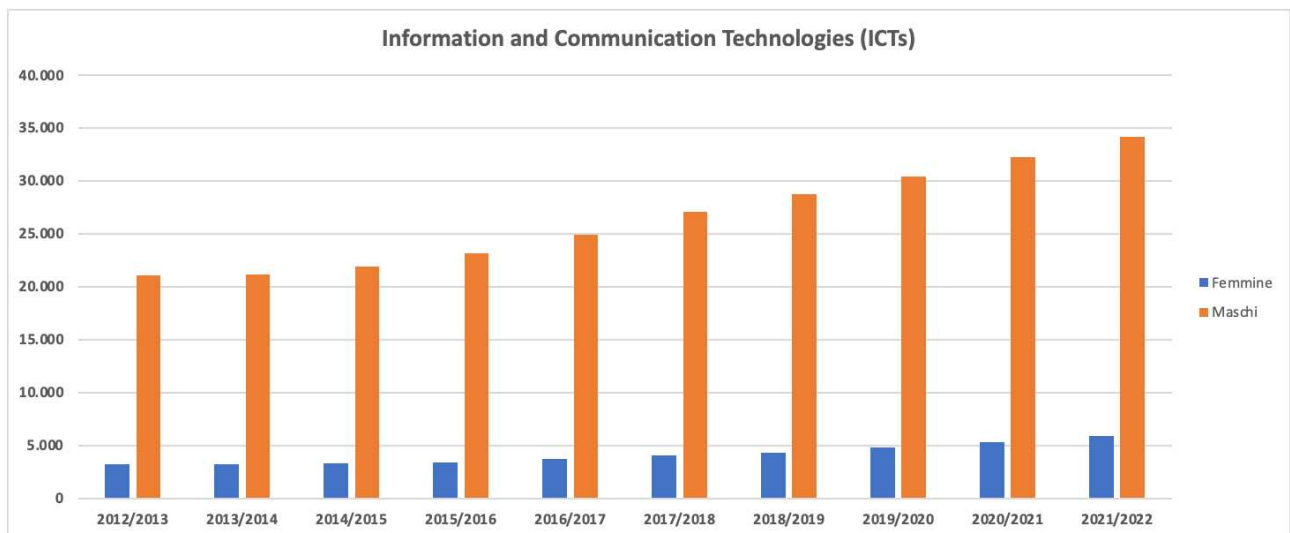


Figura 2.3: Diagramma a barre della variabile Information and Communication Technologies (ICTs) diviso per genere dall'a.a. 2012/2013 all'a.a. 2021/2022.

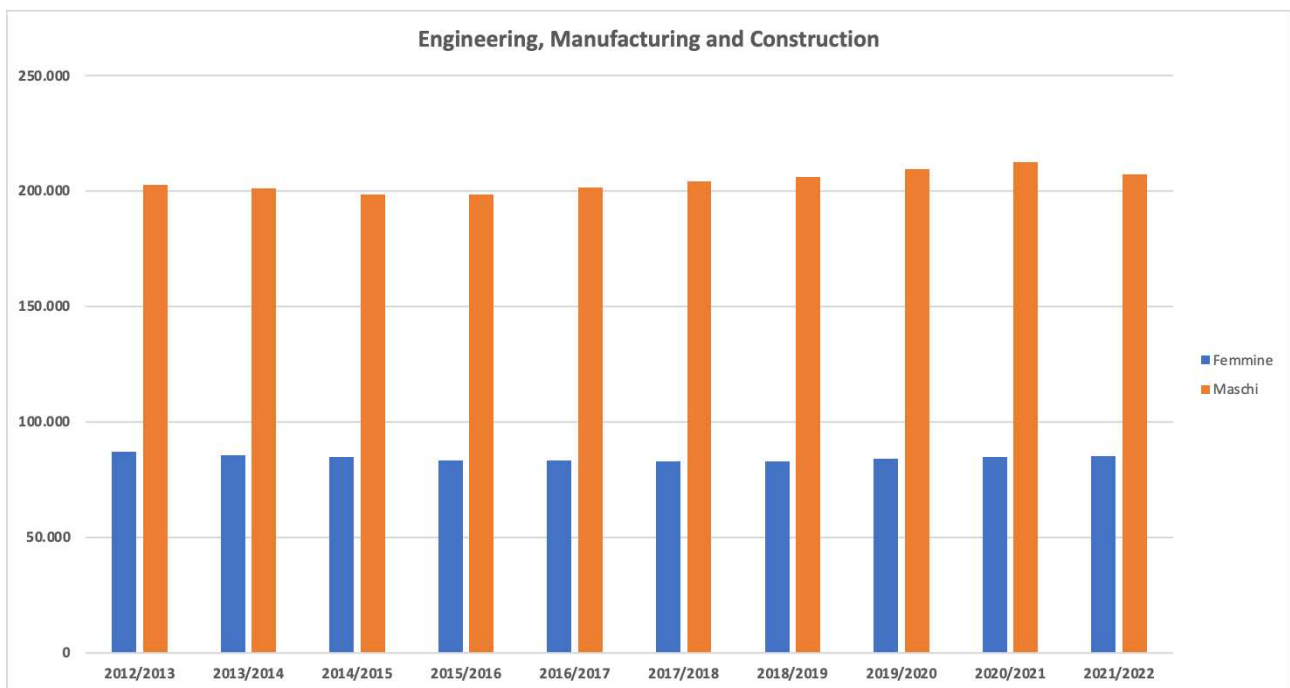


Figura 2.4: Diagramma a barre della variabile Engineering, Manufacturing and Construction diviso per genere dall'a.a. 2012/2013 all'a.a. 2021/2022.

Grazie alle variazioni relative delle femmine iscritte nelle discipline STEM (tabella 2.6) emerge che nel corso degli anni accademici, nell'ambito Natural sciences, Mathematics and Statistics, si verificano oscillazioni che vanno da un minimo di -3.95% ad un massimo di 5.18%.

Nell'ambito di Information and Communication Technologies (ICTs) si ha l'oscillazione più bassa con lo -0.59% e quella più alta con 12.06% rilevata negli ultimi due anni accademici presi in esame che porta a pensare che ci sia un interesse crescente da parte delle femmine in questo ambito.

Invece, in Engineering, Manufacturing and Construction, si può notare che nel corso del tempo la variazione relativa passa dall'essere negativa, toccando il picco a -1.80% (a.a. 2013/2014-2012/2013), all'essere positiva, superando l'1.00% nelle coppie di a.a. 2018/2019, 2019/2020 e 2020/2021.

Le variazioni relative percentuali degli iscritti maschi nelle discipline STEM (tabella 2.8) mostrano alcune tendenze interessanti durante gli anni accademici considerati.

In Natural sciences, Mathematics and Statistics si ha una crescita in tutti gli anni accademici considerati. Le variazioni relative positive indicano un interesse sostenuto degli studenti maschi in queste discipline.

In Information and Communication Technologies (ICTs) si ha un'oscillazione, infatti, si ha un aumento del 5.81% nell'anno accademico 2021/2022 che indica un crescente interesse degli studenti maschi in quest'area, sebbene sia seguito da una diminuzione del -2.48% nel settore dell'Information and Communication Technologies (ICTs).

Le variazioni relative divergenti tra le discipline STEM suggeriscono che ci sono diversi fattori in gioco e questi possono includere cambiamenti nei programmi di studio, nuove opportunità lavorative e cambiamenti nelle percezioni sociali delle diverse discipline.

<b>Variazioni Assolute - Iscritti Femmine</b>	<b>Natural sciences, Mathematics and Statistics</b>	<b>Information and Communication Technologies (ICTs)</b>	<b>Engineering, Manufacturing and Construction</b>
a.a. 21-22/ a.a. 20-21	1.358	636	329
a.a. 20-21/ a.a. 19-20	4.456	498	1.014
a.a. 19-20/ a.a. 18-19	2.117	448	858
a.a. 18-19/ a.a. 17-18	2.770	281	-56
a.a. 17-18/ a.a. 16-17	3.463	291	-158
a.a. 16-17/ a.a. 15-16	2.175	380	-181
a.a. 15-16/ a.a. 14-15	845	98	-1.348
a.a. 14-15/ a.a. 13-14	-2.262	60	-725
a.a. 13-14/ a.a. 12-13	-3.160	-19	-1.563

*Tabella 2.5: Variazioni assolute delle femmine iscritte nelle discipline STEM dall'a.a. 2012/2013 all'a.a. 2021/2022.*

<b>Variazioni Relative - Iscritti Femmine</b>	<b>Natural sciences, Mathematics and Statistics</b>	<b>Information and Communication Technologies (ICTs)</b>	<b>Engineering, Manufacturing and Construction</b>
a.a. 21-22/ a.a. 20-21	1,50%	12,06%	0,39%
a.a. 20-21/ a.a. 19-20	5,18%	10,43%	1,21%
a.a. 19-20/ a.a. 18-19	2,52%	10,35%	1,03%
a.a. 18-19/ a.a. 17-18	3,42%	6,94%	-0,07%
a.a. 17-18/ a.a. 16-17	4,46%	7,75%	-0,19%
a.a. 16-17/ a.a. 15-16	2,88%	11,26%	-0,22%
a.a. 15-16/ a.a. 14-15	1,13%	2,99%	-1,59%
a.a. 14-15/ a.a. 13-14	-2,94%	1,86%	-0,85%
a.a. 13-14/ a.a. 12-13	-3,95%	-0,59%	-1,80%

*Tabella 2.6: Variazioni relative delle femmine iscritte nelle discipline STEM dall'a.a. 2012/2013 all'a.a. 2021/2022.*



<b>Variazioni Assolute - Iscritti Maschi</b>	<b>Natural sciences, Mathematics and Statistics</b>	<b>Information and Communication Technologies (ICTs)</b>	<b>Engineering, Manufacturing and Construction</b>
a.a. 21-22/ a.a. 20-21	675	1.876	-5.282
a.a. 20-21/ a.a. 19-20	1.936	1.855	3.125
a.a. 19-20/ a.a. 18-19	2.112	1.714	3.505
a.a. 18-19/ a.a. 17-18	1.868	1.625	1.735
a.a. 17-18/ a.a. 16-17	2.467	2.132	2.605
a.a. 16-17/ a.a. 15-16	2.524	1.793	3.181
a.a. 15-16/ a.a. 14-15	1.441	1.213	-120
a.a. 14-15/ a.a. 13-14	-74	759	-2.607
a.a. 13-14/ a.a. 12-13	-1.085	73	-1.441

*Tabella 2.7: Variazioni assolute dei maschi iscritti nelle discipline STEM dall'a.a. 2012/2013 all'a.a. 2021/2022.*

<b>Variazioni Relative - Iscritti Maschi</b>	<b>Natural sciences, Mathematics and Statistics</b>	<b>Information and Communication Technologies (ICTs)</b>	<b>Engineering, Manufacturing and Construction</b>
a.a. 21-22/ a.a. 20-21	1,01%	5,81%	-2,48%
a.a. 20-21/ a.a. 19-20	2,97%	6,10%	1,49%
a.a. 19-20/ a.a. 18-19	3,35%	5,97%	1,70%
a.a. 18-19/ a.a. 17-18	3,06%	6,00%	0,85%
a.a. 17-18/ a.a. 16-17	4,21%	8,54%	1,29%
a.a. 16-17/ a.a. 15-16	4,50%	7,74%	1,60%
a.a. 15-16/ a.a. 14-15	2,63%	5,53%	-0,06%
a.a. 14-15/a.a. 13-14	-0,14%	3,58%	-1,30%
a.a. 13-14/ a.a. 12-13	-1,94%	0,35%	-0,71%

*Tabella 2.8: Variazioni relative dei maschi iscritti nelle discipline STEM dall'a.a. 2012/2013 all'a.a. 2021/2022.*

Un'ulteriore analisi è quella dei tassi (tabella 2.9) che conferma una netta e costante sottorappresentazione delle donne nei campi STEM per gli anni accademici presi in considerazione.

Tasso	2012/2013	2013/2014	2014/2015	2015/2016	2016/2017
Femmine	17,35%	17,33%	17,33%	17,56%	17,75%
Maschi	37,63%	37,75%	37,83%	38,20%	38,49%
Totale	26,09%	26,20%	26,28%	26,65%	26,96%
Tasso	2017/2018	2018/2019	2019/2020	2020/2021	2021/2022
Femmine	17,91%	17,94%	17,81%	17,55%	17,77%
Maschi	38,65%	38,82%	38,94%	38,47%	38,97%
Totale	27,16%	27,25%	27,19%	26,77%	27,00%

Tabella 2.9: Tasso di iscrizione per genere nelle discipline STEM dall'a.a. 2012/2013 all'a.a. 2021/2022.

Di seguito viene riportato il grafico (figura 2.5) che mostra l'andamento delle iscrizioni nelle discipline STEM suddivise per genere dall'a.a. 2012/2013 all'a.a. 2021/2022.

Si può osservare un andamento in decrescita per i maschi a partire dall'a.a. 2012/2013 fino circa all'a.a. 2014/2015 per poi crescere fino all'a.a. 2020/2021 e successivamente svilupparsi in direzione opposta nell'anno accademico seguente.

Un andamento simile si può notare tra le iscritte femmine, infatti, inizialmente c'è un diminuire di iscritti fino all'a.a. 2015/2016 poi però, a differenza degli iscritti maschi nelle discipline STEM, c'è una costante crescita.

La diminuzione di iscritti relativa al triennio 2012-2014 si può condurre al periodo di crisi economica che ha afflitto l'Italia dal 2008 al 2014, e questo può portare ad ipotizzare che la situazione finanziaria precaria abbia esercitato un notevole impatto sulla propensione delle persone a iscriversi all'università.

La recessione economica, infatti, potrebbe aver creato una serie di ostacoli finanziari rendendo più difficile per le famiglie sostenere i costi dell'istruzione universitaria.

Questi costi includono non solo le tasse universitarie, ma anche spese correlate come alloggio, trasporti e materiale didattico.

Di conseguenza, molte famiglie potrebbero aver dovuto rivalutare le loro priorità finanziarie, comprese quelle legate all'educazione dei loro figli.

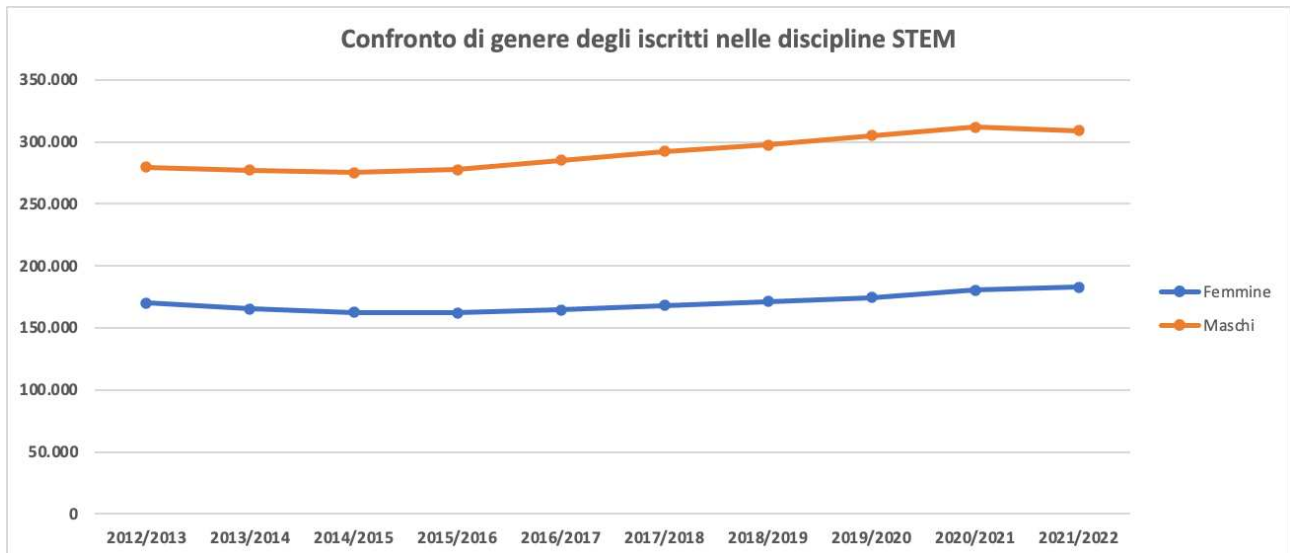


Figura 2.5: Andamento delle iscrizioni nelle discipline STEM suddivise per genere dall'a.a. 2012/2013 all'a.a. 2021/2022.

Dopo aver svolto le analisi esplorative si prosegue analizzando la serie storica proponendo anche delle previsioni per i prossimi cinque anni.

L'obiettivo è quello di individuare un modello probabilistico per ogni serie storica annuale analizzata che possa descrivere l'evoluzione del fenomeno e che permetta di utilizzarlo per scopi previsivi.

Ci si pone la domanda se la singola serie analizzata presenti ciclicità o trend e per poter rispondere è necessario controllare se il processo stocastico è stazionario.

Un processo stocastico si dice stazionario se la media e la varianza non mostrano variazioni sistematiche e se la sua dinamica non presenta cambiamenti strettamente periodici. Questo implica che i dati mantengono una consistenza del comportamento nel tempo senza mostrare trend.

Un processo stocastico può essere stazionario in senso stretto o debolmente stazionario. È stazionario in senso stretto se le distribuzioni congiunte di  $(Y_{t_1}, \dots, Y_{t_n})$  e di  $(Y_{t_1+\tau}, \dots, Y_{t_n+\tau})$  per ogni insieme di istanti  $t_1, \dots, t_n$  e per ogni  $\tau$  sono uguali mentre è debolmente stazionario o stazionario del secondo ordine se la sua media è costante e la sua funzione di autocovarianza dipende solo da  $k$ .

Esiste anche il caso in cui è ragionevole ipotizzare un processo generatore non stazionario e in alcuni casi è possibile eliminare la non stazionarietà mediante differenziazione.

Il processo che diventa stazionario a seguito di  $d$  differenziazioni successive si dice che è un processo non stazionario omogeneo di grado  $d$ .

Quando non è possibile ottenere stazionarietà attraverso la differenziazione, come nel caso in cui la serie presenti stazionarietà nella media ma non nella varianza, è necessario ricorrere ad altre trasformazioni.

È possibile controllare la stazionarietà di un processo stocastico attraverso l'osservazione dell'autocorrelazione (ACF) e dell'autocorrelazione parziale (PACF). Queste metriche misurano la dipendenza lineare tra le osservazioni attuali  $Y_t$  e le osservazioni passate  $Y_{t-k}$ .

Un'ulteriore verifica viene eseguita attraverso il test di stazionarietà KPSS (Kwiatkowski, Phillips, Schmidt e Shin). In questo test, l'ipotesi nulla è che la serie storica sia stazionaria, se il p-value del test è maggiore di un certo livello di significatività, per questo elaborato  $\alpha = 0.05$ , si può concludere che la serie è stazionaria intorno a una tendenza.

Viene per prima analizzata la variabile Natural Sciences, Mathematics and Statistics.

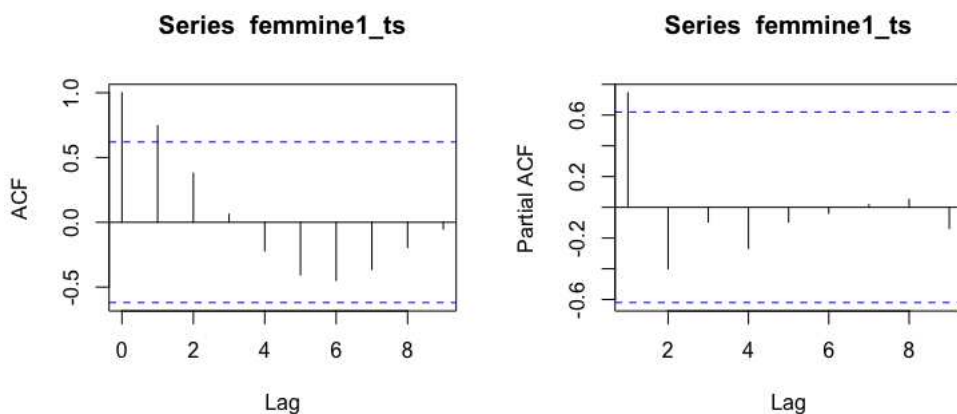


Figura 2.6: Correlogrammi relativi a ACF e PACF per la variabile Natural Sciences, Mathematics and Statistics per il genere femminile.

Dai correlogrammi delle iscritte femmine per la variabile in esame (figura 2.6) la serie sembra essere stazionaria dato che ACF e PACF decrescono in tempi ravvicinati all'aumentare di  $k$ . Si può notare, inoltre, che nell'ACF escono dalle bande di riferimento due valori corrispondenti ai lag in posizione 0 e 1 mentre nella PACF esce solamente il valore lag0.

Per confermare l'ipotesi di stazionarietà si esegue il test KPSS che riporta un p-value di 0.087 che risulta essere maggiore rispetto al livello di significatività fissato ed è per questo che non si può rifiutare l'ipotesi nulla e la serie risulta stazionaria.

Si passa ora ad adattare un modello ARIMA attraverso il criterio di selezione Akaike e il criterio di parsimonia. Il criterio di Akaike (AIC) si basa sulla log-verosimiglianza del modello e il modello con l'AIC più basso è preferibile perché indica una migliore adattabilità dei dati. Invece, il criterio di parsimonia sostiene che, quando si confrontano modelli che spiegano adeguatamente i dati, è preferibile scegliere il modello più semplice.

Il modello risultante è un ARIMA (0,1,0) con drift, stimato 1306.889, ovvero una componente lineare positiva che indica una crescita costante nel tempo della variabile presa in considerazione.

Il processo stimato è un processo Random Walk con drift:

$$Y_t = \mu + Y_{t-1} + \varepsilon_t \quad \text{con } Y_0 = c$$

$$Y_t = c + \mu t + \sum_{j=1}^t \varepsilon_j$$

Dove:

- $Y_t$  è il valore della serie temporale al tempo  $t$ ;
- $c$  è l'intercetta che rappresenta il valore iniziale della serie temporale;
- $\mu$  è il coefficiente di drift che rappresenta la crescita o la decrescita costante nel tempo;
- $t$  è il tempo;
- $\varepsilon_t$  rappresenta l'errore residuo al tempo  $t$  ed è un processo white noise di media 0 e varianza  $\sigma_\varepsilon^2$ .

Per verificare se il modello probabilistico scelto sia quello migliore si guarda se nel grafico della ACF e della PACF dei residui (figura 2.7) i ritardi sono tutti o quasi contenuti e in questo caso tale vincolo è stato rispettato.

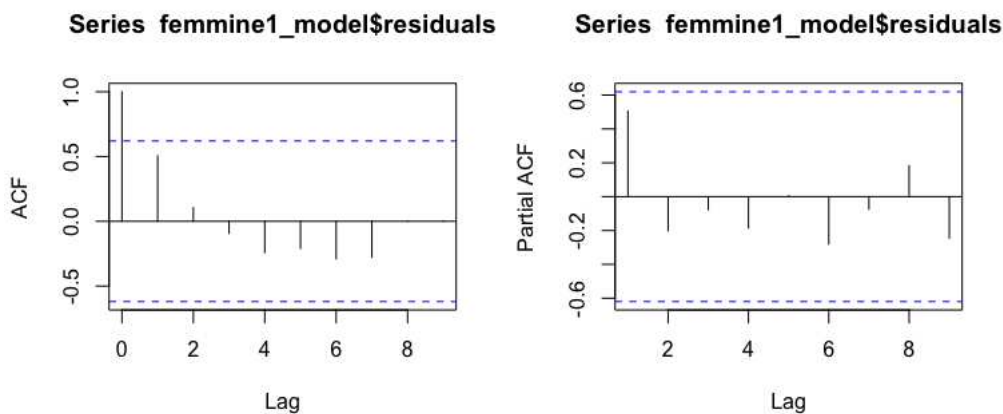


Figura 2.7: Correlogrammi relativi a ACF e PACF dei residui per la variabile *Natural Sciences, Mathematics and Statistics* per il genere femminile.

Si è ottenuta poi una previsione con un orizzonte di previsione quinquennale e un livello di confidenza del 95%.

Dalla tabella (tabella 2.10) e dal grafico (figura 2.8) presentati, è evidente una crescita costante nelle iscrizioni femminili nel settore delle Scienze Naturali, Matematica e Statistica. Questo significa che nel corso del tempo, il numero di donne che si iscrivono a corsi in queste discipline è in aumento regolare e continuativo.

	Point Forecast	Lower	Higher
2022	93.101,89	88.151,84	98.051,94
2023	94.408,78	87.408,35	101.409,21
2024	95.715,67	87.141,93	104.289,40
2025	97.022,56	87.122,46	106.922,66
2026	98.329,44	87.260,80	109.398,09

Tabella 2.10: Previsione per i prossimi cinque anni per la variabile Natural Sciences, Mathematics and Statistics per il genere femminile.

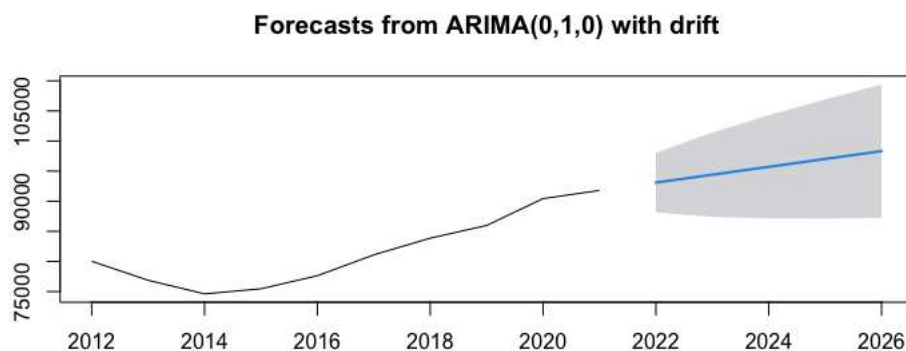


Figura 2.8: Previsione per i prossimi cinque anni per la variabile Natural Sciences, Mathematics and Statistics per il genere femminile.

Si considera ora la serie per gli iscritti di sesso maschile.

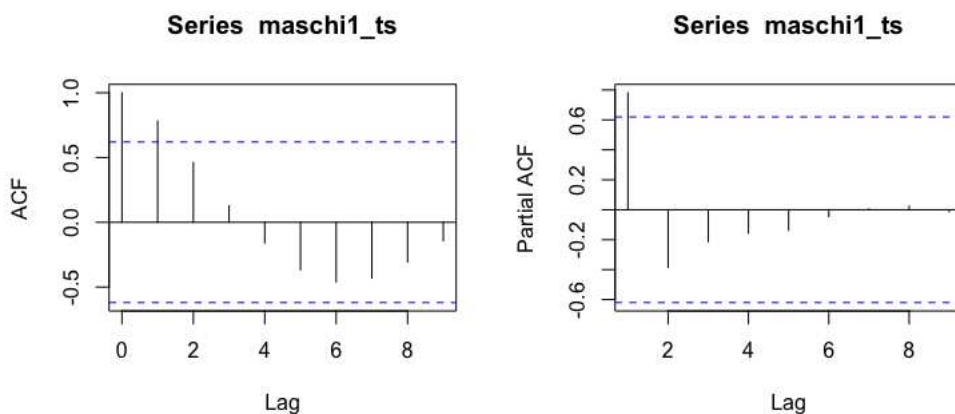


Figura 2.9: Correlogrammi relativi a ACF e PACF per la variabile Natural Sciences, Mathematics and Statistics per il genere maschile.

Entrambi i correlogrammi (figura 2.9) mostrano un andamento simile o addirittura quasi identico a quello precedente; dunque, la serie sembra essere stazionaria e dalle bande di riferimento escono soltanto i lag in posizione 0 e 1 per l'ACF e in posizione 0 per la PACF.

Si esegue il test KPSS, che riporta un p-value di 0.068, maggiore del livello di significatività fissato, che conferma la stazionarietà della serie.

Viene stimato il modello probabilistico e quello migliore, selezionato guardando l'AIC e tenendo conto del principio di parsimonia, risulta un ARIMA (1,1,0). Un'ulteriore conferma si verifica poiché i ritardi nei grafici ACF e PACF dei residui (figura 2.10) sono prevalentemente contenuti entro limiti accettabili.

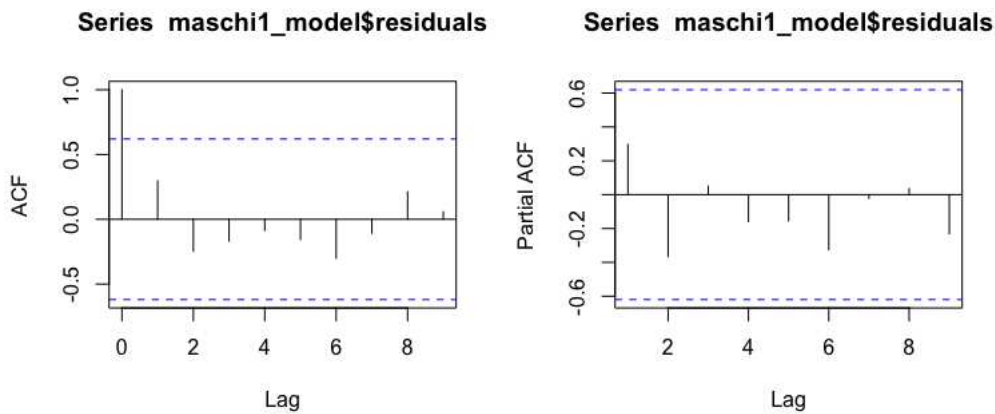


Figura 2.10: Correlogrammi relativi a ACF e PACF dei residui per la variabile Natural Sciences, Mathematics and Statistics per il genere maschile.

Facendo una previsione per i prossimi cinque anni emerge che anche la serie di iscrizioni maschili per l'area disciplinare presa in esame cresce in modo costante e stabile nel tempo.

	Point Forecast	Lower	Higher
2022	68.274,96	66.514,90	70.035,02
2023	68.742,81	65.068,46	72.417,17
2024	69.132,32	63.365,03	74.899,60
2025	69.456,59	61.514,09	77.399,10
2026	69.726,56	59.584,81	79.868,31

Tabella 2.11: Previsione per i prossimi cinque anni per la variabile Natural Sciences, Mathematics and Statistics per il genere maschile.



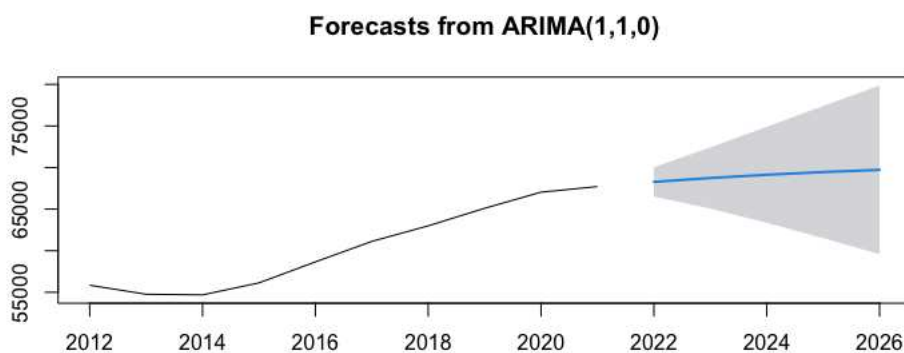


Figura 2.11: Previsione per i prossimi cinque anni per la variabile Natural Sciences, Mathematics and Statistics per il genere maschile.

La seconda variabile presa in esame è Information and Communication Technologies (ICTs).

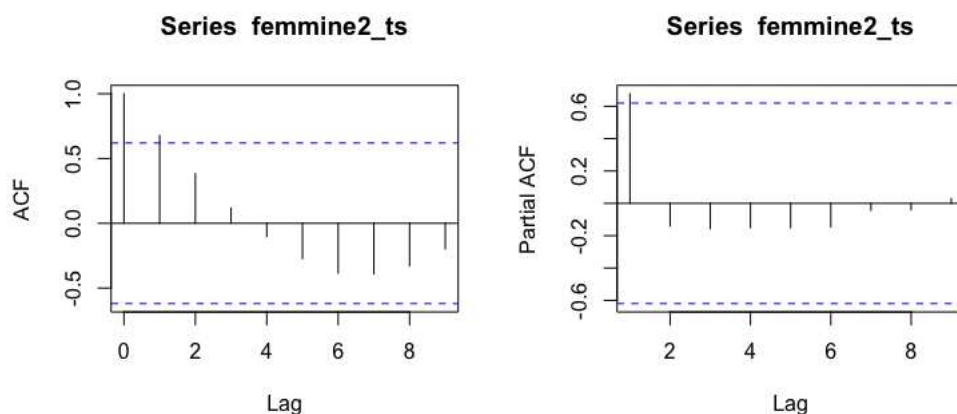


Figura 2.12: Correlogrammi relativi a ACF e PACF per la variabile Information and Communication Technologies (ICTs) per il genere femminile.

Dall'analisi dei correlogrammi delle iscrizioni femminili per la variabile in esame (figura 2.12), emerge un quadro che suggerisce la stazionarietà della serie. Questo si evince dal fatto che nell'ACF due valori escono dai limiti di riferimento, lag0 e lag1, mentre nella PACF esce solamente il valore lag0.

Per consolidare l'ipotesi di stazionarietà, è stato condotto il test KPSS che restituisce un p-value pari a 0.061, valore che indica che non vi è sufficiente evidenza per respingere l'ipotesi nulla. Di conseguenza, si può affermare che la serie è stazionaria.

Viene adattato il modello per la serie storica con un ARIMA (0,1,0) con drift, ovvero un Random Walk con drift stimato 297.

A confermare che il modello selezionato è quello migliore sono i grafici delle funzioni di autocorrelazione (ACF) e autocorrelazione parziale (PACF) dei residui (figura 2.13) nei quali i ritardi sono quasi tutti compresi all'interno delle bande di confidenza.

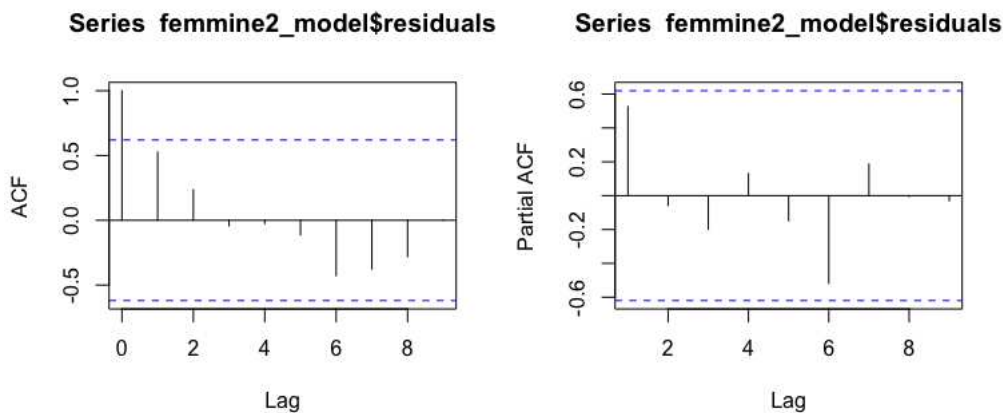


Figura 2.13: Correlogrammi relativi a ACF e PACF dei residui per la variabile Information and Communication Technologies (ICTs) per il genere femminile.

Viene poi fatta una previsione a breve medio termine, di cinque anni, con livello di significatività fissato del 95% per le donne iscritte nell'ambito Tecnologie dell'informazione e della comunicazione.

Spicca subito, dalla tabella (tabella 2.12) e dal grafico (figura 2.14), che si è verificato un incremento costante nel numero di donne nel campo di studio considerato. Questo indica che, nonostante il numero di iscrizioni femminili in questa area sia attualmente limitato, c'è una tendenza in aumento nel tempo.

	Point Forecast	Lower	Higher
2022	6.207	5.779,31	6.634,69
2023	6.504	5.899,15	7.108,85
2024	6.801	6.060,22	7.541,78
2025	7.098	6.242,62	7.953,38
2026	7.395	6.438,65	8.351,35

Tabella 2.12: Previsione per i prossimi cinque anni per la variabile Information and Communication Technologies (ICTs) per il genere femminile.

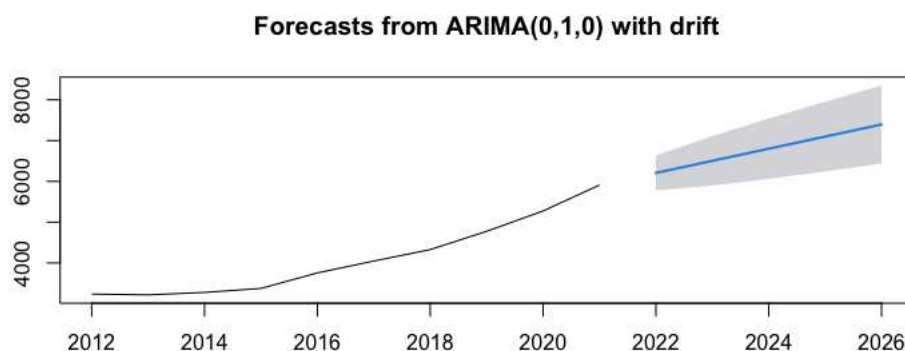


Figura 2.14: Previsione per i prossimi cinque anni per la variabile Information and Communication Technologies (ICTs) per il genere femminile.

Si prosegue analizzando la serie per le iscrizioni maschili.

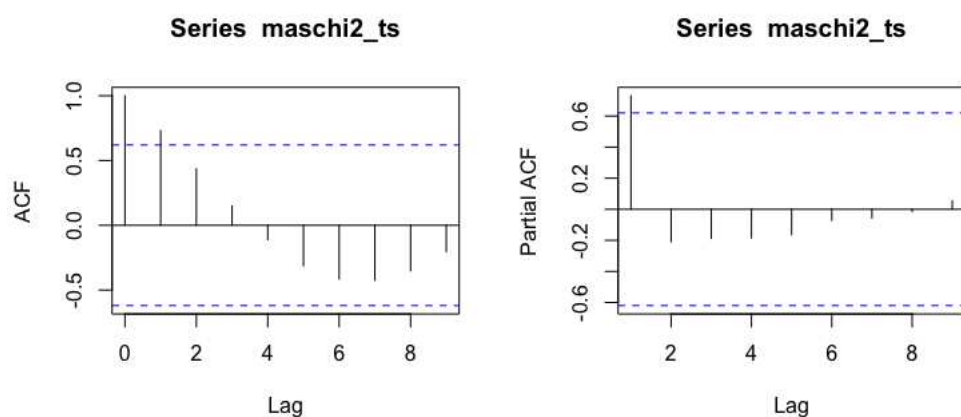


Figura 2.15: Correlogrammi relativi a ACF e PACF per la variabile Information and Communication Technologies (ICTs) per il genere maschile.

Si osservano i grafici dell'autocorrelazione e dell'autocorrelazione parziale (figura 2.15) e la serie degli iscritti maschi sembra presentare un andamento stazionario e l'unico valore che fuoriesce dalle bande di riferimento è il lag in posizione  $k=0$ .

Conferma l'ipotesi di stazionarietà il test KPSS con p-value di 0.058.

Anche in questo caso il modello probabilistico adottato è un ARIMA (0,1,0) con drift, stimato 1448.889. Risulta il modello migliore perché, oltre a utilizzare il criterio di selezione di Akaike e il criterio di parsimonia, guardando i grafici dell'ACF dell'PACF dei residui (figura 2.16) tutti i ritardi, a parte nell'ACF il valore lag0, sono compresi tra le bande di confidenza.

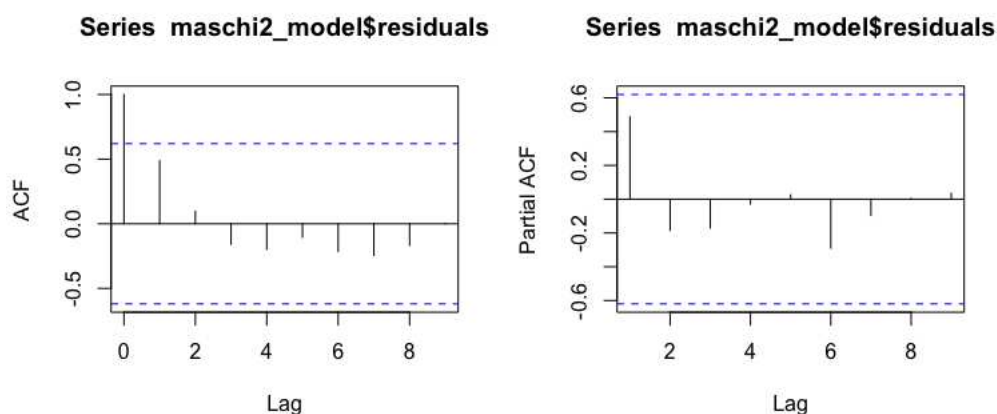


Figura 2.16: Correlogrammi relativi a ACF e PACF dei residui per la variabile Information and Communication Technologies (ICTs) per il genere maschile.

Successivamente si sono ottenute le previsioni per gli anni dal 2022 al 2026 ad un livello di confidenza del 95%. Analogamente a ciò che è stato osservato nella serie riguardante le iscrizioni femminili nell'area Tecnologie dell'informazione e della comunicazione, anche per questa specifica categoria si prospetta un incremento costante nel numero di iscritti nei prossimi cinque anni. Tale andamento positivo suggerisce un aumento graduale e continuativo degli uomini che si iscriveranno nel corso del tempo.

	Point Forecast	Lower	Higher
2022	35.605,89	34.317,16	36.894,62
2023	37.054,78	35.232,24	38.877,32
2024	38.503,67	36.271,52	40.735,82
2025	39.952,56	37.375,09	42.530,02
2026	41.401,44	38.519,75	44.283,14

Tabella 2.13: Previsione per i prossimi cinque anni per la variabile Information and Communication Technologies (ICTs) per il genere maschile.

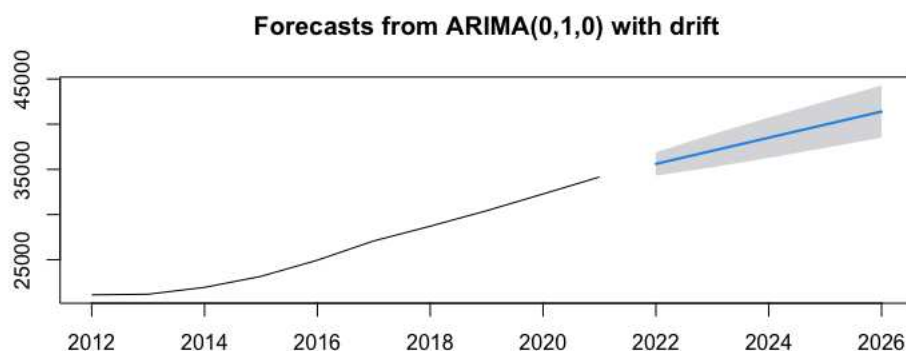


Figura 2.17: Previsione per i prossimi cinque anni per la variabile Information and Communication Technologies (ICTs) per il genere maschile.

Viene analizzata l'ultima variabile Engineering, Manufacturing and Construction.

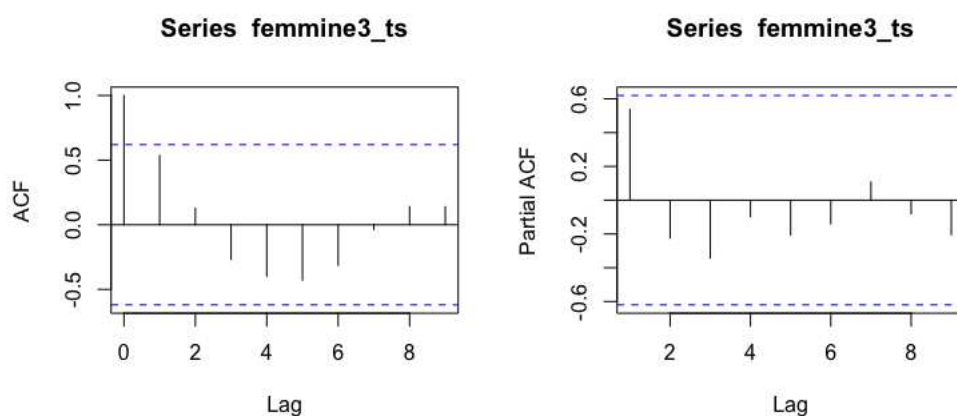


Figura 2.18: Correlogrammi relativi a ACF e PACF per la variabile Engineering, Manufacturing and Construction per il genere femminile.

Analizzando attentamente i correlogrammi delle funzioni di autocorrelazione e autocorrelazione parziale (figura 2.18), si nota che solo il valore associato al ritardo  $k = 0$  si discosta dai limiti di confidenza predefiniti.

Per determinare la stazionarietà della serie, è stato condotto il test KPSS. Il p-value ottenuto da tale test è pari a 0.1, che supera il livello di significatività stabilito di  $\alpha = 0.05$ . Di conseguenza, si può affermare che la serie è stazionaria.

L'analisi dettagliata dei dati ha portato all'identificazione del modello ARIMA (1,0,1) come la migliore rappresentazione di questa serie storica. Per confermare la correttezza di questa scelta, è stato eseguito un esame dei correlogrammi dell'ACF della PACF dei residui del modello (figura 2.19). Si nota che, in entrambi i grafici, tutti i ritardi, ad eccezione del valore associato al ritardo  $k = 0$  nel caso dell'ACF, rientrano all'interno delle bande di confidenza. Questo risultato conferma che il modello selezionato interpreta in modo adeguato la serie temporale presa in esame.

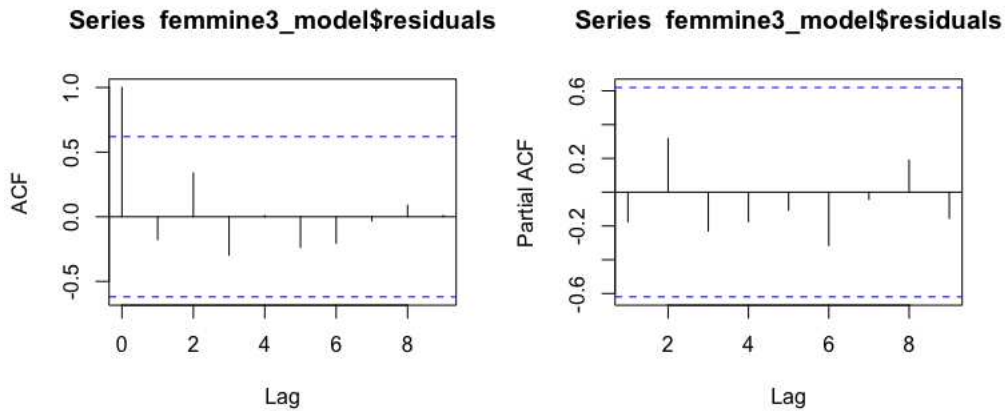


Figura 2.19: Correlogrammi relativi a ACF e PACF dei residui per la variabile Engineering, Manufacturing and Construction) per il genere femminile.

Effettuando una previsione per il prossimo quinquennio relativamente alle iscrizioni femminili nel campo dell'Ingegneria, Produzione e Costruzione, emerge un leggero declino. Questo andamento, se proiettato nel futuro, suggerisce una riduzione graduale nel numero di donne che scelgono di iscriversi a corsi di queste discipline. Tale tendenza potrebbe essere influenzata da vari fattori socioculturali ed economici e rappresenta un aspetto importante da monitorare.

	Point Forecast	Lower	Higher
2022	85.177,87	83.510,10	86.845,63
2023	85.168,74	82.211,68	88.125,80
2024	85.159,61	81.324,81	88.994,41
2025	85.150,48	80.604,51	89.696,45
2026	85.141,35	79.981,43	90.301,28

Tabella 2.14: Previsione per i prossimi cinque anni per la variabile Engineering, Manufacturing and Construction per il genere femminile.

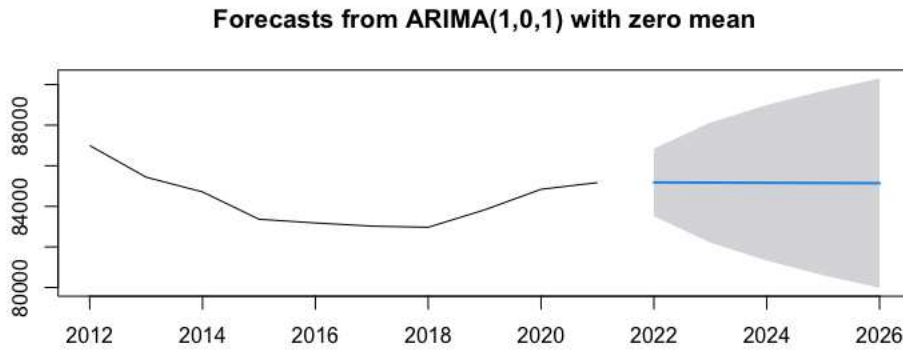


Figura 2.20: Previsione per i prossimi cinque anni per la variabile Engineering, Manufacturing and Construction per il genere femminile.

Si conclude analizzando la serie delle iscrizioni maschili.

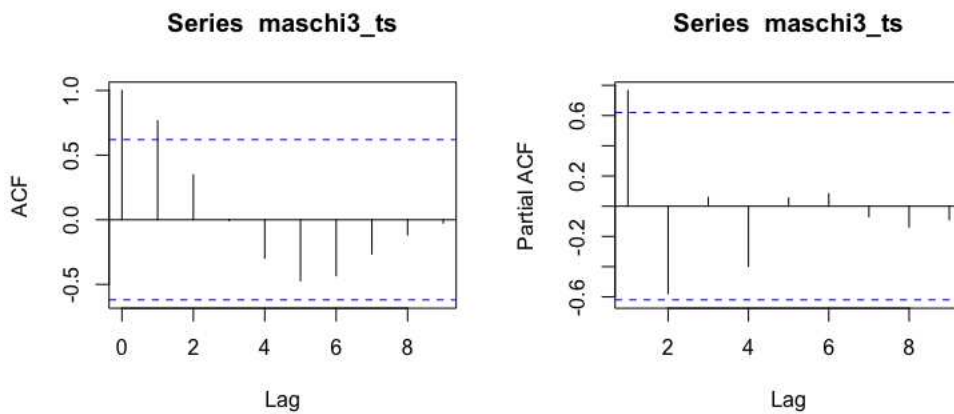


Figura 2.21: Correlogrammi relativi a ACF e PACF per la variabile Engineering, Manufacturing and Construction per il genere maschile.

Dai correlogrammi degli iscritti maschi per la variabile presa in considerazione (figura 2.21), si nota che nel grafico dell'ACF, i valori per i lag in posizione 0 e 1 escono dalle bande di riferimento, mentre nella PACF esce solo il valore per il lag 0.

Si esegue il test KPSS per capire se la serie storica è stazionaria, il cui risultato ha restituito un p-value pari a 0.1. Dato che questo valore è maggiore del livello di significatività predefinito, non si può rifiutare l'ipotesi nulla; perciò, la serie viene considerata stazionaria.

È stato individuato il modello ARIMA (0,1,0) come la migliore rappresentazione per questa serie storica. Il processo risultante può essere descritto come un Random Walk:

$$Y_t = Y_{t-1} + \varepsilon_t \quad \text{con } Y_0 = \mu$$

$$Y_t = \mu + \sum_{i=1}^t \varepsilon_i$$

Dove:

- $Y_t$  è il valore della serie temporale al tempo  $t$ ;
- $\mu$  è l'intercetta che rappresenta il valore iniziale della serie temporale;
- $t$  è il tempo;
- $\varepsilon_t$  rappresenta l'errore residuo al tempo  $t$  ed è un processo white noise di media 0 e varianza  $\sigma_\varepsilon^2$ .

Per confermare la scelta di questo modello, sono stati esaminati i correlogrammi delle funzioni di autocorrelazione (ACF) e autocorrelazione parziale (PACF) dei residui del modello (figura 2.22). Emerge che solo nel grafico dell'ACF il ritardo in posizione  $k = 0$  esce dalle bande di confidenza. Questo risultato suggerisce che il modello selezionato rappresenta adeguatamente la serie temporale in questione.

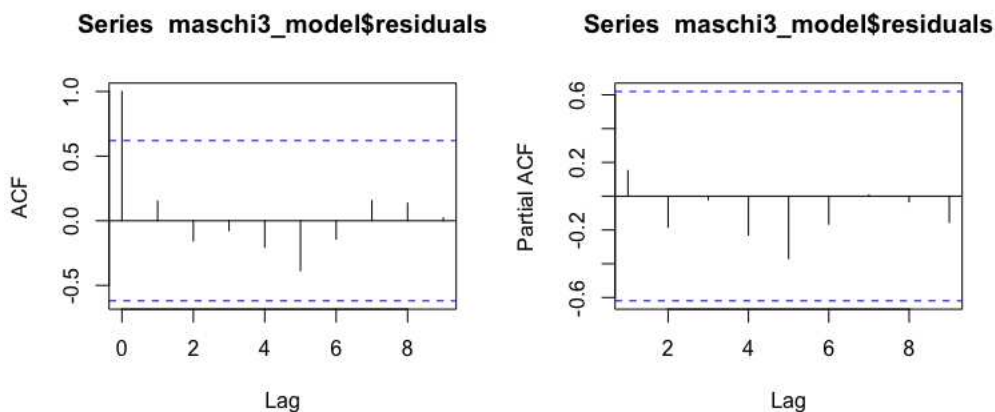


Figura 2.22: Correlogrammi relativi a ACF e PACF dei residui per la variabile Engineering, Manufacturing and Construction) per il genere maschile.

Nell'analizzare attentamente i dati relativi agli iscritti maschi nel campo dell'Ingegneria, Produzione e Costruzione, proiettati per i prossimi cinque anni, emerge un quadro particolare. Esaminando la tabella (tabella 2.15) e il grafico (figura 2.23) forniti qui di seguito, si nota che il numero di iscritti in questo settore rimane costante nel corso del tempo, senza variazioni significative. Questo implica che, almeno secondo i dati raccolti finora e le previsioni a breve medio termine, la partecipazione degli studenti maschi in questo ambito specifico di studi non cambierà in modo rilevante.



	Point Forecast	Lower	Higher
2022	207.310	201.510,7	213.109,3
2023	207.310	199.108,6	215.511,4
2024	207.310	197.265,4	217.354,6
2025	207.310	195.711,5	218.908,5
2026	207.310	194.342,5	220.277,5

Tabella 2.15: Previsione per i prossimi cinque anni per la variabile Engineering, Manufacturing and Construction per il genere maschile.

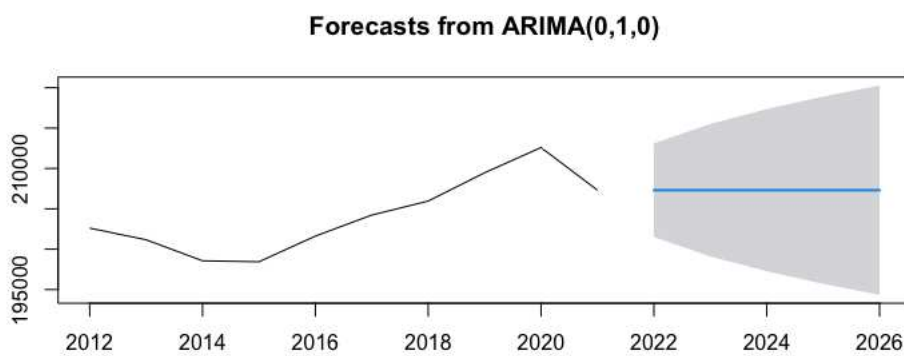


Figura 2.23: Previsione per i prossimi cinque anni per la variabile Engineering, Manufacturing and Construction per il genere maschile.



## Conclusioni

L'analisi dettagliata condotta sulle disparità di genere nelle professioni STEM nel corso dell'ultimo decennio ha fornito una visione chiara e approfondita della situazione attuale. I dati e le tendenze esaminate rivelano una realtà complessa e persistente in cui le donne continuano a essere sottorappresentate in modo significativo in queste importanti discipline accademiche e professionali. Tuttavia, nonostante le sfide evidenziate, emergono anche segnali di progresso e una crescente consapevolezza dell'importanza di affrontare questa disparità.

In particolare, durante l'analisi dei dati, è emerso un costante aumento delle iscrizioni femminili nei corsi di Scienze Naturali, Matematica e Statistica, segnalando un crescente interesse delle donne in queste discipline. Le previsioni per il futuro indicano una continuazione di questa tendenza positiva sia per le donne che per gli uomini, promuovendo l'equità di genere nell'istruzione superiore.

Nel settore delle Tecnologie dell'informazione e della comunicazione, invece, si è notato un aumento costante delle iscrizioni femminili nel breve-medio termine, suggerendo un interesse crescente, seppur limitato al momento.

Tuttavia, nell'ambito dell'Ingegneria, Produzione e Costruzione, è previsto un lieve calo nelle iscrizioni femminili, indicando una graduale riduzione della partecipazione delle donne. Questa tendenza potrebbe essere influenzata da vari fattori socioculturali ed economici, richiedendo un monitoraggio costante nel tempo. Al contrario, non sono previste variazioni significative nelle iscrizioni maschili, indicando una stabilità nella partecipazione degli uomini nel breve-medio termine in questo settore di studi.

Esistono diverse strategie che possono essere adottate per ridurre il divario di genere nelle discipline STEM. In primo luogo, è cruciale incoraggiare le ragazze sin dalla giovane età ad interessarsi alle materie STEM, al fine di sfatare gli stereotipi di genere associati e fornire loro modelli di riferimento positivi. Inoltre, è fondamentale promuovere la visibilità di figure femminili di successo nelle STEM, sia attraverso l'assunzione che attraverso le promozioni all'interno delle organizzazioni, al fine di ispirare altre donne e ragazze ad intraprendere carriere in questi settori.

I benefici derivanti dalla riduzione del divario di genere nelle STEM sono molteplici. In primo luogo, favorisce la diversità e l'inclusione, contribuendo a stimolare soluzioni innovative per le sfide sociali. Inoltre, aiuta ad affrontare il divario di competenze nella forza lavoro STEM, promuovendo la crescita economica e la creazione di nuovi posti di lavoro. L'aumento della diversità di genere nei campi STEM può anche migliorare la qualità della ricerca, poiché le prospettive diverse portano a un maggiore livello di creatività e intuizione nelle ricerche scientifiche.

Inoltre, una maggiore diversità di genere nelle STEM può portare allo sviluppo di prodotti e servizi più adatti alle esigenze di tutti i consumatori, migliorando così la qualità complessiva dei prodotti offerti sul mercato.

Infine, ridurre il divario di genere nelle STEM contribuisce al progresso sociale, contribuendo a creare una società più giusta ed equa per tutti.

Per affrontare questa sfida complessa è essenziale adottare un approccio che vada dall'incoraggiare le ragazze sin dall'infanzia, fino alla creazione di incentivi per le donne per intraprendere e mantenere una carriera nelle STEM.

Perciò con sforzi collettivi e un impegno continuo si può rispondere alle sfide globali in modo più efficace ed equo.

# Appendice A

## Formule Excel

Le variazioni assolute sono state calcolate sottraendo al valore di una cella di Excel di uno specifico anno accademico il valore della cella dell'anno precedente.

Le variazioni relative, invece, sono state calcolate dividendo il valore della variazione assoluta per la cella dell'anno precedente a quello di riferimento. Il risultato è stato poi espresso come percentuale con due cifre decimali, fornendo così un'indicazione chiara e precisa delle variazioni percentuali tra gli anni accademici consecutivi nelle discipline STEM.

Esempio:

	A	B	C	D
1		Iscritti femmine	Variazione assoluta	Variazioni relative
2	a.a. 2021/2022	91.795	1.358	1,50%
3	a.a. 2020/2021	90.437		

Variazione assoluta: “=B2-B3”

Variazione relativa: “=[(B2-B3) / B3] \*100”

I tassi sono stati calcolati dividendo il numero di iscritti con il numero della popolazione di riferimento. Il risultato è stato poi espresso come percentuale.

Esempio:

	A	B	C	D
1		Iscritti femmine	Popolazione	Tasso
2	a.a. 2021/2022	182.874	1.028.829	17,75%

Tasso: “=(B2/C2) \*100”

## Comandi RStudio

Di seguito vengono riportati i comandi utilizzati per analizzare le varie serie storiche:

- vengono richiamate le librerie “library(forecast)” e “library(tseries)”;
- viene trasformata la serie storica in serie storica annuale:  
`femmine1_ts <- ts(femmine1, start=2012, end=2021, frequency=1);`
- si controlla se la serie è stazionaria attraverso i grafici dell’ACF e PACF:  
`acf(femmine1_ts)`  
`pacf(femmine1_ts);`
- si conduce il test Kwiatkowsky-Phillips-Schimidt (KPSS) per capire se la serie è stazionaria:  
`kpss.test(femmine1_ts)`  
Se il valore del p-value risulta maggiore rispetto al livello di significatività fissato,  $\alpha = 0.05$ , non si può rifiutare l’ipotesi nulla di stazionarietà;
- si adatta il modello ARIMA:  
`femmine1_model=Arima(femmine1_ts, order=c(0,1,0),`  
`include.mean=T,include.drift=T)`  
`stat.mod(femmine1_model);`
- si controlla la stazionarietà del modello attraverso i grafici dei residui dell’ACF e della PACF:  
`acf(femmine1_model$residuals)`  
`pacf(femmine1_model$residuals)`  
Se nessun valore esce dalle bande il modello è stazionario;
- infine, si fa una previsione per i prossimi cinque anni:  
`previsioni_F1=forecast(femmine1_model, level = 0.95, h=5)`  
`plot(previsioni_F1);`

## Bibliografia e sitografia

- Di Fonzo T., Lisi F. (2005). *Serie storiche economiche: analisi statistiche e applicazioni*. Roma: Carocci Editore
- Zahidi S. (2023). *Global Gender Gap Report 2023*.  
<https://www.weforum.org/reports/global-gender-gap-report-2023/in-full>
- Mariani V, Montanaro P, e Soncin S. (2015). *Meno iscrizioni all'università per colpa della crisi*. <https://lavoce.info/archives/33494/meno-iscrizioni-alluniversita-per-colpa-crisi/>
- *Ministero dell'Istruzione e Merito*. <https://www.miur.gov.it/web/guest/home>
- *Ministero dell'Università e Ricerca*. <https://www.mur.gov.it/it>
- *Portale dei dati dell'istruzione superiore*. <https://ustat.mur.gov.it/>
- Rogers K. (2010). *The 100 Most Influential Scientists of all Time*. New York: Britannica Educational Publishing.
- Piloto C. (2023). *The Gender Gap in STEM: Still Gaping in 2023*.  
<https://professionalprograms.mit.edu/blog/leadership/the-gender-gap-in-stem/>
- Kahn S. e Ginther D. (2017). *Women and STEM*.  
[https://www.nber.org/system/files/working\\_papers/w23525/w23525.pdf](https://www.nber.org/system/files/working_papers/w23525/w23525.pdf)

