

Dipartimento di Filosofia, Sociologia, Pedagogia e Psicologia Applicata

Corso di laurea magistrale in

SCIENZE DELLA FORMAZIONE CONTINUA



**UNIVERSITÀ
DEGLI STUDI
DI PADOVA**

Tesi di Laurea Magistrale

*“Alfabetizzazione ai dati nel contesto dell’istruzione superiore:
un modello fattoriale per i fabbisogni formativi e professionali”*

Relatore: Prof.ssa Juliana Elisa Raffaghelli

Laureando: Dott. Flaviano Battistig

Matricola: 2005782

A.A. 2022/2023

Indice

1. Introduzione

1.1. [Scenari pre e post pandemici](#)

1.1.1. [La rivoluzione dei dati](#)

1.1.2. [Implicazioni della trasformazione digitale](#)

1.2. [Dalla *datafication* alla *platformization*, fino al *post-digitale*](#)

1.2.1. [Il modello di business delle piattaforme](#)

1.2.2. [Le “Big Tech” tra disillusione e post-digitale](#)

1.3. Dalla *datafication* alla *platformization*, fino al *post-digitale*: la digitalizzazione dell'educazione

1.3.1. [Le tecnologie digitali come risposta ai problemi dei sistemi educativi](#)

1.3.2. [L'istruzione tra digitale e post-digitale](#)

1.3.3. [Approcci tecnologici alla riforma delle università](#)

1.4. [Le definizioni di competenza digitale nel contesto internazionale e in Italia: un excursus](#)

1.4.1. [I precursori della Data Literacy: Statistical Literacy e Numeracy](#)

1.4.2. [Le competenze digitali in Europa](#)

1.4.3. [Le competenze digitali in Italia](#)

2. [Alfabetizzazione ai dati per il Faculty Development](#)

2.1. [Competenze emergenti nel contesto dell'istruzione superiore](#)

2.2. [L'alfabetizzazione ai dati dei docenti universitari](#)

3. Metodologia della ricerca

3.1. [Epistemologia, ontologia e assiologia della ricerca](#)

3.2. Metodi

3.2.1. [Il questionario](#)

3.2.2. [Caratteristiche del questionario](#)

3.3. Strumenti informatici

3.3.1. [L'ambiente di R](#)

3.3.2. [Una piattaforma "home-made"](#)

4. Risultati

4.1. [Statistiche descrittive univariate](#)

4.1.1. [Caratteristiche generali del campione](#)

4.2. Analisi statistiche inferenziali

4.2.1. [Test di affidabilità: il Cronbach's \$\alpha\$](#)

4.2.2. [Premesse per l'analisi della varianza ANOVA](#)

4.2.3. [ANOVA per gruppo "Genere"](#)

4.2.4. [ANOVA per gruppo "Età"](#)

4.2.5. [ANOVA per gruppo "Nazione"](#)

4.2.6. [ANOVA per gruppo "Area Disciplinare"](#)

4.2.7. [ANOVA per gruppo "Esperienza d'insegnamento"](#)

4.2.8. [ANOVA per gruppo "Esperienza di ricerca"](#)

4.3. Test multivariato di analisi della varianza MANOVA

4.3.1. [Premesse per il test MANOVA](#)

4.3.1.1. [Test per la presenza di outlier](#)

4.3.1.2. [Test di normalità Shapiro-Wilk](#)

4.3.1.3. [Test non-parametrico di Levene per l'omoschedasticità](#)

4.3.1.4. [Test di multicollinearità](#)

4.3.2. [Test multivariato di analisi della varianza: MANOVA](#)

4.3.2.1. [MANOVA per gruppo "Genere"](#)

4.3.2.2. [MANOVA per gruppo "Età"](#)

4.3.2.3. [MANOVA per gruppo "Nazione"](#)

4.3.2.4. [MANOVA per gruppo "Area Disciplinare"](#)

4.3.2.5. [MANOVA per gruppo "Esperienza d'insegnamento"](#)

4.3.2.6. [MANOVA per gruppo “Esperienza di ricerca”](#)

4.3.2.7. [Dimensione dell'effetto \(“Eta squared”\)](#)

4.3.2.8. [Test Post-Hoc di Games-Howell](#)

4.4. [Analisi fattoriale](#)

4.4.1. [L'Analisi Fattoriale Esplorativa \(AFE\) e Confermativa \(CFA\)](#)

5. [Discussione](#)

6. [Conclusioni](#)

7. [Bibliografia](#)

8. [Allegati](#)

8.1. Script di R per l'analisi dei dati

8.1.1. Script per l'analisi univariata e bivariata

8.1.2. Script per MANOVA

8.1.3. Script per l'Analisi Fattoriale

1. Introduzione

1.1. Scenari pre e post pandemici

1.1.1. La rivoluzione dei dati

La recente pandemia di Covid-19 ha, con tutta probabilità, accelerato processi già in atto nelle società avanzate. La sempre maggiore diffusione dell'Information Technology e di reti sempre più performanti in termini di velocità di scambio di dati ha profondamente modificato usi, costumi, abitudini di vita e lavoro in maniera sempre più massiccia e pervasiva. Da quella che Castells (1996, citato da Raffaghelli et al., 2020) definiva la “*network society*”, la “società della rete”, si è passati ad una “*datafied society*”, una società “datificata” in cui «*vaste quantità di dati digitali [...] stanno guidando nuove pratiche sociali*» (Raffaghelli et al., 2020, traduzione a cura dell'autore). Così, ad una crescente *digitalizzazione* di prodotti e servizi ha fatto seguito, a partire dalla seconda metà degli anni 2000, un processo di *datificazione* di ogni aspetto dell'esperienza umana.

Se la digitalizzazione viene definita da Williamson e Hogan (2010, p. 6) come «*l'uso di tecnologie programmate al computer per compiere attività che in precedenza sarebbero state eseguite attraverso altre forme analogiche o azioni umane, o ai nuovi compiti non possibili senza il calcolo elettronico*», la datificazione viene descritta dagli autori come «*l'utilizzo delle tecnologie digitali per registrare aspetti della vita umana, dei processi sociali ed economici, delle attività istituzionali, e così via, sotto forma di informazioni digitalizzate, spesso a fini di analisi, previsione e intervento di un qualche genere*» (ibidem, p. 6). Si deduce quindi che, tra i compiti in precedenza non possibili senza il calcolo computazionale, vi è anche questa massiccia raccolta e analisi di dati prodotti costantemente da ogni apparato digitale in possesso dell'utente, sia esso un personal computer, uno smartphone, ma anche una televisione “smart” o un “assistente personale digitale”.

Subito dopo aver esposto queste prime definizioni, è tuttavia imperativo chiarire il concetto di “dati”, in quanto di per sé prono a differenti interpretazioni. Raffaghelli (2017, p. 302), citando Christine Borgman, ci ricorda infatti che «*i dati non sono oggetti naturali, che esistono per sé. I dati [...] possono essere più facilmente definiti attraverso esempi variegati, includendo fatti, numeri, lettere, simboli. Quindi, sono solo le definizioni operative che mettono il dato in contesto e ne fanno emergere un senso trasversale per gli aspetti che tentiamo di ancorare attraverso una definizione*». La natura molteplice ed emergente dei dati, infatti, si evince dalla loro differenziazione per «*i gradi di elaborazione, le origini e la preservazione, l'appartenenza a collezioni*» (ivi, p. 303).

Nel contesto contemporaneo, i dati – o, più esattamente, i “*Big Data*” – si caratterizzano inoltre per altre proprietà, elencate dal “*Gartner IT dictionary*”: «*Alto volume, alta velocità, alta varietà di informazioni che compongono le masse digitali*» (ivi, p. 304). Per “alto volume” ci si riferisce al tracciamento continuo delle attività degli utenti che produce un’enorme massa di dati; con “alta velocità” ci si riferisce al fatto che la loro generazione e raccolta avviene in tempo reale; infine, per “alta varietà” si indica la molteplicità di applicazioni e tecnologie, in grado di produrre e trasmettere informazioni.

Può, infatti, trattarsi di dati estratti da piattaforme di *online collaboration* inerenti all’attività professionale di un individuo, o le informazioni sul suo stato di salute raccolte dal suo *smartwatch*, o anche la mappatura della sua abitazione eseguita da un’aspirapolvere robotico, o persino dati sulla tipologia di elettrodomestici posseduti e il loro consumo energetico. Queste ultime due, nello specifico, sono rese possibili dall’implementazione del cosiddetto “Internet delle cose” (*Internet-of-Things*), in cui intere abitazioni sono, a tutti gli effetti, in Rete, producendo una lunga serie di dati ad uso e consumo dell’utente – e non solo.

1.1.2. Implicazioni della trasformazione digitale

Questo capitolo introduttivo presenta il *background* che fa da sfondo alla ricerca e da cui si parte per affrontare il tema della trasformazione digitale nella società e nell’educazione.

Il presente paragrafo affronta, innanzitutto, una prima disamina del fenomeno della *datificazione*, approfondendo l’impatto del processo di *digitalizzazione* della società – i cui esordi si possono far risalire agli anni ’80 del secolo scorso – con particolare riferimento alle competenze professionali emergenti richieste dal nuovo scenario. Viene qui, inoltre, introdotto e discusso il tema delle Intelligenze Artificiali, elencandone potenzialità e criticità, nonché gli effetti sul mercato del lavoro e sulle professioni attualmente esistenti.

Il paragrafo seguente (1.2) discuterà il tema delle piattaforme e, nello specifico, le narrazioni ad esse correlate, il loro modello di business (1.2.1), la percezione di queste da parte dell’utente del post-digitale, e le questioni etiche collegate al fenomeno della “piattaformizzazione”; seguirà poi il paragrafo (1.3), con cui si argomenterà più nello specifico l’influenza del digitale nel contesto educativo. Infine, nel paragrafo conclusivo del capitolo (1.4), si tratterà la questione delle competenze digitali nel contesto europeo ed italiano.

Data la complessità dei fenomeni analizzati e discussi – digitalizzazione, piattaforma, post-digitale, per elencarne alcuni – caratterizzati da molteplici sfaccettature e ramificazioni, nonché da una certa pervasività ed ubiquità, si è scelta una struttura argomentativa che segmentasse,

nella maniera più razionale possibile, gli argomenti oggetto d'indagine. In questo modo, si cercherà, da un lato, di facilitare l'analisi e l'esposizione e, dall'altro lato, si evidenzieranno i benefici e gli aspetti problematici, per molti versi ancora irrisolti.

Leitmotiv di questo lavoro, infatti, saranno tanto gli aspetti controversi e critici della *digitalization* e *datafication* – rispetto della privacy, della libertà accademica, delle possibilità di scelta individuale – quanto evidenziare l'impreparazione dell'Università nel suo complesso ad affrontare un processo che, ben lungi dall'essere puramente tecnico e tecnologico, solleva questioni etiche, sociali, economiche, ma anche politiche e geopolitiche.

Partendo da una consensuale definizione generale sulla natura dei dati, si può procedere con l'analisi delle implicazioni del processo di *datificazione*. Tali innovazioni, infatti, non solo hanno modificato profondamente i modi di produrre e consumare beni e servizi, creando nuovi bisogni, ma hanno messo in moto la comparsa di nuove professionalità, di nuove esigenze di formazione e di aggiornamento professionale, legate all'uso delle moderne tecnologie informatiche.

Se, da un lato, la digitalizzazione ha automatizzato e semplificato numerose operazioni non più richieste all'agente umano, dall'altro ha prodotto un fabbisogno di nuove figure professionali e di nuove competenze, nonché ridotto o addirittura eliminato la necessità di certe mansioni. Tale processo ha conosciuto una ancor più marcata accelerazione grazie alla sopra citata *datificazione*, in atto nella nostra società. Un esempio di nuove professioni affermatesi con la trasformazione digitale è riconducibile a figure quali lo sviluppatore di applicativi Web, il programmatore, il tecnico informatico, o, in tempi più recenti, il *data scientist*, o “scienziato dei dati”, e il *social media manager*.

Più in generale, la trasformazione del mondo del lavoro ha creato l'esigenza, virtualmente in (quasi) ogni contesto lavorativo, di saper impiegare in maniera funzionale strumenti di comunicazione elettronica (come le e-mail, i sistemi di messaggistica istantanea, o le piattaforme di video-conferencing, come Zoom e/o Skype) e anche di *office automation* (che comprendono i programmi di videoscrittura, presentazione di diapositive digitali, fogli di calcolo).

Non a caso, queste competenze divengono importanti nel contesto contemporaneo sin dall'introduzione a livello europeo, nel 1997, della Patente Europea del Computer (ECDL: *European Computer Driving License*), che, cavalcando il parallelismo con la patente dell'automobile, certificava la capacità dell'utente di conoscere e saper utilizzare, almeno ad un livello di base, computer e applicativi informatici, nonché di possedere una conoscenza essenziale circa le caratteristiche ed il funzionamento delle tecnologie digitali. Dal 2019, questa certificazione è stata rinominata in ICDL (*International Certification of Digital Literacy - Certificazione Internazionale di Alfabetizzazione Digitale*), in cui l'impiego del termine *alfabetizzazione* è già di

per sé emblematico: la competenza digitale è divenuta, infatti, talmente necessaria da essere equiparata alle conoscenze di base fondamentali per l'individuo, come saper leggere, scrivere e “far di conto” (*reading, writing e numeracy*).

Un'ulteriore sfida formativa per i lavoratori proviene di certo dal crescente impiego dell'Intelligenza Artificiale (IA) nei luoghi di lavoro e dagli innovativi sviluppi della robotica, sempre più vicini a ideare sostituti dell'essere umano in diversi ambiti. In primo luogo, l'IA viene definita da Lane e Saint-Martin (2021, p. 17, TdA), come:

«Un sistema IA è un sistema machine-based che può, per un determinato insieme di obiettivi definiti dall'uomo, fare previsioni, raccomandazioni o assumere decisioni che influenzano ambienti reali o virtuali. Utilizza input macchina e/o umani per percepire ambienti reali e/o virtuali; astrae tali percezioni in modelli (in modo automatizzato, ad es. con Machine Learning (ML) [apprendimento automatico, NdA] o manualmente); e utilizza l'inferenza del modello per formulare opzioni per l'informazione o l'azione. I sistemi di intelligenza artificiale sono progettati per funzionare con diversi livelli di autonomia.»

L'Intelligenza Artificiale viene inoltre definita dall'OCSE (2019, citato da Lane e Martin, 2021, p. 19, TdA), come una “tecnologia d'uso generale”, o “tecnologia multiuso” (“*General Purpose Technology*”, o “*GPT*”). Si tratta, in pratica, di soluzioni tecnologiche che possono trovare impiego in svariati contesti e che hanno il potenziale per modificare radicalmente il lavoro, l'economia e la società: i primi esempi riscontrabili di questo genere di tecnologie sono il motore a vapore, l'elettricità e il calcolo computazionale (ibidem).

Come riportato da Lane e Saint-Martin (ibidem, p. 9, TdA), inoltre, *«le occupazioni giudicate come più esposte all'IA includono professioni altamente qualificate che coinvolgono compiti cognitivi non routinari, come tecnici di laboratorio, ingegneri e attuari»*, a dimostrazione che tali tecnologie “cognitive” siano sempre più mature e capaci di essere impiegate in contesti nuovi, grazie a nuove capacità di problem-solving, ragionamento logico e abilità percettive (ivi, p. 20).

A tal proposito, gli autori suggeriscono che *«i lavoratori potrebbero avere bisogno di riqualificazione (“reskilling”, NdA), o di sviluppare le proprie competenze (“upskilling”, NdA), al fine di adattarsi alla riorganizzazione dei compiti e all'emergere di nuovi, e di anticipare la potenziale perdita di posti di lavoro e passare a nuove occupazioni. Questo non significa solo acquisire competenze legate all'IA, ma anche acquisire competenze in settori che l'IA non può svolgere così bene, come l'intelligenza creativa e sociale, le capacità di ragionamento e l'affrontare l'incertezza»* (ivi, p. 9, TdA).

Sebbene Intelligenza Artificiale e tecnologie correlate (come, per esempio, la Robotica) possano sollevare l'uomo da compiti ripetitivi, dannosi o pericolosi, o anche potenziarne l'azione fornendo supporto all'agente umano stesso (ivi, p. 10), è altresì vero come molte occupazioni siano esposte alla sostituzione tecnologica. Tra queste possiamo annoverare anche professioni ad elevato valore aggiunto nell'industria e nella ricerca, come l'ingegnere nucleare, chimico o civile, il tecnico di laboratorio clinico, lo statistico e l'epidemiologo, ma anche figure impiegate come il contabile, l'analista di credito o, infine, le professioni nel settore produttivo legate ai controlli ispettivi e di qualità (ivi, p. 24).

Tra i risvolti negativi, inoltre, è possibile annoverare anche il pericolo di un uso improprio dell'IA. Essa, infatti, potrebbe essere usata a scopo di sorveglianza e vigilanza, così come essere impiegata con lo scopo precipuo di tagliare i costi, sostituendo la forza lavoro con sistemi automatizzati, nonché per assumere decisioni – sulla base delle previsioni da essa prodotte – che siano rilevanti per l'intera organizzazione e per i collaboratori, secondo modalità opache e non condivise. Tutto ciò potrebbe avere notevoli effetti negativi per il benessere psico-fisico dei lavoratori, generando stress, sfiducia, insicurezza e paura. (ivi, p. 10).

Nel tentativo di arginare il potenziale distruttivo di questa tecnologia, Elon Musk (imprenditore sudafricano, fondatore di “Tesla” e “SpaceX”) e Sam Altman (presidente dell'acceleratore di startup “Y Combinator”), hanno dato vita nel 2015 ad OpenAI, una società che si ripromette di creare intelligenze artificiali “amichevoli”, che possano portare beneficio a tutta l'umanità. L'organizzazione opera collaborando con altri soggetti, sia pubblici che privati, inizialmente anche rendendo ricerche e brevetti aperti al pubblico. L'organizzazione, nata come *no-profit*, è stata trasformata in società “a scopo di lucro” già nel 2019, abbandonando le pratiche di condivisione tecnologica che ne avevano ispirato la fondazione. A settembre 2020, l'azienda ha siglato una *partnership* con Microsoft Corporation al fine di concedere a quest'ultima l'utilizzo *in via esclusiva* della sua intelligenza artificiale generativa denominata GPT-3; a febbraio 2023, Microsoft ha incorporato la chatbot chatGPT (“*Chat Generative Pre-trained Transformer*”, letteralmente “*Trasformatore Pre-addestrato Generatore di conversazioni*”), un modello linguistico capace di scrivere testi in linguaggio naturale, all'interno del suo motore di ricerca Bing. All'inizio del 2023, Microsoft ha inoltre sottoscritto un accordo pluriennale con OpenAI del valore, secondo Bloomberg, di circa 10 miliardi di dollari, per sostenerne i progetti e rafforzare il ruolo di Microsoft, in quanto fornitore di soluzioni tecnologiche di OpenAI.

Sulla base di quanto affermato da vari osservatori indipendenti, chat-GPT si dimostrerebbe capace di risposte convincenti e coerenti; essa si dimostra in grado di scrivere articoli tecnico-scientifici consistenti, codice per programmi informatici, sintesi di testi e anche componenti

poetici. Tuttavia, basandosi su un modello di apprendimento statistico del linguaggio, essa non ha alcuna percezione del vero o del falso, né alcuna “idea” sul mondo che la circonda, fornendo talvolta informazioni palesemente errate o inventate. Trattandosi di una “fonte primaria” di informazioni – ovvero, una fonte che non cita l’origine delle notizie da essa riportate – essa può provocare effetti negativi, inquadrabili nel fenomeno delle “fake news”.

Inoltre, essendo stata addestrata su un *corpus* di dati abbastanza variegato, essa può, per imitazione, ripetere insulti o espressioni razziste (linguaggio cosiddetto “tossico”), o fornire risposte inappropriate. La tecnologia, pertanto, si dimostra promettente ma ancora acerba, nonostante lo sviluppo e l’apprendimento continuo realizzati anche grazie agli input degli utenti (Floridi, 2023).

Si può quindi affermare che, ancora una volta, dietro all’IA si celino sia opportunità – in primo luogo quello di ampliare e semplificare l’azione dell’essere umano – che risvolti potenzialmente dannosi: la chatbot ha mostrato di poter generare *codice malevolo* (virus informatici e simili), *spam* (posta elettronica indesiderata) ed e-mail di *phishing* decisamente credibili. Il suo impiego, inoltre, mette a repentaglio lavori di scrittura più “tecnici”, come quello giornalistico o enciclopedico. Non dimentichiamoci che essa potrebbe anche prestarsi al plagio e alla violazione del diritto d’autore, essendo stata addestrata su materiale coperto da diritto d’autore senza il consenso dei detentori di tali diritti.

Si può dire, quindi, che la trasformazione digitale non sia avvenuta in maniera indolore. Possiamo vederne le tracce anche su dinamiche sociali preesistenti, talvolta accentuate da ciò, come sottolineato anche da Knox (2019). L’autore, infatti, sostiene che: «...*tali tecnologie devono essere considerate non solo come “adattate a” contesti politici ben consolidati, ma stanno anche lavorando per amplificare e intensificare tali contesti con precisione e trasmissività senza precedenti.*» (Knox, 2019, p. 281, TdA).

Ancora più significativo è l’intervento riportato poco dopo: «*modelli storici di discriminazione e classificazione, che spesso costruiscono rappresentazioni dannose di persone basate su differenze percepite, si riflettono nelle ipotesi e nei dati che informano i sistemi di IA, spesso con conseguenti danni allocativi. Questa prospettiva richiede di andare oltre l’individuazione di pregiudizi in un algoritmo o set di dati, e di considerare “il ruolo dell’IA nelle rappresentazioni dannose dell’identità umana,” e il modo in cui tali rappresentazioni dannose sono sia modellate, sia danno forma, alle nostre comprensioni sociali e culturali di noi stessi e gli uni degli altri.*» (Whittaker et al., 2018, TdA).

Knox introduce così la questione della regolamentazione etica delle Intelligenze Artificiali (IA), in quanto responsabili della riproduzione sistematica di bias e stereotipi fautori di disuguaglianze storiche (Lane e Saint-Martin, 2021, p. 10). È il caso, per esempio, della riproduzione di *bias*

discriminatori nella selezione del personale, o le disuguaglianze di accesso alla rete causate da ritardi infrastrutturali, il cosiddetto *digital divide*, particolarmente sentito durante i lockdown pandemici e che ha provocato l'isolamento di intere comunità, o reso impossibile fruire della Didattica-a-Distanza, aggravando problemi socioeducativi preesistenti. Ad essere in discussione sono, quindi, la giustizia sociale e l'inclusività, in opposizione alle disuguaglianze, alla discriminazione e alla marginalizzazione sociale.

All'interno di questa cornice trovano luogo anche le perplessità scaturite dalla sempre maggiore intrusività di tali sistemi di raccolta dati nella privacy individuale (Stewart, 2023) e i timori per l'impatto che le nuove infrastrutture tecnologiche – come i datacenter, gli enormi centri di elaborazione dati – hanno a livello ambientale, in quanto responsabili di un grande consumo di energia e risorse e conseguente elevato livello di emissioni di gas serra.

A questi timori di carattere etico-ecologico si aggiungono pertanto i temi del consumo delle limitate risorse planetarie e della produzione di rifiuti elettronici di ogni tipo. È necessario, infatti, ricordare come ogni dispositivo elettronico richieda materiali talvolta rari, come il litio, che vengono estratti dal sottosuolo esacerbando ingenti danni agli ecosistemi locali. Come se non bastasse il tasso di sviluppo tecnologico rende presto obsolete le tecnologie in uso, generando una folle corsa al consumo di dispositivi sempre più nuovi ed aggiornati (Crawford, 2022).

I timori politici, infine, si intersecano con questioni di *sovranità digitale* e *sicurezza nazionale*: permettere a soggetti stranieri di entrare nel mercato delle infrastrutture e servizi di telecomunicazioni – così come di possedere, in maniera totale o parziale, diretta o indiretta, società operanti in questo campo – dà adito a timori e preoccupazioni legati all'effettivo potere esercitato dagli stati per mantenere al sicuro dati sensibili di ogni genere, dalle informazioni personali dei cittadini a dati di rilevanza strategica per le nazioni. L'Agenzia Europea per la Cyber Sicurezza sottolinea come, nel 2022, il maggior numero di attacchi informatici sia avvenuto a danno di soggetti pubblici istituzionali (ENISA, 2022, p. 17), di cui un numero significativo con l'obiettivo del furto di dati sensibili e come movente lo spionaggio o l'influenza geopolitica (ibidem, p. 20, 29). I gruppi ritenuti responsabili di tali azioni sono ricondotti all'operato di governi stranieri (ivi, p. 30).

Una minaccia particolare per le democrazie emerge, inoltre, dalle attività di profilazione a carico dell'utente, a volte svolte da soggetti interessati ad influenzare i processi democratici per i propri scopi. Ancora una volta, l'ENISA evidenzia disinformazione e mistificazione tra le maggiori minacce alla sicurezza digitale nel corso del 2022 (ivi, p. 5); infatti, ottenendo un profilo del soggetto che utilizza il servizio, è possibile indirizzare comunicazioni personalizzate, *tailor-made*, per condizionarne pensieri e decisioni. Poco importa che tali informazioni siano vere oppure false,

come nel caso delle *fake-news*: l'importante è che siano congruenti con gli schemi di pensiero del destinatario, in modo tale da avere la massima efficacia.

All'interno di questo quadro preesistente rispetto alla fase pandemica, si inserisce di forza la diffusione globale del virus del Covid-19. La pandemia che ne è seguita ha segnato una forte discontinuità nella storia del mondo e nella diffusione delle tecnologie informatiche tra individui, aziende e istituzioni, affrettando e amplificando processi già in atto. La dilagante chiusura di scuole, università, uffici e attività commerciali ha difatti generato una enorme e dilagante necessità di servizi fruibili da remoto: all'interno di questo *trend* generalizzato si è riscontrata una crescita vertiginosa degli e-commerce – i siti di commercio elettronico – particolarmente apprezzati durante i lockdown, in quanto semplice e praticabile via d'accesso al mercato di consumo; nonché l'esplosione dello *smart-working*, il “lavoro agile”, compiuto da remoto grazie alla rete Internet; o ancora quella che potremmo chiamare la “metamorfosi” di scuole e università, che da luoghi deputati alla didattica in presenza si sono trasformati in spazi virtuali per processi educativi a distanza.

In particolare, gli ultimi due fenomeni citati si possono considerare delle assolute novità, sia all'interno del contesto italiano che europeo. Il telelavoro – o lavoro da remoto – era infatti una prassi fino alla pandemia sostanzialmente sconosciuta, nonché relegata al livello della mera sperimentazione; mentre per la Didattica-a-Distanza è stata la prima volta nelle aule di università e scuole pubbliche di tutto il mondo.

1.2. Dalla datafication alla platformization, fino al post-digitale

1.2.1. Il modello di business delle piattaforme

Come precedentemente accennato, i servizi delle Big Tech (le grandi aziende operanti nell'industria digitale: Google, Apple, Microsoft, Amazon, Facebook) hanno radicalmente cambiato molti settori del quotidiano: in particolare, si può citare lo sviluppo di reti sociali e professionali, attraverso social network quali Facebook e LinkedIn; la ricerca di informazioni e fruizione dei servizi sul Web con Google; fino all'acquisto di beni tramite piattaforme quali Amazon. Tali soggetti vengono indicati come *piattaforme*, secondo la definizione proposta da Dijck et al. (2018), nel libro "The Platform Society": «una "piattaforma" online è un'architettura digitale programmabile pensata per organizzare le interazioni tra gli utenti, non solo gli utenti finali, ma anche le entità aziendali e gli enti pubblici» (TdA). Ogni aspetto della quotidianità è ormai pervaso dalla presenza dell'ICT, secondo un fenomeno che è stato definito "platformization" ("piattaformizzazione"), in riferimento alle sopracitate piattaforme digitali.

Dijck et al. (ibidem, p. 1, TdA), a riguardo, afferma: «la promessa delle piattaforme è quella di offrire servizi personalizzati e di contribuire all'innovazione e alla crescita economica, aggirando efficacemente le organizzazioni esistenti, regolamentazioni onerose e spese inutili». A questa premura di partecipare ai processi di innovazione tecnologica e di efficienza globale, tuttavia, non sempre hanno fatto seguito azioni a beneficio della collettività. Se da un lato «il mercato utopico basato su Internet consentirebbe agli individui di offrire prodotti e servizi "direttamente" senza il bisogno di fare affidamento su intermediari "offline", siano essi statali o aziendali», (ivi, p. 2, TdA) creando nuove opportunità di business e di crescita economica, dall'altro causano «una profonda disputa tra il profitto privato rispetto al beneficio pubblico, in una società in cui la maggior parte delle interazioni avviene via Internet. Mentre le piattaforme, ipoteticamente, potenziano i benefici individuali e il guadagno economico, mettono contemporaneamente pressione sui mezzi collettivi e sui servizi pubblici» (ivi, TdA).

Pertanto, nonostante gli indubbi vantaggi e le nuove opportunità che tale sviluppo ha generato, non si può dire che tali benefici siano stati "a costo zero", né per la collettività – che deve sostenere i costi di attività economiche poco o per nulla regolamentate – né per l'utente finale: il prezzo di servizi online – spesso proposti come "gratuiti" – viene invece pagato attraverso una sempre maggiore intrusione nella privacy degli individui da parte di queste società, che scandagliano sistematicamente dispositivi e azioni eseguite dagli utenti a fini di profilazione. Come viene affermato da Dijck et al. (ivi, p. 4, TdA), infatti, la piattaforma «è orientata verso la raccolta

sistematica, l'elaborazione algoritmica, la circolazione e la monetizzazione dei dati degli utenti». Ogni interazione attraverso mezzi digitali viene tradotto in *dati*, unità grezze di informazione, che vanno a costituire un flusso costante e di cui l'utente di solito non è consapevole, essendo così impossibilitato ad esprimere una propria agentività (Raffaghelli, 2017, p. 304).

Tale processo di raccolta dati su larga scala si avvantaggia, da un lato, della sostanziale ubiquità delle tecnologie digitali nella nostra esistenza e, dall'altro, delle enormi risorse sia finanziarie che tecnologiche di cui le Big Tech dispongono, le quali permettono anche, in ultima analisi, di elaborare enormi flussi informativi all'interno di strutture che raccolgono centinaia – se non migliaia – di server dedicati a questo scopo, i *datacenter*.

La raccolta di queste informazioni, che avviene con ogni dispositivo *smart* – ovvero dotato di funzionalità avanzate quali la possibilità di collegarsi ad internet – avviene perlopiù all'insaputa degli utenti, come sottolineano vari autori (tra cui Raffaghelli (2019), Selwyn e Jandrić (2020), Atenas et al. (2020), Raffaghelli et al. (2020, 2021)). Gli individui appaiono ignari – a volte disinteressati – al fatto che ogni loro gesto, scelta, post o foto inserito online diventi di fatto proprietà delle grandi piattaforme. Codificato e anonimizzato, tale insieme di dati viene elaborato attraverso specifici algoritmi, i quali sono definiti come «*un set di istruzioni automatizzate che trasformano i dati in input in un output desiderato*» (Gillespie 2014; Pasquale 2015, citato da Dijck et al., 2018, p. 9). Lo scopo, l'*output desiderato*, è quello di produrre un profilo dell'utente che utilizza il servizio; il profilo ottenuto potrà poi essere utilizzato dalla piattaforma stessa, per esempio per finalità di marketing, oppure venduto ad altri soggetti economici, ai quali viene garantito l'accesso alla piattaforma (“*mercattizzazione*”, o “*marketization*”). In quest'ultimo caso, come lo scandalo *Cambridge Analitica* ha dimostrato, i dati personali degli utenti possono essere impiegati per interessi di parte, per esempio condizionando l'opinione pubblica, diffondendo fake news attraverso inserzioni mirate.

In generale, l'interesse delle Big Tech sembra di natura prettamente economica ed orientata alla crescita dell'audience e alla promozione dei propri beni e servizi. L'obiettivo appare, pertanto, quello di individuare – con sempre maggior precisione – interessi e gusti del pubblico, cercando poi di influenzarlo con azioni specificamente orientate a tale scopo. In questa direzione, rivestono particolare rilevanza sia «*i recommender system, che basandosi sulle suddette forme di analisi fanno comparire raccomandazioni di servizi o pubblicità nelle pagine web visitate da un utente (in particolare social network)*» (ivi, p. 305, TdA), sia le tecniche di analisi predittiva (“*predictive analytics*”), che permettono di elaborare previsioni su comportamenti ed eventi futuri.

L'efficacia di questo modello di business – la raccolta (“*data gathering*”, o anche “*data mining*”), l'analisi (“*data analysis*”) e l'eventuale vendita a terzi di ingenti quantità di dati

(“*monetizzazione*”) – si evince dalla crescita vertiginosa dei fatturati di queste aziende negli ultimi vent'anni. Non a caso, ci si riferisce a queste società con l'espressione “*Over-The-Top*” (“OTT”), sia per il loro posizionamento – quasi monopolistico – nei mercati dell'industria digitale, sia per il volume dei loro fatturati e delle dimensioni aziendali.

A titolo di esempio, Google detiene una posizione dominante nel mercato dei motori di ricerca ed una quota rilevante del mercato dei sistemi operativi per smartphones grazie al sistema Android; quest'ultimo, in particolare, è stato un veicolo estremamente importante per diffondere a livello globale le applicazioni di Google in tutte le loro svariate forme, posizionandoli in prima linea all'interno degli smartphone di milioni di persone e, di conseguenza, accedendo ad un insieme tanto eterogeneo, quanto vario, di dati personali.

Utilizzando un qualsiasi servizio Google (per i quali è necessaria, in primo luogo, una registrazione da parte dell'utente), infatti, la società si riserva il diritto di accedere ed utilizzare qualsiasi informazione l'utente inserisca, siano esse e-mail, documenti condivisi con altri utenti, immagini, ricerche sul web, o altro ancora. Con le dovute differenze, tale schema viene replicato con successo dalle altre grandi società dell'ICT.

Emblematico, in questo senso, è il caso di Windows 11: all'interno dell'ultima *release* del noto sistema operativo di Microsoft (per utilizzare il quale è richiesto un account Microsoft), i processi di sistema che raccolgono in maniera automatica i dati degli utenti e li inviano in rete senza un esplicito consenso sono diverse centinaia. All'opposto, nel caso di Windows XP (sistema operativo uscito nel lontano 2001), esisteva un solo processo che, in *background*, si collegava a Internet: si trattava, in questo caso, dell'utility di sistema Windows Update, responsabile di rilevare la presenza degli ultimi aggiornamenti disponibili e di installarli nel computer dell'utente. Pertanto, oltre allo stato del sistema, nessun'altra informazione veniva condivisa con Microsoft.

1.2.2. Le “Big Tech” tra disillusione e post-digitale

Oltre alla dimensione tecno-economica, è necessario sottolineare un'altra caratteristica delle piattaforme – forse meno evidente – ma di grande impatto: si tratta della dimensione sociopolitica. «*Le piattaforme non sono costrutti neutri o privi di valore; sono dotate di specifiche norme e valori inscritti nelle loro architetture*» (Dijck et al., 2018, p. 3, TdA). Le piattaforme si accompagnano ad una specifica cultura e ad una loro ideologia, che ne costituisce la narrazione fondamentale e orienta le interazioni degli utenti tra loro e con la piattaforma stessa.

Questa narrazione può riguardare, per esempio, il primato del libero mercato sullo Stato, l'uguaglianza delle relazioni orizzontali contro le gerarchie verticali, l'enfasi sulle reti sociali come

driver di sviluppo professionali o di autopromozione, o anche la possibilità di maggiore controllo che verrebbe posto nelle mani dell'utente, grazie a capacità avanzate di raccolta e visualizzazione di dati all'interno di applicazioni specificamente progettate.

La rilevanza *politica* di narrazioni ideologiche si può facilmente comprendere nel momento in cui si prende coscienza del fatto che le infrastrutture su cui le piattaforme globali – nonché i loro clienti/utenti – si appoggiano, sono *di proprietà privata*: qualora si verifici un conflitto tra piattaforme e utenze, in cui i secondi esprimono posizioni contrarie alle prime, queste possono revocare l'utilizzo della piattaforma sulla base di specifici accordi chiamati *Termini di Servizio* (“*Terms of Service*”, o “*ToS*”), (Williamson e Hogan, 2021, p. 2). Nella pratica, un utente che esprima, in maniera più o meno diretta, opinioni e vedute contrarie a quelle della piattaforma dominante può vedersi sospeso il servizio potenzialmente in qualsiasi momento e a tempo indefinito.

Pertanto, come ci ricorda Knox (2019, p. 281, TdA), è «*la separazione tra politica ed economia ad essere in discussione*». Secondo l'autore è avvenuto uno spostamento dal *digitale* al *post-digitale*, in primo luogo nei sentimenti delle persone verso le piattaforme, «*dai primi potenziali emancipatori, alla successiva disillusione e crescente corporativizzazione*» (ibidem, p.281, TdA). Sarebbe avvenuto un cambiamento radicale nella *percezione* delle Big Tech, da soggetti “progressisti” e fautori di una maggiore libertà personale, ad attori economici globali con tendenze marcatamente privatistiche e corporativistiche, portatori di interessi economici che spesso operano in “zone grigie” sotto il profilo legale e, pertanto, *non responsabili* delle ricadute negative delle azioni intraprese.

Si stanno pertanto sviluppando sentimenti negativi verso le piattaforme, un fenomeno noto come “*techlash*”. Selwin e Jandrić (2020, p. 994, TdA), a tal proposito, riporta: «*Ho sempre utilizzato il termine “post-digitale” nell'accezione di “guardare oltre” l'apparente progresso tecnologico e la novità, per riconoscere invece il nostro senso di inquietudine e disillusione nei confronti dell'attuale società ipertecnologica*» [lett. “*piena di tecnologia*”, NdA].

A questa disillusione, quindi, è seguita la critica del digitale e delle tecnologie collegate, nel tentativo di svelare la complessa rete di relazioni economiche, sociali, politiche e finanche geopolitiche che ruota attorno ad essi. Infatti, sempre secondo Knox (ivi, p. 281, TdA), «*parte della spinta critica del post-digitale è quello di tentare di operare un decentramento dalla tecnologia del digitale in sé, per fare in modo che le sue relazioni con strutture più ampie siano portate alla luce*». E subito dopo: «*parte di questo interesse nello svelare la geopolitica delle tecnologie digitali deriva dall'idea che l'industria tecnologica stessa sembra convenzionalmente trattare le questioni etiche come una questione di ingegneria proprietaria*».

Oltre a ciò che Knox definisce la “*geopolitica della tecnologie digitali*” c’è anche la questione etica, che riguarda in primo luogo il fatto che le società tecnologiche trattano gli aspetti etico-legali da un punto di vista puramente tecnico-ingegneristico. Whittaker et al (2018), citato da Knox (2019, TdA), «*mette in guardia contro la tendenza [delle aziende tecnologiche, NdA] a riformulare le questioni politiche come preoccupazioni tecniche*».

Come già accennato in precedenza, ad essere in gioco non sono solo interessi economici, o la dicotomia pubblico-privato, ma equilibri di potere globali tra nazioni, aziende e gruppi sociali: degli esempi sono rappresentati dal *soft-power* che le piattaforme possono esercitare nei confronti degli utenti, siano essi individui, gruppi o istituzioni; ma anche, e soprattutto, dalla rete di relazioni emergenti che si nasconde dietro alla produzione tecnologica. Più esplicitamente, si può citare l’industria degli *smartphone*, i quali sono progettati e sviluppati nei paesi industrializzati, ma costruiti in paesi a basso costo del lavoro e grazie a materie prime estratte e raffinate in nazioni del Sud globale; tra queste risorse possiamo annoverare, ad esempio, le “*terre rare*”, un gruppo di 17 elementi chimici indispensabili per il funzionamento dei circuiti elettronici. L’importanza geopolitica di tali risorse si spiega con la loro assoluta centralità per l’industria elettronica e ad elevata tecnologia: esse trovano impiego in diverse applicazioni *hi-tech*, tra cui magneti, superconduttori, laser, fibre ottiche e batterie. Leader mondiale in questo campo è la Cina, che ospita sul suo territorio l’80% circa dei giacimenti esistenti di tali risorse, e ne produce il 95% a livello globale (Crawford, 2022).

Queste relazioni di *interdipendenza* ridisegnano i rapporti di potere globali, in cui nazioni e organizzazioni possono subire l’influenza di attori esterni soprattutto tramite la minaccia di ritorsioni che, spesso, prendono la forma di sanzioni dirette verso gli stati o verso specifiche imprese e istituzioni, con ricadute potenzialmente distruttive sull’economia e l’occupazione.

1.3. Dalla *datafication* alla *platformization*, fino al *post-digitale*: la digitalizzazione dell'educazione

1.3.1. Le tecnologie digitali come risposta ai problemi dei sistemi educativi

All'interno di questo scenario, fatto di luci ed ombre, ha fatto irruzione nei primi mesi del 2020 la pandemia da Covid-19, a cui le autorità nazionali hanno risposto con restrizioni via via più ferree alla circolazione delle persone, culminate spesso con l'imposizione di chiusure totali, quali i *lockdown*.

Quest'ultima misura, certamente drastica, ha avuto l'indubbio effetto di abbassare la curva dei contagi entro limiti sostenibili per i sistemi sanitari nazionali; ciò ha comportato, tuttavia, l'impossibilità di lasciare il proprio domicilio se non per ragioni comprovate, la chiusura di molte attività commerciali e, in generale, la soppressione di qualsiasi spostamento non essenziale, compresi quelli per motivi di studio e lavoro. In particolare, sono due i fenomeni che hanno caratterizzato tale nuovo stato di cose: da un lato, l'introduzione di nuove modalità di lavoro da remoto (quali lo *smart working*, o "lavoro agile"), e la Didattica-a-Distanza (DAD), attraverso la quale docenti e studenti eseguivano online tutte le attività in precedenza svolte nell'aula fisica dell'istituto d'appartenenza.

Se l'insegnamento a distanza era già balzato agli onori delle cronache grazie alle piattaforme private di MOOC ("*Massive Open Online Courses*", "Corsi Online Aperti di Massa"), per le istituzioni pubbliche il ricorso a modalità di didattica virtuale era una assoluta novità. Ciò ha determinato l'introduzione di tecnologie e metodologie del digitale in maniera pressoché inedita, nonché massificata, in ambiti in cui gli spazi fisici dell'aula accademica o della scuola avevano ancora un rilievo indiscusso: come sottolineato da Fiebig et al. (2021), «*mentre il software educativo per l'insegnamento a distanza aveva già ricevuto attenzione prima della pandemia [...], il Covid-19 ha aumentato l'importanza delle infrastrutture di apprendimento come strumenti di video-chat e soluzioni di streaming*» (Fiebig et al, 2021, traduzione a cura dell'autore).

Questo processo di adozione delle tecnologie di *online communication* e *online collaboration* nelle scuole e nelle università è stato talmente massiccio da spingere alcuni autori a parlare di "zoomificazione dell'educazione".

Sempre Fiebig et. al (2021), a riguardo, afferma: «*Gli strumenti per la video-chat e le soluzioni VoIP hanno già assunto da lunga data una certa rilevanza nelle comunicazioni professionali, in particolare sotto forma di Skype-for-Business (SfB). Tuttavia, con l'emergere del Covid-19, questi strumenti, specialmente Zoom, hanno guadagnato un'attenzione pubblica significativa, con attività*

accademiche di base – insegnamento, incontri di ricerca e conferenze basate fortemente su questi strumenti. Infatti, la discussione intorno alla dipendenza delle università e dell'istruzione dalle infrastrutture commerciali spesso inquadra questo come fenomeno come la “zoomificazione” dell'istruzione».

L'utilizzo di Zoom per la Didattica-a-Distanza in periodo pandemico è stato, per l'appunto, un esempio emblematico di tale fenomeno, così come il contemporaneo uso di applicazioni per la *supervisione* e la *valutazione* degli studenti (ibidem, p. 2); tuttavia, la migrazione delle università verso soluzioni *cloud* può essere retrodatata almeno al 2015, sebbene con intensità e proporzioni variabili all'interno dei diversi contesti nazionali (ivi, p. 4). A titolo esemplificativo, negli Stati Uniti e nel Regno Unito l'utilizzo di servizi in cloud appartenenti ai “*Big Three*” (Amazon, Google, Microsoft), era già piuttosto comune a metà decade 2010, con il 95% degli atenei statunitensi già operanti su queste piattaforme, mentre in Gran Bretagna erano il 75%; nell'Europa continentale, la nazione con il sistema accademico più esposto verso il cloud era l'Olanda, con una percentuale di atenei pari al 50%, mentre all'opposto l'Austria non ne aveva alcuno (ivi, p. 5). Tali percentuali, sebbene con velocità differenti, sono cresciute ovunque nel corso degli ultimi anni.

Si può pertanto parlare di “*piattaformizzazione dell'educazione*”, intendendo con tale espressione descrivere il graduale e crescente ricorso alle piattaforme da parte delle istituzioni educative, a sua volta accelerato in periodo pandemico. Raffaghelli (2021, p. 3), a riguardo, afferma: «*durante la prima fase critica della pandemia COVID19, la rincorsa alle soluzioni rapide ha dato luogo ad un fenomeno di prevalenza di tecnologie “GAMAF” (Google, Amazon, Microsoft, Apple, Facebook) senza una approfondita riflessione sulla concentrazione abnorme di dati e suo conseguente sfruttamento a favore delle suddette piattaforme private (Bozkurt et al., 2020) e a discapito di ottimi progetti di infrastruttura nazionale (per esempio rete GARR) ed istituzionale (Moodle gestiti dagli Atenei)».*

Tale fenomeno, tuttavia, ha radici remote, e scaturisce dalla precedente digitalizzazione delle istituzioni educative. Infatti, Feenberg (2019, p. 8, TdA), parlando di digitalizzazione educativa, sottolinea: «*Quello che abbiamo visto negli ultimi 30 anni è un'ossessione per i computer e Internet, soprattutto da parte degli amministratori scolastici e universitari*».

L'adozione di tecnologie digitali ed informatiche, nel contesto educativo, è stato visto come un'opportunità di innovazione ed obiettività (Raffaghelli e Stewart, 2020, p. 438) e riforma delle istituzioni scolastiche a tutti i livelli, in quanto capace di scardinare i tradizionali paradigmi didattico-educativi e portare a maggiore efficienza e razionalità nell'uso delle risorse.

In questo senso, la digitalizzazione è stata proposta come risposta alle ristrettezze finanziarie in cui versavano i sistemi d'istruzione, nonché come una svolta intrinsecamente positiva e

rivoluzionaria nella storia dell'educazione. A riguardo, Feenberg (2019, p. 8, TdA) sottolinea: «*La tecnologia promette la modernizzazione di istituzioni che si suppone essere legate alla tradizione. Tecnologia contro tradizione giustifica enormi investimenti e pressioni per il cambiamento. Questo è il significato del digitale, non come realtà tecnica ma come fenomeno ideologico*».

Williamson e Hogan (2021, p. 16, TdA), citando Marres (2017), ci ricordano che: «*La recente rapida digitalizzazione di tutti i settori è spesso associata alla convinzione che i computer sono sempre la soluzione ai problemi sociali, sono più obiettivi della ragione umana e promettono miglioramenti ovunque vengano impiegati*».

Si può parlare, a riguardo, di *tecno-soluzionismo* o *soluzionismo tecnologico*, intendendo con tale espressione la concezione della tecnologia come soluzione di tutti i problemi delle istituzioni educative, siano essi finanziari, organizzativi o didattico-pedagogici (Williamson e Hogan, 2021, p. 2). Tale prospettiva è stata avanzata, in particolare, da soggetti privati portatori di interessi particolari, orientati alla “modernizzazione” dell’istruzione come opportunità di business.

In tal senso, sono aumentate le pressioni verso la politica per intraprendere una sempre maggiore *privatizzazione* dell’educazione, a tutti i livelli: questo processo viene definito da Williamson e Hogan (2021, p. 2, TdA) come indirizzato a « [...] *realizzare i presunti benefici della mercatizzazione per i sistemi di istruzione – come l’efficienza e l’efficacia – attingendo alle tecniche e ai valori del settore privato, introducendo la partecipazione del settore privato, e/o rendendo l’istruzione pubblica più simile a un’impresa*».

Tale processo di privatizzazione, pertanto, non solo introduce strategie di gestione tipicamente imprenditoriali, né riguarda solamente la trasformazione dell’istruzione da bene pubblico a prodotto mercificato, ma investe le istituzioni nel loro insieme e ne modifica le modalità di funzionamento: «*Il perseguimento del vantaggio commerciale, l’imposizione di interessi commerciali come motori del cambiamento educativo e l’introduzione dei valori di mercato di efficienza, concorrenza e miglioramento delle prestazioni possono influenzare la missione delle istituzioni educative, rimodellare il giudizio e l’attenzione professionale degli educatori e riconfigurare i curricula e i metodi e le pratiche pedagogiche*» (Verger, Fontdevila e Zancajo (2016), citato in Williamson e Hogan (2021), pag. 6; TdA).

Il rischio dietro all’apparente “neutralità” dei concetti di efficienza ed efficacia è quello di sostituire gli obiettivi e la missione del sistema educativo, riorientandolo dal suo potenziale emancipatorio ed inclusivo, nonché promotore d’uguaglianza, responsabilità e consapevolezza, agli scopi definiti dal sistema imprenditoriale e dalla società di mercato: «*La privatizzazione e la mercatizzazione possono erodere gli obiettivi fondamentali dell’istruzione pubblica e statale, come*

lo sviluppo personale, l'indipendenza intellettuale e l'orientamento sociale, politico e culturale degli studenti come cittadini informati, responsabili e attivi» (Biesta (2015), ibidem).

Gli interessi privatistici e gli orientamenti al libero mercato non sono una novità per il sistema scolastico ed universitario: fin dagli anni '80 e '90 del secolo scorso, attraverso politiche pubbliche mirate a rinsaldare i legami tra istruzione pubblica e mondo imprenditoriale ed industriale, i governi miravano a conservare il vantaggio competitivo delle nazioni all'interno della società globale della conoscenza (Williamson e Hogan (2021), p. 13).

Tuttavia, la pervasività dell'influenza di soggetti privati ad ogni livello d'istruzione è cresciuta enormemente durante la pandemia, come risposta dei governi alle criticità emerse con la crisi economica da questa provocata: riduzione dei finanziamenti pubblici, rischio di bancarotta, modelli di business insostenibili, esigenza di dimostrare il ritorno economico della laurea (Marshman et al. (2020), citato in Williamson e Hogan (2021), p. 7), sono solo alcune delle urgenze a cui il sistema ha dovuto far fronte.

1.3.2. L'istruzione tra digitale e post-digitale

I tentativi di rendere l'istruzione un fenomeno prettamente virtuale si sono tuttavia scontrati con le reticenze di insegnanti e studenti ad abbandonare l'idea dell'educazione come fenomeno faccia-a-faccia (Feenberg, 2019, p.8), dimostrando che la pura e semplice innovazione tecnica non è sufficiente a rivoluzionare le realtà educative. Sempre Feenberg (ivi, TdA) aggiunge: *«Il digitale è integrato ed embricato nelle nostre azioni e interazioni quotidiane. Questa è la tendenza dell'istruzione, dove progetti ambiziosi orientati all'automazione hanno lasciato il posto a tentativi più modesti di integrare gli strumenti digitali nei tradizionali corsi faccia-a-faccia. Questa "blended education" sembra un buon modello di post-digitalizzazione»*.

Il paradigma di "automazione dell'educazione" è stato pertanto sostituito da un approccio complementare, *post-digitale*, che vede l'introduzione di supporti informatici (documenti elettronici, diapositive, video, filmati) all'interno della tradizionale aula scolastica ed accademica, talvolta accompagnati da ambienti virtuali di discussione e condivisione (forum, wiki, lavagne digitali, ecc...), (ivi).

Pertanto, afferma ancora Feenberg (ivi, TdA) – contestando la nozione di post-digitale – *«le nozioni di "digitale" e "post-digitale" appaiono artificiali»*. L'approccio complementare adottato in tempi recenti sembra configurarsi maggiormente come una modalità "predigitale" di impiegare le tecnologie informatiche a supporto della – o in sostituzione de – l'azione educativa tradizionale.

A sostegno di ciò, Hayes et al. (2020), citato da Selwyn e Jandrić (2020, TdA), pone l'accento sulla coesistenza tra realtà virtuale e mondo fisico: «*Non siamo tutti post-digitali ma co-digitali, che ci piaccia o no*». La modalità “digitale”, infatti, prevedeva l'esclusione dell'elemento umano (“*de-umanizzazione*”) in favore di una radicale digitalizzazione dei processi educativi, realizzati se possibile “a distanza” e tramite le risorse rese disponibili su Internet (Feenberg, 2019, p. 9), un fenomeno noto anche come “*de-istituzionalizzazione*”.

1.3.3. Approcci tecnologici alla riforma delle università: una panoramica

All'interno della “società della conoscenza”, e nello specifico in un'ottica di competitività dei sistemi economici, al sistema universitario sono state riconosciute come fondamentali le funzioni di ricerca e insegnamento: la prima, infatti, risponde alla necessità di innovazione tecnica e di progresso scientifico delle “*economie della conoscenza*”, mentre la seconda permette di formare professionalità ad alto valore aggiunto, capaci di navigare la complessità e di operare con alti gradi di competenza ed autonomia.

Si è affermato, quindi, l'ideale dell'università “*knowledge-intensive*” (Williamson e Hogan (2021), p. 13). Assieme a questo nuovo approccio, che inseriva l'istruzione superiore nella cornice del più ampio sistema capitalistico, sono state associate le esigenze di sviluppare alcune misure di performance delle università, prima per stabilire e poi per raggiungere obiettivi (Olssen e Peters (2007), citato da Williamson e Hogan (2021), p. 13). Sono state, quindi, incentivate le istituzioni che operavano in conformità a tali misure e che ottenevano i risultati desiderati dal sistema (Espeland e Sauder (2016), *ibidem*). Un esempio di ciò sono i *ranking* universitari nazionali ed internazionali, le cui metriche sono state cristallizzate e rese pubbliche.

Con la progressiva digitalizzazione della società – e, all'interno di questa, dell'educazione – si è quindi assistito alla nascita del concetto di “università digitale” (“*digital-first university*”) ad “alta densità di dati” (“*data-intensive*”), come evoluzione del precedente ideale di università “ad alta concentrazione di sapere”, funzionale alle logiche del sistema economico. Si può, quindi, affermare che digitalizzazione e datificazione abbiano esaltato le prerogative intrinseche di questi paradigmi a forte connotazione neoliberista, ovvero l'efficacia, l'efficienza, la competizione, l'auditing e l'accounting, la misurazione delle performance, la gestione della qualità, la mercatizzazione, la commercializzazione e la privatizzazione (Selwyn (2014), *ivi*, p. 14).

L'università, infatti, agisce sempre più con un approccio orientato al mercato e alla creazione di valore, attraverso competizione, classifiche di rendimento, domanda di consumo e ritorno sull'investimento (Busch (2016), *ivi*).

I promotori della “università datificata” sostengono che le “università smart” sono «*istituzioni in grado di utilizzare le enormi quantità di dati che generano per migliorare l'esperienza di apprendimento degli studenti, migliorare lo sforzo di ricerca, sostenere un'efficace sensibilizzazione della comunità e far progredire l'infrastruttura del campus*» (Lane e Finsel (2014), p. 4, *ivi*; TdA).

Tuttavia, per ottenere tali risultati si è necessariamente fatto ricorso all'*expertise*, ai prodotti e servizi tecnologici di fornitori privati, i quali hanno soddisfatto tale domanda di informatizzazione e crescente datificazione coprendo tutti gli ambiti (“segmenti di mercato”) del funzionamento universitario: immatricolazioni, sistemi di gestione degli studenti, biblioteche digitali e servizi informativi, *learning analytics* (“LA”) e *learning management systems* (“LMS”), valutazione online, rilevamento plagio e, non ultime, le stesse infrastrutture digitali su cui le università si appoggiano (Williamson e Hogan (2021), p. 15).

Questa tendenza ha fatto nascere quella che è stata definita la “industria globale dell'istruzione superiore”, a cui hanno anche preso parte università pubbliche e private, fornitori di servizi digitali e di dati (Komljenovic and Robertson (2017), citato da Williamson e Hogan (2021), p. 14). Attraverso i prodotti informatici a supporto del sistema di istruzione terziaria, inoltre, i fornitori di tecnologie digitali e di sistemi di dati hanno ottenuto un'elevata influenza sul sistema stesso (Williamson e Hogan (2021), p. 15), plasmando l'evoluzione dell'università in direzione di modalità maggiormente orientate al mercato (Komljenovic and Robertson (2016), citato da Williamson e Hogan (2021), p. 14).

La quantificazione, la misurazione, la comparazione e la valutazione delle performance di università, staff e studenti hanno raggiunto dimensioni senza precedenti e si stanno espandendo rapidamente (Williamson, Bayne and Shay (2020), *ibidem*); le università impiegano i prodotti digitali per misurare e migliorare le prestazioni sia a livello organizzativo che individuale, e le aziende mettono a frutto i dati ottenuti da queste per perfezionare e sviluppare ulteriormente tali prodotti (Komljenovic (2019), *ivi*).

Assieme alla transizione digitale si sono manifestate diverse criticità, successivamente acuite dalla pandemia: tra queste si possono annoverare forme emergenti di esclusione digitale e di disuguaglianze educative, nonché la sempre maggiore influenza delle piattaforme nel determinare l'agenda delle politiche educative e il futuro dell'istruzione, dell'insegnamento e dell'apprendimento (Selwyn et al. (2020), *ivi*, p. 15).

Inoltre, l'adozione di tecnologie informatiche può «*rendere le cose più complicate, produrre effetti non intenzionali o dannosi nei sistemi sociali in cui sono implementati, riflettere la visione*

del mondo di un ristretto gruppo di specialisti informatici, o servire gli interessi commerciali delle società commerciali su altre esigenze» (Broussard (2018), ivi; TdA).

A riguardo, Williamson e Hogan (2020, p. 16), specificano: *«insieme, digitalizzazione e datificazione sono state implicate in forme di profilazione, distorsione e discriminazione basate sulla razza e sul genere, la riproduzione e l'esacerbazione delle disuguaglianze sociali ed economiche esistenti, l'ascesa di politiche estremiste, l'allargamento della polarizzazione politica, e minacce a diritti, libertà, sicurezza e lavoro»*.

Il combinato disposto di questi due processi, infatti, introduce logiche di mercato laddove principi e valori di pubblico interesse erano egemoni, di fatto depotenziandoli e ponendoli “in secondo piano”; tra questi vi sono la promozione della democrazia, dell'equità nell'accesso alle opportunità, così come dell'auto-realizzazione e della crescita economica.

Al loro posto, vengono introdotti l'automatizzazione dell'educazione, la massimizzazione della raccolta di informazioni e la codifica dell'apprendimento sulla scorta di un numero limitato di parametri e credenziali definiti dai dati (Zeide e Nissbaum (2018), citato da Williamson e Hogan (2021), p. 17). Se, da un lato, le tecnologie educative (“*EdTech*”) hanno il potenziale di promuovere lo sviluppo accademico, l'equità e forme migliorate di insegnamento e apprendimento (Bayne et al. (2020), ibidem), dall'altro possono portare alla standardizzazione educativa, al riduzionismo digitale per mezzo dei dati, alla privatizzazione dell'istruzione e alla delegittimazione e de-professionalizzazione dell'insegnamento (Williamson e Hogan (2021), p. 16).

Oltre alle difficoltà legate all'implementazione del digitale nell'educazione, vi è una ulteriore criticità, evidenziata da diversi autori (Knox (2019); Feenberg et al. (2019); Selwyn e Jandrić (2020)): si tratta dell'approccio prevalentemente *tecnicistico* e *apolitico* che è stato adottato durante questo processo. Il sistema universitario, infatti, si è mostrato impreparato di fronte alla brusca ed inaspettata insorgenza del Covid-19, e di ciò che ne è conseguito – con particolare riferimento alla necessità di traslare nel mondo virtuale attività che, fino a poco prima, venivano svolte in presenza.

Sebbene *«negli ultimi dieci anni, si sia assistito ad uno spostamento delle operazioni informatiche verso l'uso delle infrastrutture cloud»* (Fiebig et al., 2021, p.1; TdA), il mondo dell'università ha dovuto, *ex abrupto*, innestare inedite soluzioni digitali all'interno della propria struttura, per di più in un lasso di tempo estremamente breve.

A riguardo, Raffaghelli (2021, p. 1), afferma: *«La “datificazione” di processi e servizi è un fenomeno emergente [...]. Le due missioni principali dell'università sono entrate in pieno nelle dinamiche della datificazione a partire dai processi di digitalizzazione che hanno accompagnato la modernizzazione dell'università e più recentemente la pedagogia della pandemia»*.

L'autrice, inoltre, sottolinea come *«le due missioni dell'istruzione superiore – insegnamento e ricerca – abbiano attraversato diversi processi di digitalizzazione [...]». Nell'insegnamento, i dati sull'apprendimento e sugli studenti, raccolti su scale senza precedenti, hanno dato origine al data mining educativo e in particolare all'analisi dell'apprendimento (Siemens and Long 2011)»* (Raffaghelli et al., 2020, p. 1; traduzione dell'autore).

Raffaghelli e Stewart (2020, p. 439), citando De Rosa (2017), ricordano che *«la progressiva adozione dei “learning analytics” era originariamente diretta verso la comprensione di dati digitali al fine di creare interventi significativi, ma – con la comparsa dei predictive analytics – è avvenuto uno spostamento da una visione retrospettiva ad una previsionale»* (TdA).

L'entusiasmo iniziale per tali innovazioni, tuttavia, non era giustificato da un adeguato sostegno teorico, nonché empirico, riguardo alla loro capacità di poter apportare reali miglioramenti alla didattica tradizionale. Sempre Raffaghelli e Stewart (ibidem), infatti, spiegano come *«la ricerca abbia svelato ipotesi pedagogiche ingenuie, o addirittura scarse, riguardanti la capacità degli algoritmi di prevedere, supportare e indirizzare l'apprendimento, che erano connesse ad approcci techno-deterministici ai dati (Ferguson 2019; Perrotta and Williamson 2018; Selwyn 2019)»* (TdA).

Secondo la prospettiva del “determinismo tecnologico”, infatti, è il progresso tecnico a generare e plasmare i valori di una data società, determinandone – per l'appunto – lo sviluppo. Tale corrente di pensiero è stata aspramente criticata; tra queste, lo strumentalismo afferma che la tecnologia non rappresenta, necessariamente, un beneficio per l'umanità, e che l'adozione di mezzi sempre più evoluti non comporti un progresso ineluttabile.

Valore ed utilità della tecnologia, pertanto, dipendono da come questa viene incorporata ed impiegata all'interno delle società; essa è mero strumento nelle mani dell'uomo («Determinismo tecnologico», 2023).

A questa rinnovata consapevolezza, riguardo alle opportunità e ai limiti del progresso tecnico, hanno pertanto fatto seguito disincanto e disillusione; a ciò, tuttavia, non ha corrisposto una *revisione critica* delle pratiche digitali all'interno del mondo dell'istruzione, né una riflessione riguardo all'influenza crescente delle piattaforme all'interno della dimensione accademica.

Al contrario, l'adozione di tali soluzioni tecnologiche è avvenuta in maniera sostanzialmente acritica, mossa dalla necessità, o da interventi guidati da innovatori “ingenui”: a riguardo, infatti, Moore (2017, citato da Atenas et al., 2020; TdA) afferma: *«anche l'istruzione superiore (HE) è stata relativamente acritica di questi fenomeni di spietata quantificazione»*. Ciò è particolarmente rilevante poiché, secondo Noble (2018), (citato da Raffaghelli e Stewart, 2020, p. 435): *«I sistemi su cui ci affidiamo per la “digital scholarship” e l'istruzione sono sempre più datificati, anche quando*

i pregiudizi razziali e di genere incorporati nel processo decisionale algoritmico diventano sempre più evidenti».

Oltre al tema della riproduzione di bias discriminatori, incorporati all'interno dei sistemi algoritmici, vi sono le questioni legate alla *privacy* e al *controllo* realisticamente esercitabile su infrastrutture che, nell'hardware e nel software, sono completamente private (Fiebig et al., 2021, p. 1), e viceversa all'opportunità che queste *tecno-strutture* hanno di condizionare e limitare la *libertà accademica* al fine di tutelare i propri interessi (ibidem, p. 2).

1.4. Definizioni di competenza digitale nel contesto internazionale e in Italia: un excursus

1.4.1. I precursori della *Data Literacy: Statistical Literacy e Numeracy*

Da quanto emerso nel corso della precedente trattazione, emerge con una certa chiarezza la rilevanza delle capacità informatiche nel contesto odierno, sia esso privato o professionale. Calvani et al. (2009, p. 7) propone una sintetica definizione di ciò che si può intendere per competenza digitale: «*La competenza digitale consiste nel saper esplorare ed affrontare in modo flessibile situazioni tecnologiche nuove, nel saper analizzare selezionare e valutare criticamente dati e informazioni, nel sapersi avvalere del potenziale delle tecnologie per la rappresentazione e soluzione di problemi e per la costruzione condivisa e collaborativa della conoscenza, mantenendo la consapevolezza della responsabilità personali, del confine tra sé e gli altri e del rispetto dei diritti/doveri reciproci*».

All'interno di questa concettualizzazione risaltano gli elementi di appropriazione attiva e critica dell'ambiente digitale, attraverso il quale raccogliere, scambiare e generare informazioni in maniera autonoma, competente e responsabile. Tuttavia, è da notare come la questione delle *digital skills* (competenze digitali) sia stato affrontato, almeno inizialmente, secondo una prospettiva prettamente tecnicistica, che trascura l'acquisizione individuale e critica degli strumenti informatici a supporto della promozione di quella *agentività* che solo la consapevolezza e la riflessione autonoma possono offrire.

Innanzitutto, prima di esaminare la trasformazione del concetto di competenza digitale nei diversi contesti (ai fini della nostra trattazione si è scelto di discutere l'ambito nazionale italiano e quello sovranazionale europeo), è necessario indagare i suoi precursori, ovvero i concetti di *statistical literacy* (“*alfabetizzazione statistica*”) e *numeracy* (traducibile in italiano con la perifrasi “*far di conto*”).

All'interno della moderna società industriale, fondata sul sapere scientifico, sul progresso tecnico e sulla quantificazione, è diventata presto evidente l'importanza della diffusione e dell'appropriazione degli strumenti matematici, almeno ad un livello di base, all'interno della popolazione; successivamente, con l'evoluzione sociale e culturale nella direzione della democratizzazione e dell'informazione di massa, si è sostenuta l'esigenza di divulgare una maggiore conoscenza collegata alla statistica, nonché al processo di ragionamento ad essa sottostante.

Raffaghelli (sd, p. 4) delinea l'evoluzione storica del termine *numeracy*, ricordando come tale espressione sia stata introdotta nel 1959 all'interno del "Rapporto Crowther" nella cornice dell'alfabetizzazione fondamentale nel Regno Unito.

Come riportato dall'autrice (Raffaghelli, sd, p. 4), la popolarità del concetto di competenza matematica cresce progressivamente – conoscendo il suo apice negli anni '90 – quando il riconoscimento dell'importanza delle discipline STEM (*"Science, Technology, Engineering and Mathematics"* – *"Scienza, Tecnologia, Ingegneria e Matematica"*) per la crescita economica e la competitività, porta alla sua introduzione all'interno dei programmi d'istruzione nazionali; nonché, come logica conseguenza, alla necessità di valutare l'efficacia di tali programmi.

Nascono così prove standardizzate come il PISA (*"Performance International Students' Assessment"*, OECD (2017), come riportato da Raffaghelli (sd, p. 4) per gli studenti dell'obbligo scolastico, e il PIAAC (*"Programme for the International Assessment of Adult Competencies"*, OECD (2012), citato da Raffaghelli (sd, p. 4) per gli adulti.

All'interno di questi progetti di valutazione delle competenze, il concetto di *numeracy* viene esteso all'applicazione di metodi matematici per la risoluzione di problemi quotidiani; come sottolinea Raffaghelli (sd, p. 5): *«Nelle più recenti definizioni PISA e PIAAC risulta evidente lo sforzo di andare oltre il concetto di conoscenza matematica come dimensione acquisita solo attraverso ambienti formali e trasmissivi, dove la matematica risulta assai astratta e lontana dalla risoluzione contestuale di problemi»*. Sempre Raffaghelli (2019, p. 106), evidenzia il superamento del *"concetto di competenze matematiche (come conoscenza di procedure formali all'interno di operazioni aritmetiche e algebriche)"* nella direzione di *"concetti applicati in ambienti autentici che richiedono competenze di problem solving»*.

Tuttavia, le limitazioni insite nel concetto originario – sostanzialmente riferito alle abilità aritmetiche fondamentali (somma, sottrazione, moltiplicazione e divisione), e sulle operazioni cognitive più semplici di misurazione applicata a dimensioni e tempi (Raffaghelli, sd) – determinano l'incompatibilità con la competenza statistica, ben più complessa e sofisticata.

Pertanto, autori come Gould (2017), citato da Raffaghelli (2019, p. 106), rilevano come *«nella società contemporanea è necessario acquisire competenze per interagire con le informazioni statistiche»*, poiché esse sono *«sempre più presenti in tutti i tipi di informazione, come report e articoli giornalistici, comunicazione politica e social media»*.

Secondo l'autore, la *statistical literacy* deve comprendere, pertanto, anche la comprensione della comunicazione basata su dati statistici (Raffaghelli, sd, p. 5; 2019, p. 106) e delle sue dimensioni essenziali – tra queste, di particolare rilevanza sono:

- Conoscere e capire i concetti di campionamento e di errore campionario;

- L'essere in grado di impiegare la statistica descrittiva per sintetizzare i dati;
- Il saper riconoscere la differenza tra la statistica descrittiva ed inferenziale;
- Il comprendere la differenza tra correlazione e causalità;
- Capire i processi alla base delle rappresentazioni grafiche di informazioni statistiche.

Tuttavia, con l'emergere di nuove forme di elaborazione di dati, basate su algoritmi bayesiani anziché su procedure inferenziali (Gould (2017), citato da Raffaghelli (sd, p. 5), e su raccolte dati “*massive, continue e crowd-sourced*” (ibidem), anche il concetto di *statistical literacy* appare ormai superato; gli algoritmi, infatti, utilizzando variabili *proxy* riescono a estrapolare pattern di comportamento abituali e, sulla base di questi, realizzare previsioni su eventi futuri (Raffaghelli, 2019, p. 106).

Pertanto, si è ritenuto necessario elaborare una definizione che meglio descrivesse le competenze necessarie ad operare con abilità e consapevolezza nel contesto del digitale: tale definizione è quella di alfabetizzazione digitale (“*data literacy*”).

Come sottolinea Raffaghelli (2019, p. 7), citando Risdale (Ridsdale et al., 2015), essa viene definita «come “l’abilità di raccogliere, gestire, valutare e usare dati in contesti applicati attraverso una lente critica”».

Sempre Raffaghelli (ibidem) descrive le prerogative fondamentali della *data literacy*, di cui alcune più tipicamente – anche se non esclusivamente – tecniche, come l'estrazione, la gestione e l'elaborazione delle informazioni, ed altre legate alla consapevolezza individuale e alla responsabilità verso l'uso dei dati, ovvero all'«*approccio etico e critico al trattamento dei dati*»; attorno a questi concetti fondamentali si identificano, quindi, una serie di competenze specifiche, che l'autrice, citando una rassegna compiuta da Maybee e Zilinski (2015) (ivi), elenca; esse possono essere distinte in due macro-aree:

- Capacità di riflessione e autoriflessione rispetto ai dati:
 - Consapevolezza nell'uso dei dati, intesa come capacità di comprenderli e di comprenderne la funzione all'interno della società contemporanea;
 - Impiego etico dei dati, che consiste nel saper discernere tra diverse fonti di dati e nella consapevolezza rispetto ai rischi potenziali legati al loro uso e trattamento;
- Competenza nella gestione, trattamento, uso, conservazione e comunicazione dei dati:
 - Capacità di gestione dei dati, pianificando come affrontare ogni questione inerente alla conservazione di questi (sicurezza, integrità, organizzazione e analisi);
 - Preservazione dei dati, ovvero abilità e cognizione delle modalità di uso, riuso e conservazione dei dati a lungo termine;

- Competenza nell'accedere ai dati, intesa come la capacità di identificare una fonte di dati (per es., un dataset) e di comprenderne il contenuto;
- Essere coinvolti nell'uso dei dati: si tratta della capacità di agire attivamente nei confronti dei dati, attraverso processi di valutazione, organizzazione, analisi e interpretazione, anche ai fini della presa di decisioni;
- Comunicare i dati e attraverso i dati, grazie a rappresentazioni grafiche e a strumenti di sintesi.

Come si evince dalla presente trattazione, il concetto di alfabetizzazione digitale ha subito un'evoluzione lunga e articolata, avendo le sue radici nelle competenze aritmetiche di base, arricchite da una preparazione matematica e statistica più approfondita e ad un approccio critico ai dati. Di seguito, invece, viene presentata la trasformazione di tale nozione all'interno del contesto europeo.

1.4.2. Le competenze digitali in Europa

In parallelo con la digitalizzazione della società, è cresciuta la necessità di una alfabetizzazione digitale di massa e, ancora prima di questa, di definire in maniera univoca la natura di questa competenza. Il termine *competenza digitale* fa la sua apparizione già nel 2006, all'interno del framework "Competenze chiave per l'apprendimento permanente" stilato da Parlamento e Consiglio europei (*Competenze chiave per l'apprendimento permanente*, 2006), incentrato sulle otto competenze chiave per la piena partecipazione dei cittadini alla vita sociale, economica e politica dell'Unione Europea.

In primo luogo, quindi, le competenze vengono più generalmente definite come «una combinazione di conoscenze, abilità e attitudini, in altre parole, sono composte da concetti e fatti (ad esempio, conoscenze), descrizioni di abilità (ad esempio, la capacità di svolgere processi) e atteggiamenti (ad esempio, un'attitudine, un approccio all'azione). Le competenze chiave vengono sviluppate durante l'intero arco di vita». (Vuorikari et al., 2022, p. 3; traduzione dell'autore).

Più specificatamente, le competenze chiave sono quelle abilità necessarie all'individuo per «la realizzazione personale, uno stile di vita sano e sostenibile, l'occupabilità, la cittadinanza attiva e l'inclusione sociale» (ibidem, p. 6; TdA) e, tra queste, possiamo annoverare la *Digital competence*, che in prima battuta viene intesa come il «saper utilizzare con dimestichezza e spirito critico le tecnologie della società dell'informazione (TSI) per il lavoro, il tempo libero e la comunicazione. Essa è supportata da abilità di base nelle TIC: l'uso del computer per reperire, valutare, conservare, produrre, presentare e scambiare informazioni nonché per comunicare e partecipare a

reti collaborative tramite Internet» (Calvani et al., 2009); pertanto, all'introduzione del concetto ha fatto seguito, nel 2013, il *DigComp*, ovvero il “*Digital Competence Framework for Citizens*” (“Quadro di Competenza digitale per la cittadinanza”).

Aggiornato nel 2016 all'edizione 2.0, al suo interno venivano identificate cinque aree di competenza (Alfabetizzazione ai dati e all'informazione, Comunicazione e Collaborazione, Creazione di contenuti digitali, Sicurezza, Problem Solving), tralasciando tuttavia le competenze legate alla «*comprensione, analisi e trattamento dei dati*» (Raffaghelli, 2021, p. 6); queste ultime verranno aggiunte solamente l'anno dopo, nella versione 2.1 del Quadro, che inserisce la *data literacy* all'interno della più ampia area dell'*information literacy* (ibidem).

Seguendo la rapida trasformazione del panorama tecnologico ed informativo, già nel 2022 giunge una nuova revisione del *framework*, il “*Digital Competence Framework for Citizens*” (“*DigComp 2.2*”). Come viene ricordato al suo interno, «*La competenza digitale è una delle competenze chiave per l'apprendimento permanente. [...] La competenza digitale implica l'uso sicuro, critico e responsabile delle tecnologie digitali per l'apprendimento, il lavoro e la partecipazione alla società. Comprende l'alfabetizzazione ai dati e all'informazione, la comunicazione e la collaborazione, l'alfabetizzazione mediatica, la creazione di contenuti digitali (compresa la programmazione), la sicurezza (compreso il benessere digitale e le competenze relative alla sicurezza informatica), le questioni relative alla proprietà intellettuale, al problem solving e al pensiero critico*». (Vuorikari et al., 2022, p. 6; TdA).

Inoltre, aggiunge Vuorikari et al. (ibidem, TdA): «*Tutte le competenze chiave sono complementari e interconnesse tra loro. In altre parole, le competenze essenziali per un settore sosterranno lo sviluppo delle competenze in un altro. Ciò riguarda anche il rapporto tra le competenze digitali e le altre competenze chiave*».

Ad una più approfondita elaborazione teorica, quindi, vengono associati esempi concreti in cui è necessario applicare le competenze digitali in maniera critica e consapevole: tra queste, vengono ricompresi fenomeni di disinformazione e mistificazione su social media e siti d'informazione; la consapevolezza rispetto alla datificazione di prodotti e servizi, con particolare riferimento allo sfruttamento dei dati personali degli utenti; conoscenza delle tecnologie emergenti (Intelligenza Artificiale, *Internet-of-Things*, Robotica e Realtà Aumentata) e delle implicazioni ad esse correlate; sostenibilità ambientale e nuovi contesti emergenti (lavoro da remoto o ibrido) (ivi, p. 5; TdA).

Anche a livello europeo, quindi, si osserva un graduale mutamento di paradigma: ad un'idea di competenza digitale che, sebbene ritenuta fondamentale per la piena partecipazione alla vita collettiva, veniva intesa come abilità perlopiù tecnica, ha fatto seguito un ampliamento che ha ricompreso capacità di riflessione, valutazione, e presa di decisione personale rispetto ai contesti

digitali, ed una interazione consapevole, organica e critica, con dati ed informazioni di varia natura, siano essi in forma numerica, visiva o testuale, nonché provenienti da fonti diverse (siti di social media, informazione, ecc...).

Nel paragrafo seguente, invece, viene introdotto e descritto il panorama italiano delle competenze digitali, all'interno del quale si cerca, non senza difficoltà, di dare impulso alla diffusione delle *digital skills*.

1.4.3. Le competenze digitali in Italia

Sulla scia delle ricerche inerenti alla *numeracy*, il legislatore italiano si era concentrato su azioni rivolte ad un migliore insegnamento ed apprendimento delle abilità logico-matematiche (Trincherò, 2016, citato da Raffaghelli, 2019, p. 108).

La questione della *data literacy* restava in secondo piano, sebbene specifiche iniziative (Indicazioni Ministeriali del 2012; gruppo di lavoro sui Big Data del 2016; ibidem) riportassero l'attenzione – rispettivamente:

- all'insegnamento della statistica come mezzo per comprendere la complessità della realtà contemporanea;
- alla necessità di elaborare percorsi d'istruzione e d'apprendimento, lungo l'intero arco di vita, che avessero nella *data literacy* il loro fulcro, o quantomeno ne riconoscessero la rilevanza: a riguardo, veniva sottolineato il contributo degli istituti di Alta Formazione (ITS) ai fini della preparazione di professionalità innovative ed emergenti come quelle del *data scientist*, nonché si dedicava attenzione all'elaborazione di proposte d'intervento per integrare i consueti percorsi scolastici con attività indirizzate alla comprensione e manipolazione di dati digitali (ivi, p. 109).

Nel rapporto del 28 luglio 2016 redatto dal gruppo sui Big Data del Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR), inoltre, veniva sottolineata la necessità di integrare l'istruzione universitaria, a tutti i livelli, con la formazione alla scienza dei dati ("*data science*"), nonché ad accrescere la cura, a livello scolastico, nell'insegnare la *data literacy*, all'interno del più grande quadro della formazione alle competenze digitali; inoltre, evidenziava l'esigenza di promuovere la ricerca inerente ai Big Data all'interno dei progetti nazionali, e di migliorare le capacità di comunicazione del MIUR rispetto al suo patrimonio informativo, come punto di partenza per la governance e lo sviluppo dell'intero sistema scolastico e della ricerca in ambito educativo (Raffaghelli, 2017, p. 14).

Citando il Rapporto MIUR del 2016, Raffaghelli (2017, p. 309), sottolineava come tra il 2014 e il 2017, in *almeno* 4 università venivano proposti corsi di laurea, sia triennali che magistrali, con riferimenti ai Big Data; inoltre, evidenziava la presenza di 4 lauree magistrali e 11 curricula dedicati interamente ai Big Data, 5 lauree magistrali e 7 curricula rivolti alla cybersicurezza (“*cybersecurity*”), nonché numerosi insegnamenti su queste tematiche all’interno dei programmi accademici. Tali corsi, ed i relativi insegnamenti, erano perlopiù presenti all’interno dei dipartimenti di informatica e di ingegneria informatica, e sostanzialmente assenti in tutti gli altri (ibidem).

Sempre nel 2016 – all’interno del *Piano Nazionale per la Scuola Digitale* – veniva lanciato il *Bando sui curricula digitali*, con l’intenzione di sollecitare l’innovazione e la progettazione di percorsi di formazione alle competenze digitali, con particolare riferimento a Big e Open Data (ibidem, p. 15).

Rispetto al tema della diffusione di cultura e pratiche basate sui dati, è da segnalare il progetto “A Scuola di Open Coesione” (ASOC) il quale, anche grazie al supporto dell’osservatorio europeo della rete EDIC (Centri d’Informazione Europe Direct), mira a educare all’uso dei dati aperti come risorsa educativa (OER, “*Open Educational Resource*”) e a costituire una comunità di apprendimento professionale formata dagli insegnanti e dagli esperti ASOC. Il tutto, rivolto a stimolare l’attivismo civico e il monitoraggio dell’utilizzo dei fondi pubblici (ivi, p. 12).

Nonostante tali iniziative – sicuramente meritorie e degne di nota – all’interno del contesto italiano sembra mancare un progetto organico di integrazione della cultura dei dati e del digitale con i curricula educativi tradizionali, nonché un piano di rinnovamento delle competenze professionali, in ambito pubblico e privato, sia in un’ottica di *data literacy* che in quello – certamente più ampio – delle *digital skills*.

Tale dinamica conosce un momento di rottura profonda con l’insorgenza della pandemia da Covid-19 e la successiva introduzione delle misure di contenimento, in particolare le chiusure forzate di attività pubbliche e commerciali. La scuola italiana si dimostra impreparata a far fronte a tale evento, soprattutto per quanto concerne le abilità informatiche di studenti e educatori (Carretero Gomez et al., 2021; citato da Ranieri, 2022, p. 9).

Le pesanti ripercussioni, economiche e sociali, del periodo pandemico hanno l’effetto di spingere le istituzioni europee a varare piani di sostegno (all’interno della cornice del Next Generation EU) alle nazioni più colpite; tali piani, senza precedenti nella storia del continente, sono stati finanziati per la prima volta attraverso debito pubblico comune, e sono focalizzati su investimenti e progetti di riforma volti a colmare le lacune degli stati e delle rispettive economie, indirizzandole verso percorsi di crescita sostenibili e duraturi.

Il Piano Nazionale di Ripresa e Resilienza italiano – la cui elaborazione è terminata ad aprile 2021 (PNRR Italia, 2022, p. 16) – trova uno dei suoi capisaldi nella transizione digitale, a cui – come stabilito dalle istituzioni europee – va destinato obbligatoriamente almeno il 20% delle risorse del piano (ibidem).

A testimoniare il ritardo italiano nelle tecnologie digitali, nonché l'improcrastinabile necessità di attuare investimenti e progetti a molteplici livelli, è l'importante quota ad esse rivolte, circa 48 sui 191,5 miliardi del piano (ivi, p. 15); attraverso lo sviluppo infrastrutturale, la digitalizzazione delle amministrazioni statali e parastatali, la promozione delle competenze digitali nell'istruzione e nel lavoro, il PNRR cerca di colmare i ritardi accumulati dal contesto italiano.

Per quanto concerne la dimensione dell'istruzione e della formazione, particolare attenzione viene rivolta (Missione 4) al rafforzamento degli ITS (Riforma 1.2) e all'insegnamento delle discipline STEM (Scienza, Tecnologia, Ingegneria e Matematica) (Investimento 3.1), nonché alla formazione del personale scolastico e universitario alle competenze digitali (Investimento 2.1), al fine di operare la transizione digitale del sistema d'istruzione nel suo insieme (ivi, p. 190).

Il riferimento alla formazione alle competenze digitali dei docenti è di sicura rilevanza, poiché, come osserva Ranieri (2022, p. 10), anche i nuovi insegnanti ricevono un'alfabetizzazione digitale incerta e frammentaria: a titolo di esempio, l'autrice cita i Corsi di Studio a Ciclo Unico di Scienze della Formazione Primaria, i quali non prevedono un *syllabus* che tenga in adeguata considerazione le nuove competenze necessarie ad una didattica sempre più digitale e connessa.

Pertanto, all'interno del Piano si sottolinea la volontà di creare un «*sistema [nazionale, NdA] multidimensionale per la formazione continua dei docenti*» sulle *digital skills* (ivi), per integrare i processi didattici e di apprendimento in conformità con i quadri di riferimento europei sulle abilità informatiche (DigComp 2.1 per gli studenti, e DigCompEdu per i docenti) (ivi). Tale sistema dovrebbe, nelle intenzioni, essere composto da – almeno – tre “*Teaching and Learning Centers*” (TLC) per la formazione dei docenti di tutti i gradi d'istruzione, e di tutte le discipline, all'insegnamento delle competenze informatiche, e da tre “*Digital Education Hubs*” (DEH), aventi lo scopo di migliorare la capacità del sistema universitario di trasmettere competenze digitali a staff e studenti (ivi, p. 192-193).

A livello di ateneo, si segnala il progetto di *faculty development* “Teaching4Learning@Unipd” dell'Università di Padova, il quale pone tra i suoi obiettivi il miglioramento della qualità dell'insegnamento, anche attraverso lo sviluppo della didattica tramite strumenti digitali (Fedeli & Tino, 2019; De Rossi e Fedeli, 2022).

Degno di nota è anche il progetto “*PRODID*”, “Preparazione alla professionalità Docente e Innovazione Didattica”, all'interno del quale sono state avviate esperienze di *faculty development* in

diversi atenei italiani (Felisatti et al., 2020, p. 33), e dal quale è nata l'associazione ASDUNI, “*Associazione italiana per la promozione e lo sviluppo della didattica, dell'apprendimento e dell'insegnamento in Università*”, che opera nella direzione di sviluppare la professionalità docente e promuovere la qualità nell'istruzione superiore (ibidem, p. 20).

Infine, a livello nazionale, è da evidenziare il gruppo di lavoro SIPED (Società Italiana di Pedagogia), il cui fine è promuovere ricerca e innovazione didattica universitaria, elaborando un quadro teorico di riferimento e modelli per il *faculty development* (*Faculty development e didattica universitaria*, 2023).

Questa breve rassegna – che non può considerarsi completamente esaustiva – descrive un panorama italiano che, come già in precedenza sottolineato, evolve con maggiore lentezza rispetto a quello sovranazionale, il quale dimostra una sensibilità ed una prontezza ben differenti rispetto al tema delle competenze informatiche e, tra queste, alla *data literacy*.

Il capitolo successivo – Alfabetizzazione ai dati per il Faculty Development – si occupa nel dettaglio della trasformazione in atto nell'istruzione superiore; l'avvento della pandemia ha, infatti, imposto alle istituzioni accademiche un cambio di passo radicale rispetto all'introduzione delle tecnologie digitali, e alla loro applicazione per la didattica e l'apprendimento.

Esse – benché impreparate – si sono dirette giocoforza verso forme d'insegnamento ibride e a distanza, non senza criticità e contraddizioni.

2. Alfabetizzazione ai dati per il Faculty Development

2.1. Competenze emergenti nel contesto dell'istruzione superiore: una revisione della letteratura

I fenomeni di crescente *digitalizzazione*, *piattaformizzazione* e *datificazione* dell'università hanno messo in luce la questione delle competenze digitali all'interno della comunità accademica.

In primo luogo, Raffaghelli e Stewart (2020, p. 438; traduzione a cura dell'autore), fanno notare che con il concetto di *Faculty Development*, di solito, si fa riferimento «*all'apprendimento professionale all'interno dell'istruzione superiore*», ricordando come, da un punto di vista storico, questo filone di ricerca si sia dedicato all'efficacia dei programmi di sviluppo professionale e, più recentemente, alle competenze richieste dall'insegnamento tramite mezzi digitali. Tuttavia, molti di questi studi sono stati criticati a causa di uno scarso fondamento teorico o concettuale, e per la mancanza di un legame con la pratica (Webster e Wright, 2009, citato da Raffaghelli e Stewart, 2020, p. 438).

Alcuni autori, viene fatto notare, concepiscono i docenti come soggetti che apprendono all'interno di un contesto sociale e situato, nel tempo e nello spazio (Boud, 1999; Cox, 2004; citati da Raffaghelli e Stewart, 2020, p. 438), evidenziando i cambiamenti che le attività di formazione introducono a livello organizzativo e professionale (Raffaghelli e Stewart, 2020, p. 438). In riferimento a ciò, Raffaghelli e Stewart propongono un *faculty development* che sia «*radicato nell'apprendimento professionale e connesso, e rivolto a sviluppare le conoscenze necessarie per gestire le culture dei dati nell'istruzione superiore*» (ibidem, p. 436; TdA).

Tuttavia, essi affermano, gli attuali modelli di alfabetizzazione ai dati si occupano di set di abilità specifiche, perlopiù tecniche, evitando di considerare questioni controverse come la profilazione, l'etica nell'uso dei dati, e le questioni di potere ad esse correlate; inoltre, essi riducono «*un insieme aggrovigliato di sistemi ai segmenti e ai risultati che, di questi, possono essere osservati o misurati*» (ivi; TdA). Come conseguenza, si può constatare che la letteratura sul tema abbia affrontato la questione da un punto di vista prettamente tecnico e/o di gestione istituzionale, secondo una logica che vede la *data literacy* nell'ambito dell'istruzione come «*la capacità di raccogliere o estrarre dati sull'educazione con l'intenzione di offrire un supporto alle decisioni istituzionali che riguardano gli studenti*» (Hartong & Förschler, 2019, citato da Raffaghelli e Stewart, 2021, p. 2; TdA).

Tale «*prospettiva tecnica*» è risultata, a lungo, prevalente all'interno della ricerca accademica: a sostegno di questa affermazione, Raffaghelli e Stewart (2020), in una revisione sistematica della letteratura scientifica riguardante l'alfabetizzazione digitale degli educatori, sottolineano come la

maggior parte dei *paper* analizzati – 78 su 137 – adottassero una “*cornice strumentale*” (“*instrumental framing*”) nel trattare l’argomento, considerando le competenze informatiche dei docenti da un punto di vista prettamente tecnico; tralasciando, invece, aspetti quali la valutazione critica dello strumento impiegato, o la consapevolezza rispetto alle implicazioni etiche – come il rispetto della riservatezza degli studenti – conseguenti all’uso di Learning Analytics e altri sistemi di “automazione didattica”.

La capacità di gestire in sicurezza i dati (“*data safety*”) era il secondo argomento più discusso, con 35 articoli, mentre meno numerosi erano gli articoli inerenti all’uso dei dati a supporto degli studenti (5), al data hacking (11), e ad approcci critici ai dati (7). Lo scenario che si delinea mette in rilievo la consolidata tendenza della letteratura scientifica ad occuparsi di abilità e competenze tecniche specifiche, anziché di proporre una visione onnicomprensiva dell’alfabetizzazione ai dati (Raffaghelli e Stewart, 2020, p. 444).

Da sottolineare, ai fini della nostra trattazione, l’ambito epistemologico e il livello d’istruzione a cui i *paper* facevano riferimento: sui 136 articoli complessivi, 33 erano rivolti allo sviluppo professionale degli insegnanti, 27 alla ricerca sui processi di apprendimento, 14 alle Scienze dell’Educazione; il Faculty Development, viceversa, era un argomento poco dibattuto, con soli 8 paper inerenti al tema.

Infine, per quanto concerne il livello d’istruzione, 48 erano ascrivibili all’Istruzione Superiore (“*Higher Education*”), e 32 alla Formazione dei docenti (ibidem).

Pertanto, osservano gli autori, esiste una lacuna importante riguardo al tema dell’approccio critico ai dati nella dimensione dell’istruzione superiore; oltre a ciò, evidenziano la necessità di superare una prospettiva meramente tecnica, orientata alle performance, ai risultati e all’efficienza, all’interno della discussione riguardo le competenze digitali e l’impiego dei dati educativi (ivi).

A riguardo, quindi, propongono un nuovo inquadramento concettuale (Raffaghelli e Stewart, 2021, p. 3). Esso si basa sulla contrapposizione “complicato/complesso” – già avanzata da Snowden e Boone (2007, ibidem) all’interno del *Cynefin framework* – per distinguere tra i vari approcci di *Data Literacy*: i modelli di competenze tecnico-strumentali ricadono nel dominio del “complicato”, all’interno del quale vi è la convinzione che esistano “risposte giuste identificabili” («*knowable right answers*») al problema dello sviluppo delle competenze digitali degli educatori, le quali possono essere applicate attraverso l’uso dell’*expertise* (ibidem, p. 440).

Viceversa, l’orientamento al “complesso” riconosce l’incertezza intrinseca all’agire umano e, nella fattispecie, alla ricerca di soluzioni ai problemi: se una visione “complicata” della realtà fa riferimento a schemi già noti per interpretarla, ed utilizza soluzioni comprovate ai problemi che si presentano, una mentalità orientata al “complesso” richiede risposte inedite ad eventi, percepiti

nella loro molteplicità e globalità, come frutto di articolate reti di relazioni (ivi). Le narrazioni inerenti alla datificazione, al contrario, fanno leva sulle idee di “certezza” e di “misurabilità”, per il potere e il fascino che esse possiedono (ivi, p. 3).

Di conseguenza, sostengono gli autori, l'istruzione superiore, nel suo insieme, potrebbe trarre vantaggio da un'alfabetizzazione ai dati degli educatori che sia *critica ed orientata alla complessità* (ivi); nello specifico, affermano che «*un approccio pedagogico, che sia complesso e critico, alla progettazione del faculty development per l'alfabetizzazione ai dati dovrebbe includere quadri di competenze, contesti informativi, risorse, strategie e politiche istituzionali, studio di casi, coaching e pratiche accademiche personali in relazione ai sistemi di dati. Dovrebbe includere anche un'analisi critica della governance dell'istruzione superiore e un esame dei regimi di potere e conoscenza rappresentati dal dominio del complicato e dalle narrazioni sul pensiero computazionale*» (Raffaghelli e Stewart, 2020, p. 449).

Atenas et al. (2020), nello specifico, affermano che i programmi di formazione ai dati dovrebbero includere l'etica dei dati, aspetti politici inerenti alla raccolta e l'uso dei dati, leggi e regolamenti riguardo uso, presentazione e conservazione dei dati, la gestione dei dati, l'analisi dei dati, teoria e metodi di narrazioni dei dati, e la visualizzazione dei dati.

Il tutto, nel tentativo di integrare tra loro l'alfabetizzazione statistica, mediatica e politica per lo sviluppo di *soft-skills* utili durante l'intero arco di vita, favorire le capacità di elaborazione critica delle informazioni da parte degli studenti e, infine, rendere questi ultimi consapevoli del loro ruolo di attori sociali, capaci di promuovere i valori di democrazia e di giustizia sociale.

Il paragrafo seguente descrive l'attuale condizione – rispetto alla *data literacy* – dei docenti universitari, esaminando aspetti teorici ed operativi di questa, nonché analizzando la situazione all'interno di contesti specifici.

2.2. L'alfabetizzazione ai dati dei docenti universitari

La questione della *data literacy* dei docenti universitari appare una questione ancora irrisolta: nel 2017 viene varato, con l'intenzione di porre rimedio alla mancanza di un riferimento teorico ed operativo alla formazione digitale di docenti, il quadro di riferimento dell'Unione Europea riguardante le competenze digitali degli educatori, il *DigCompEdu* (Ranieri, 2022, p. 2).

Al suo interno vi si trova sia un «*inquadramento concettuale per la definizione delle competenze digitali degli insegnanti, ma anche indicazioni su attività operative che possono essere realizzate nel quadro della competenza indicata* (Lucas et al. 2021)» (ibidem, p. 6).

Esso definisce sei aree di competenza che ogni docente dovrebbe padroneggiare, e 22 competenze in totale. Ranieri (2022, pp. 6-7) le elenca brevemente:

- Area 1, “Coinvolgimento e valorizzazione professionale”: capacità di utilizzare la comunicazione digitale per scopi organizzativi, la collaborazione e la crescita personale;
- Area 2, “Risorse digitali”: capacità di individuare, condividere e creare risorse educative digitali;
- Area 3, “Pratiche di insegnamento e apprendimento”: saper organizzare ed impiegare i mezzi digitali all’interno dei processi di insegnamento e apprendimento;
- Area 4, “Valutazione dell’apprendimento”: capacità di applicare strumenti e strategie digitali ai fini del miglioramento delle pratiche di valutazione;
- Area 5, “Valorizzazione delle potenzialità degli studenti”: impiego delle tecnologie digitali per promuovere inclusione, personalizzazione dell’apprendimento e il coinvolgimento attivo degli studenti;
- Area 6, “Favorire le competenze digitali degli studenti”: capacità di supportare gli studenti nell’uso, creativo e responsabile, dei mezzi digitali nelle attività di informazione, creazione di contenuti, comunicazione, benessere personale e risoluzione di problemi.

Tuttavia, evidenzia Raffaghelli (2021, p. 8), la competenza di alfabetizzazione ai dati non è integrata in maniera chiara ed organica all’interno del framework: infatti, sostiene l’autrice, *«in nessuna delle sei aree proposte [...], viene considerata nello specifico la capacità di utilizzare dati per l’attività di programmazione o progettazione, né vengono identificati i dati come risorse educative per informare la didattica, o migliorare la valutazione»* (ibidem).

Pertanto, al fine di colmare tale lacuna (Raffaghelli, 2019), viene proposto un *«quadro di orientamento per l’alfabetizzazione ai dati nella didattica universitaria»*, in particolare attraverso l’uso di Open Data (Raffaghelli, 2021, p. 9). Secondo l’autrice, tale approccio *«può risultare d’interesse per esplorare e definire le aree di pratica professionale del docente universitario, specifiche e trasversali, volte a far fronte alla datificazione e a cogliere gli aspetti costruttivi delle pratiche data-driven»* (ibidem). Anche sulla base dei riferimenti citati, Raffaghelli (2023, p. 274) propone – per meglio definire l’alfabetizzazione digitale degli educatori – il *“Quadro per una rinnovata professionalità accademica post-digitale”*, suddiviso nei quadranti “Innovazione educativa attraverso pratiche basate sui dati” (1), “Scienza e pratiche educative aperte” (2), “Ripensamento della pratica educativa quantificata” (3), “Analisi critica e partecipativa dell’uso dei dati educativi” (4), le quali vengono poste all’interno di uno spazio quadridimensionale che varia tra gli estremi “pubblico-personale” (asse trasversale) e “proattivo-reattivo” (asse orizzontale).

Lo schema delineato definisce un processo ricorsivo che ha inizio nel primo quadrante, all'interno del contesto della pratica professionale e che, attraverso l'uso dei dati come strumento d'apprendimento (secondo quadrante) e come supporto alla valutazione (terzo quadrante), conduce all'*empowerment* degli studenti e alla riflessione – critica e collaborativa – sui dati, per giungere infine ad una alfabetizzazione digitale da, e per, gli studenti.

Tuttavia, al di là dell'astrazione concettuale, l'autrice ci ricorda come le pratiche basate sui dati siano legate ad uno specifico contesto; pertanto, in quanto legata a fattori culturali ed organizzativi, la competenza digitale non può essere definita in maniera del tutto univoca (Raffaghelli, 2023, p. 276).

In uno studio di caso sulle competenze digitali degli educatori in formazione, Raffaghelli (2019, p. 116) evidenzia come all'inizio dell'intervento gli studenti riportavano, prevalentemente, un livello di competenza in *data literacy* pari a zero (ovvero, assenza di competenza), o comunque di base. Nel dettaglio, le abilità maggiormente diffuse erano quelle di "Ricerca Dati" e "Collaborazione con i Dati", mentre le capacità di "Estrazione Dati" erano sostanzialmente nulle (ibidem, p. 117).

Inoltre, osservava come la maggior parte dei partecipanti fossero interessati ad una eventuale formazione di base alla conoscenza dei dati, mentre solo uno si dimostrava aperto all'idea di affrontare un percorso per acquisire un livello di competenza avanzato.

L'autrice, riguardo quest'ultimo aspetto, ha ipotizzato che l'attuale identità professionale dell'educatore sia percepita come distante rispetto al tema dell'alfabetizzazione ai dati (ivi, p. 118).

Raffaghelli et al. (2021, pp. 67-68), in uno studio sulle pratiche dei dati all'interno del personale dell'Università di Padova, riportava come queste fossero incentrate sull'uso dei dati per la riprogettazione didattica e curricolare, nonché legate all'uso di dati ottenuti da verifiche in itinere per il monitoraggio didattico e per dare feedback sommativi agli studenti. In generale, i punteggi medi delle scale considerate («Uso dei dati per la gestione e la qualità della didattica», e «Valutazione informata dai dati digitali») si collocavano sistematicamente sotto i 4 punti (su una scala di 5) e, nella fattispecie, all'interno di un *range* che variava da 0,99 a 3,73.

Tale risultato mette in luce una mancanza di familiarità – piuttosto generalizzata all'interno della popolazione di riferimento – per quanto riguarda le competenze di *data literacy*.

Infine, in uno studio condotto con una università italiana e una spagnola, sempre Raffaghelli (2021) ha messo in luce come le competenze di *data literacy*, soprattutto in relazione ad un approccio critico ai dati, siano ancora poco diffuse in entrambe le realtà: su una scala da 1 a 5, la media dei punteggi con cui i docenti mettevano in atto pratiche basate sui dati era di circa 1,97, e perlopiù connesse ad attività di valutazione, di miglioramento della didattica e dei processi

istituzionali, e come risorse per l'insegnamento; assai raramente è associata una riflessione critica, *con e per* gli studenti, sull'utilizzo dei dati.

I capitoli successivi discuteranno più approfonditamente quest'ultima ricerca e, sulla base dei risultati di questa, si cercherà di definire un modello fattoriale utile a stimare il fabbisogno formativo dei docenti universitari, al fine di supportarli nella direzione di uno sviluppo professionale che tenga conto delle nuove criticità legate ai dati.

3. Metodologia della ricerca

3.1. Epistemologia, ontologia e assiologia della ricerca

L'approccio ontologico che fa da cornice a questa ricerca è quella del realismo critico; si tratta di un paradigma pragmatista sorto come reazione alle logiche oggettivistiche del Positivismo, e associato in particolar modo alla Teoria Critica e alla Pedagogia Critica (Farrow et al., 2020, p. 18; Critical realism – New World Encyclopedia, 2023).

All'interno della "famiglia" degli approcci critici, il ricercatore agisce in base ai suoi ideali, valori ed interessi per mettere in luce fenomeni quali discriminazioni, disuguaglianze e iniquità all'interno dei sistemi socioeconomici, spesso combinando tecniche interpretative (quali studio di caso, interviste, focus group, ecc...) e positiviste (come nel caso di disegni di ricerca sperimentali e quasi-sperimentali, basati sull'isolamento e la misurazioni delle variabili d'interesse, nonché sull'analisi statistica, interviste strutturate e, come nel caso della presente ricerca, questionari) (ibidem).

Se, da un lato, il termine "realismo" richiama alla visione ontologica tipicamente positivista, secondo la quale esiste una realtà oggettiva che può essere studiata e analizzata con i metodi quantitativi e rigorosi delle scienze naturali (o "esatte"), dall'altro l'aggettivo "critico" riconduce alle posizioni più tradizionalmente associate alla ricerca interpretativa, in cui si afferma che non è possibile disgiungere l'atto della ricerca da influenze culturali e individuali (ivi, p. 16), esaltando la soggettività intrinseca nell'esperienza umana e cercando di "comprendere" anziché "spiegare" (ivi, p. 15). Il paradigma interpretativo, dunque, adotta spesso una posizione relativista (ivi, p. 16), nell'intenzione di elaborare un quadro d'insieme che tragga vantaggio dai diversi punti di vista che osservano un dato fenomeno, conferendo ad esso senso e significato.

Operando una sintesi intermedia tra questi due estremi, si può affermare che, da un lato, l'orientamento interpretativo e critico trasforma in risorsa il posizionamento del ricercatore all'interno del contesto socioculturale – la cui figura è, per questo, assimilabile a quella di un attivista che agisce come forza emancipatoria e trasformativa (ivi, p. 18), per orientare individui, gruppi, istituzioni, nonché la società nel suo insieme, nella direzione del progresso sociale, dell'equità, dell'uguaglianza e della democrazia – dall'altro, l'approccio "realista" afferma non solo la possibilità di conoscere la realtà in quanto oggettiva (ivi, p. 14) – assunto marcatamente positivista – ma anche di poterla comprendere, e ricomprendere, all'interno dei sistemi di relazione e di potere che agiscono all'interno della società.

A tale impostazione fondamentale è stato aggiunto, in qualità di *trait d'union*, l'atteggiamento del *pragmatismo scientifico*: all'interno di una cornice di pensiero che non si pone la questione della *conoscibilità* del reale, le domande relative alla veridicità – o meno – della conoscenza scientifica vengono *sospese*; l'adozione di posizioni pragmatiche evita, quindi, discussioni filosofiche e dibattiti accademici, dando l'opportunità al ricercatore di agire più rapidamente, e con una maggiore flessibilità, sotto pressione temporale e/o materiale. Il pragmatismo, infatti, rivendica l'uso dei modelli e dei metodi degli altri approcci di ricerca, in base agli scopi e ai vincoli che si trova ad affrontare (Farrow et al., 2020, p. 19).

L'approccio pragmatico può, inoltre, fornire un valido supporto nell'affrontare il tema della *complessità* del reale, in cui fenomeni diversi e talvolta distanti, sia nel tempo che nello spazio, interagiscono tra loro, determinando esiti spesso imprevedibili ed inattesi («Complexity», 2023): ciò è di particolare rilievo in un'ottica di ricerca scientifica, perché introduce un margine d'errore ineliminabile all'interno dei risultati, ed impone al ricercatore di agire in condizioni di perenne ambiguità ed incertezza («Uncertainty», 2023).

Infatti, se l'errore sistematico all'interno dei dati può essere quantificato e, per quanto possibile, ridimensionato attraverso l'impiego di metodi statistici, resta il problema della *cognizione limitata* dello scienziato, che necessariamente deve elaborare e ricorrere a modelli semplificati dei fenomeni, dei quali è sostanzialmente impossibile ricostruire l'intera rete di relazioni e correlazioni con gli aspetti del reale (Wikipedia, 2023). Il pragmatismo, pertanto, evitando di definire le nozioni di “*realtà*”, di “*verità*”, o di “*conoscenza*”, permette di agire in tali condizioni di variabilità ed incertezza epistemiche.

Pertanto, se da un lato il ricercatore agisce per ottenere dati il più possibile affidabili, dall'altro deve essere consapevole della “*non-neutralità*” e della *parzialità* del dato: esso non può essere considerato, infatti, né “*oggettivo*” – il dato stesso viene raccolto in base ad una precisa scelta del ricercatore, secondo modalità e tecniche da lui decise – né *definitivo*, in quanto le conclusioni che è possibile trarre da esso sono tutt'al più *probabili*, se non *verosimili*, ma, in ultima analisi, mai *vere*.

Da ciò, deriva la possibilità che nuove osservazioni e nuove interpretazioni portino ad un mutamento di paradigma e ad una revisione, anche profonda, delle ipotesi di partenza del ricercatore.

3.2. Metodi

3.2.1. Il questionario

L'orientamento pragmatico pone l'accento sulla risoluzione di un problema pressante – esattamente come può essere identificata la questione delle competenze digitali all'interno della comunità accademica. La raccolta di dati attraverso un questionario, in questo senso, offre l'opportunità al ricercatore di conoscere approfonditamente la realtà di un determinato contesto, raggiungendo un numero elevato di potenziali partecipanti anche in località distanti o non immediatamente accessibili, poiché esso può essere somministrato anche da remoto, e le risposte ad esso possono essere fornite in maniera asincrona, garantendo una elevata flessibilità d'uso sia al ricercatore che al partecipante (ibidem, p. 62). Dal punto di vista delle informazioni che è possibile ottenere, esso può fornire dati sia qualitativi che quantitativi, permettendo di eseguire comparazioni tra soggetti e tra gruppi di soggetti all'interno del campione; quest'ultimo, se statisticamente significativo, può portare a trarre conclusioni sulla popolazione di riferimento.

Tuttavia, l'uso del questionario presenta anche dei risvolti negativi – i soggetti possono fornire risposte parziali, incomplete, non veritiere – o, semplicemente, non rispondere. Inoltre, bisogna tenere in assoluta considerazione l'uso del linguaggio, che deve essere comprensibile e di chiara interpretazione da parte dei rispondenti, per evitare errori sistematici che invaliderebbero la raccolta dati.

Entrando nello specifico dello strumento analizzato in questa sede, si tratta di un questionario basato su una revisione della letteratura compiuta dall'autrice (Raffaghelli, 2019); le scale considerate facevano riferimento al “DigCompEdu”, il framework delle competenze digitali per gli educatori e, nelle intenzioni dell'autrice, doveva rendere lo strumento un utile accompagnamento all'analisi delle competenze digitali di questi.

Pertanto, prima di somministrare il questionario, è stato eseguito da parte dell'autrice (Raffaghelli, 2020), uno studio di validazione *Delphi*: si tratta di un processo basato su un ristretto *pool* di esperti che, in due o più *round*, esprimono giudizi sullo strumento, fino a raggiungere un consenso; in alternativa, il processo di revisione termina in base ad un altro criterio definito a priori, per esempio il numero di round («Delphi Method», 2023).

L'idea alla base di questa modalità di validazione è che i giudizi espressi da soggetti esperti all'interno di gruppi strutturati siano più affidabili rispetto a quelli emessi da gruppi non-strutturati: in particolare, al termine di ogni round, le valutazioni espresse dovrebbero ridursi di numero, fino al raggiungimento di un'interpretazione comune, o “corretta” (ibidem).

3.3. Caratteristiche del questionario

Il questionario, in lingua inglese, spagnola ed italiana, è stato inviato al personale docente e di ricerca delle due università prese in esame; esso era composto da 6 scale per un totale di 35 item. Le possibilità di risposta, elencate su scala Likert a 5 punti, erano volte ad indagare la frequenza dei comportamenti legati all'uso dei dati nella, e per, la didattica, secondo le seguenti opzioni: 1 (Mai), 2 (Qualche volta), 3 (Circa la metà delle volte), 4 (Il più delle volte), 5 (Sempre).

Le scale individuate erano (Raffaghelli, 2020):

- EDMQ (Educational Data use for Management and Quality) – Uso dei dati da diverse fonti per informare i processi istituzionali, la collaborazione e la costruzione di conoscenza accademica a supporto della didattica, inclusa la ricerca.
- DER (Data as Educational Resources) – Uso dei dati come risorse di apprendimento. Adozione di dati provenienti dalla propria ricerca o dati aperti pubblici per orientare e supportare attività di apprendimento.
- DTL (Data supporting Teaching and Learning) – Didattica informata dai dati digitali. Decisione docente, monitoraggio di processi di apprendimento e feed-back con uso di dati digitali dai sistemi adottati per la didattica.
- DA (Data supporting Assessment) – Valutazione informata dai dati digitali. Decisione docente e analisi dei percorsi formativi compiuti a partire da dati ottenuti/estratti tramite piattaforme digitali.
- SED (Students' Empowerment through Data) – Potenziamento degli Studenti. Coinvolgimento degli studenti nella comprensione della propria partecipazione e dei propri processi di apprendimento a partire dagli ecosistemi di dati generati dagli ambienti di didattica a distanza.
- SDL (Students' Data Literacy) – Alfabetizzazione (degli studenti) ai dati. Azioni didattiche mirate allo sviluppo di conoscenza e skills tecniche, grafiche, multimediali, critiche ed estetiche sull'uso dei dati nei contesti di vita e professionali.

La tabella sottostante fornisce un riepilogo degli item del questionario e delle scale ad essi collegate (ibidem):

Tabella 1 – Costrutti adottati nello studio e Domande di ricerca, basato su un quadro di competenze e pratiche basate sui dati legati alle stesse.

Aree di pratiche basate sull'uso di dati digitali nella didattica universitaria	N di Items	Domande di Ricerca
<p>EDMQ <i>(Educational Data use for Management and Quality)</i> - Uso di dati per la gestione e la qualità della didattica</p> <p>Uso dei dati da diverse fonti per informare i processi istituzionali, la collaborazione e la costruzione di conoscenza accademica a supporto della didattica, inclusa la ricerca.</p>	<p>8 item</p> <p>EDMQ1 Uso di dati da Report Nazionali per lo sviluppo istituzionale e la pianificazione.</p> <p>EDMQ2 Uso di dati da Report Istituzionali per lo sviluppo istituzionale e la pianificazione.</p> <p>EDMQ3 Uso di dati dalla valutazione del proprio corso per lo sviluppo istituzionale e la pianificazione.</p> <p>EDMQ4 Uso di dati del proprio corso per la progettazione curricolare.</p> <p>EDMQ5 Uso di analitiche di apprendimento per la progettazione didattica.</p> <p>EDMQ6 Uso di analitiche di apprendimento per analizzare efficacia didattica propria.</p> <p>EDMQ7 Uso di dati da social media inseriti nel mio corso per analizzare la propria efficacia didattica.</p> <p>EDMQ8 Uso di dati da social media dove partecipano gli studenti in generale per la progettazione e l'efficacia didattica.</p>	<p>Quali sono le pratiche basate sull'uso di dati digitali nella didattica più diffuse all'interno dell'Ateneo di Padova?</p> <p>Esistono differenze attraverso le varie tipologie di pratica, considerando particolarmente l'ambito disciplinare?</p>
<p>DER (Data as Educational Resources) - Uso dei dati come risorse di apprendimento.</p> <p>Adozione di dati provenienti dalla propria ricerca o dati aperti pubblici per orientare e supportare attività di apprendimento.</p>	<p>4 item</p> <p>DER1 Uso di dati di report internazionali/nazionali come risorsa didattica.</p> <p>DER2 Riutilizzo di Dati Aperti (Open Data) provenienti dalla ricerca o dalla pubblica amministrazione come risorsa didattica.</p> <p>DER3 Riutilizzo di Dati Aperti (Open Data) provenienti dalla propria ricerca come risorsa didattica.</p> <p>DER4 Riflessione su aspetti etici relative all'integrazione dei dati aperti come risorse didattiche.</p>	
<p>DTL (Data supporting Teaching and Learning) -</p>	<p>5 item</p> <p>DTL1 Integrazione di modalità tradizionali di raccolta dati come parte delle attività e compiti degli studenti.</p>	

<p>Didattica informata dai dati digitali. Decisione docente, monitoraggio di processi di apprendimento e feed-back con uso di dati digitali dai sistemi adottati per la didattica.</p>	<p>DTL2 Integrazione di tecniche estrattive di dati (massivi) come parte delle attività e compiti degli studenti. DTL3 Insegnamento di tecniche per l'elaborazione/ rappresentazione/ visualizzazione di dati. DTL4 Riflessione docente sugli aspetti di estrazione automatica dei dati per utilizzo didattico. DTL5 Integrazione di dashboards (pannelli di controllo visuali) per supportare le abilità autoregolative degli studenti.</p>	
<p>DA (Data supporting Assessment) - Valutazione informata dai dati digitali. Decisione docente e analisi dei percorsi formativi compiuti a partire da dati ottenuti/estratti tramite piattaforme digitali.</p>	<p>10 item DA1 Uso di dati ottenuti da verifiche progressive per monitoraggio dell'apprendimento. DA2 Uso di dati ottenuti da verifiche progressive per monitoraggio didattico DA3 Uso dei dati ottenuti da verifiche per dare feedback sommativi agli studenti. DA4 Uso dei dati ottenuti dalla valutazione di un corso per dare feedback formativi agli studenti. DA5 Riflessione insieme agli studenti sui processi di raccolta dei dati provenienti dalla didattica. DA6 Uso dei dati dalle piattaforme di apprendimento per monitorare/valutare la didattica. DA7 Uso di “dashboards” presenti nelle piattaforme di apprendimento per informare la mia didattica. DA8 Uso di sistemi digitali automatizzati per analizzare e valutare i risultati degli studenti (e.g. online quiz). DA9 Uso di dati dalle piattaforme di apprendimento per riflettere insieme agli studenti sulla qualità della didattica. DA10 Uso di sistemi digitali automatizzati per valutare in modo complessivo la qualità della didattica (e.g., online surveys).</p>	
<p>SED (Students' Empowerment through Data) - Potenziamento degli Studenti. Coinvolgimento degli studenti nella comprensione della propria</p>	<p>3 item SED1 Condivisione di report istituzionali per supportare la riflessione degli studenti sul progresso dei propri apprendimenti nel contesto istituzionale o sociale. SED2 Condivisione di dati provenienti dalla piattaforma di apprendimento (analitiche di apprendimento) per supportare la riflessione degli studenti sul</p>	

partecipazione e dei propri processi di apprendimento a partire dagli ecosistemi di dati generati dagli ambienti di didattica a distanza.	progresso dei propri apprendimenti nel contesto istituzionale o sociale. SED3 Discussione su analitiche estratte da social media (integrati nella didattica) con gli studenti, ove consentire loro di riflettere sul progresso dei propri apprendimenti nel contest istituzionale o sociale.	
SDL (Students' Data Literacy) - Alfabetizzazione (degli studenti) ai dati. Azioni didattiche mirate allo sviluppo di conoscenza e skills tecniche, grafiche, multimediali, critiche ed estetiche sull'uso dei dati nei contesti di vita e professionali.	5 item SDL1 Promozione dell'abilità degli studenti per valutare criticamente la affidabilità e credibilità dei dati incorporati in una fonte informativa. SDL2 Promozione di skills tra gli studenti per integrare dati in modo significativo entro report narrativi o presentazioni. SDL3 Promozione della discussione e riflessione tra gli studenti sui vincoli sociali relativi all'uso dei dati e le applicazioni digitali nell'IA come processo sociotecnico. SDL4 Promozione dell'analisi etica relativamente a raccolta e utilizzo dei propri dati come parte del processo di apprendimento. SDL5 Discussione congiunta con gli studenti sull'impatto dell'uso dei dati nella società in generale.	

Tabella 1: Costrutti adottati all'interno della ricerca

3.4. Strumenti informatici

3.4.1. L'ambiente di R

L'analisi dei dati è stata condotta tramite il software statistico R, con l'aggiunta dell'interfaccia grafica e delle funzioni di supporto di R Studio; R è sia un programma statistico, open-source e multiplatforma, che un linguaggio di programmazione; esso viene fornito dalla R Foundation, un'organizzazione *no-profit* che si occupa dello sviluppo e distribuzione dell'applicazione – in collaborazione con i *CRAN mirror*, repository da cui ottenere il programma e i “pacchetti” (*packages*) che forniscono funzioni aggiuntive – nonché di svolgere attività di formazione e organizzazione di conferenze sulla statistica computazionale.

Come riportato sul sito ufficiale (*R: What is R?*, 2023), R è un progetto GNU, ovvero rientra all'interno dell'orizzonte dei software rilasciati per funzionare all'interno dell'ambiente GNU; la distribuzione del programma è concepita sulle basi delle quattro libertà essenziali del “Software

Libero”, come stabilite dalla *Free Software Foundation*: eseguire il programma per i propri scopi (personali o professionali), modificarlo, e/o redistribuirlo liberamente, sia nella forma modificata che in quella originale (ibidem).

Da ciò, deriva che non solo il codice sorgente di R è disponibile e liberamente consultabile, ma la comunità di utilizzatori è messa nelle condizioni di elaborare nuove versioni del programma, o di proporre estensioni (i “pacchetti”) delle sue funzioni originarie, aumentandone le potenzialità (ivi). Tale lavoro viene svolto su base volontaria e gratuita, allo scopo di alimentare l’uso di R e di fornire un contributo alla comunità degli utenti.

R Studio, invece, è un ambiente di sviluppo integrato (*IDE*, “*Integrated Development Environment*”) sviluppato da un’azienda privata, la “Posit Software, PBC” (fino al 2022 nota come “RStudio, PBC”) (*Posit*, 2023).

Si tratta di una “società a beneficio pubblico”, o “di utilità comune” (“*Public Benefit Company*”) che, a differenza di altre entità operanti nell’area del software libero, è a scopo di lucro; il suo fine, infatti, è quello di elaborare prodotti open source e gratuiti a cui fornire, poi, soluzioni commerciali e commercializzabili – come servizi di supporto ad imprese e organizzazioni, o soluzioni avanzate per l’analisi dei dati – per finanziare le proprie attività, secondo una logica ricorsiva: proporre soluzioni *free* e aperte ne facilita l’adozione e la diffusione e, di conseguenza, aumenta le possibilità di offrire prodotti e servizi a pagamento; i proventi vengono, poi, destinati allo sviluppo di nuovi prodotti e al miglioramento di quelli esistenti, alimentando così un circolo virtuoso (ibidem).

R Studio rappresenta un utile complemento alla scarna interfaccia di R, fornendo una console, un editor per supportare l’utente nella stesura del codice, e strumenti per la gestione dell’ambiente di lavoro (“*workspace*”) e la creazione di rappresentazioni grafiche; esso viene accompagnato da una collezione di pacchetti pensati appositamente per l’analisi dei dati, il “*tidyverse*”, che raccoglie al suo interno vari tool, tra cui “*ggplot2*” – usato per l’elaborazione di grafici “*publication-ready*” – e “*dplyr*”, una “grammatica” per la manipolazione dei dati che ne semplifica l’analisi. Esattamente come R, R Studio e il “*tidyverse*” sono open source, gratuiti e multiplatforma (ivi).

Come già accennato, i pacchetti di funzioni per R sono numerosi, in costante crescita ed aggiornamento. Oltre al “*tidyverse*”, infatti, ai fini della stesura di questa tesi sono stati utilizzati anche (assieme alle relative contingenze):

- “*Summarytools*” per eseguire l’analisi univariata e bivariata del dataset,
- “*Flextable*”, per creare ed esportare le tabelle in formato “flextable”,
- “*Rempsyc*” per tabelle in stile “APA”,
- “*Ggpubr*” per una più semplice realizzazione di grafici “pronti all’uso”,

- “*Heplots*” per la visualizzazione dei risultati dei test multivariati,
- “*Performance*”, per testare la validità del questionario (“*Cronbach's α* ”), e la multicollinearità del dataset.
- “*Corrplot*”, per la rappresentazione grafica della matrice di correlazione,
- “*Rstatis*” per l'esecuzione dei test ANOVA, MANOVA, della dimensione dell'effetto;
- “*Psych*”, per l'analisi multivariata e fattoriale,
- “*Lavaan*”, per l'analisi delle variabili latenti,
- “*GPArotation*”, per la rotazione dei fattori,
- “*semPlot*”, per la realizzazione del “*Path Diagram*”.

3.4.2. Una piattaforma “*home-made*”

Congiuntamente a questo lavoro di tesi, al fine di strutturare un ambiente di collaborazione sicuro, poiché rispettoso della privacy degli utenti, su misura, che fosse flessibile nelle sue declinazioni e versatile nell'utilizzo, l'autore ha deciso di creare una sua personale piattaforma *Moodle* sul proprio computer domestico; essa è stata strutturata secondo le necessità del presente lavoro, e aggiungendo funzionalità ritenute utili al progetto di tesi.

Tuttavia, prima di descrivere l'esperienza compiuta nell'installare e gestire tale piattaforma personale e “su misura”, è necessario introdurre brevemente l'ambiente Moodle: si tratta di una piattaforma d'apprendimento open source, gratuita, multipiattaforma e multilingua, pensata per soddisfare le necessità di docenti, studenti e amministratori, permettendo ad istituzioni pubbliche e private – non necessariamente operanti nel campo dell'istruzione – di utilizzare ed ampliare il software di base secondo le proprie esigenze, attraverso l'impiego di plug-in liberamente installabili dal sito dello sviluppatore (*LMS Platform - Moodle LMS - Learning Management System*, 2023). Essi sono spesso forniti da contributori che, a titolo personale e gratuito, elaborano programmi e strumenti che ampliano le funzioni del software Moodle originario.

Quest'ultimo è una B-Corp (Società a Beneficio Pubblico), la *Moodle Pty Limited*, che coordina l'operato di altre 80 società partner impegnate nello sviluppo e gestione della piattaforma (*Overview | Moodle Developer Resources*, 2023).

Essendo Moodle dotato di un'architettura scalabile – esso può, con risorse hardware adeguate, servire sia l'individuo che la grande organizzazione – è stato installato sul personal computer dell'autore, su sistema operativo Windows 10; questo dettaglio è di particolare rilievo, poiché il processo di start-up della piattaforma viene notevolmente semplificato da una procedura, guidata ed automatizzata, che è stata creata per questo sistema, ma non disponibile, per esempio, per Linux.

Terminata questa prima fase, è stato possibile creare uno spazio personale – un corso Moodle intitolato “Tesi magistrale” – il quale, dopo avervi inserito il materiale di tesi prodotto, è stato integrato con numerose funzionalità. Tra queste si possono annoverare:

- Un nuovo formato per gestire i contenuti del corso in maniera più flessibile e personalizzata;
- Strumenti per la pianificazione delle attività e il monitoraggio dei progressi tramite *checklist* e barre di completamento;
- Un editor per la visualizzazione, la modifica e l’annotazione collaborativa di documenti PDF.

Questo step iniziale di configurazione della piattaforma è stato eseguito con una certa facilità; tuttavia, rappresenta solo una piccola parte del lavoro necessario per raggiungere la piena operatività: per ottenerla, infatti, è necessario attivare manualmente alcune funzionalità, come lo script “*cron*”, o la configurazione del Moodle per l’invio di e-mail di servizio, quest’ultimo passo indispensabile per permettere la registrazione alla piattaforma. Il *cron*, in particolare, svolge al posto dell’utente-amministratore una serie di operazioni indispensabili, come l’invio di e-mail riguardanti modifiche ai corsi, controlli di sicurezza di routine, backup dei contenuti della piattaforma. Tale programma deve essere configurato per operare in autonomia ad intervalli di un minuto; all’interno di Windows, tale operazione deve essere svolta all’interno dello strumento “Utilità di Pianificazione”, creando un processo apposito e definendone le modalità di funzionamento (momento dell’entrata in azione, ciclicità, ecc...). Si tratta di un procedimento non banale, che richiede una conoscenza tecnica del sistema operativo già approfondita rispetto a quella dell’utente medio.

Oltre all’attivazione di questo script è, inoltre, essenziale rendere la piattaforma accessibile *dall’esterno*, ovvero da dispositivi remoti che non appartengono alla Rete Locale (*LAN*, “*Local Area Network*”) dell’utente originario. Questo è un passaggio ben più complesso del precedente: innanzitutto, richiede di attivare presso il proprio *provider di rete* (il fornitore del servizio di collegamento ad Internet) un indirizzo IP statico e pubblico: un indirizzo IP (“*Internet Protocol*”) è un codice numerico di nove cifre che identifica, in maniera univoca, un dispositivo (sia esso un personal computer, una stampante, uno smartphone, ecc...) collegato ad una rete. Normalmente, esso è *dinamico* (cambia ad ogni nuovo collegamento alla rete) e *privato* (il dispositivo è visibile solo all’interno della rete domestica – o comunque privata – dell’utente); ogni elemento del *network* è quindi invisibile, ed irraggiungibile, dall’esterno.

Pertanto, per ovviare a questa situazione, è necessario contattare il fornitore dei servizi di connettività ed attivare un contratto per il noleggio di un indirizzo IP con le caratteristiche adeguate, ovviamente pagando un determinato corrispettivo. Benché l'esborso equivalga solamente ad una ventina di euro all'anno, è un altro elemento da tenere in considerazione nel decidere di costituire la propria piattaforma.

Oltre a ciò, un'altra considerazione è doverosa: se, in condizioni abituali, non è dato rintracciare un qualsiasi mezzo informatico di una rete domestica, è per via della sicurezza degli utenti, che altrimenti si troverebbero esposti ad un elevato numero di potenziali minacce (malware di vario genere, furto o perdita di dati perpetrati da criminali informatici, e numerosi altri). Da questo fatto ne deriva che la decisione di rendere, in qualche modo, la propria rete raggiungibile da terze parti, deve essere soppesata con attenzione.

In questo caso, anziché un'apertura indiscriminata dell'accesso al sito Moodle, si è preferito impostare una VPN (*Virtual Private Network*, una rete virtuale privata che stabilisce una connessione cifrata e sicura tra uno specifico utente e il computer di destinazione), con un nome utente ed una password specifici per ogni utilizzatore esterno (in questo caso, la sola relatrice della presente tesi, la Prof.ssa Raffaghelli). Configurare una VPN può essere complesso; è stato necessario individuare uno specifico software compatibile con il *router* (il dispositivo che permette l'accesso ad Internet), nonché perfezionare le opzioni di funzionamento di quest'ultimo per abilitare accessi esterni, ed infine creare l'utente personalizzato.

Riguardo quest'ultimo aspetto, non si possono non considerare le possibili difficoltà a cui gli utenti esterni possono andare incontro nell'utilizzo di una VPN: tale servizio non rientra nei casi d'uso tradizionali di un personal computer da parte di un utilizzatore medio, non-specialista; apprenderne l'impiego, oltre che comprenderne la necessità, richiede la pazienza, la disponibilità, la motivazione e anche l'intraprendenza di tutte le parti in causa. Benché il programma in questione permettesse di facilitare le procedure di identificazione della VPN – attraverso la creazione di un file apposito che può essere condiviso ed installato da ogni utente della rete virtuale con pochi, semplici, passaggi – rimane il problema delle competenze tecniche da mettere in atto durante questa fase, oltre all'effettivo desiderio di ogni parte di superare eventuali problematiche legate alla piattaforma.

Benché, nell'insieme, il progetto abbia avuto successo – la piattaforma Moodle è stata effettivamente utilizzata durante il lavoro di tesi, ed è stata resa accessibile ad utenti terzi – vanno annoverate criticità, di varia natura, incontrate durante il percorso; tra queste, vengono sottolineate:

- Difficoltà nell'avvio della piattaforma, a causa di processi di sistema che ne occupavano lo slot di esecuzione;

- Deterioramento critico del database di Moodle, che ha provocato in una occasione la perdita totale dei dati, e la reinstallazione e riconfigurazione della piattaforma;
- Occasionali perdite dei file caricati nella piattaforma, per cause che non è stato possibile identificare.

Ognuno di questi incidenti ha richiesto tempo ed energie per la ricerca di soluzioni; la documentazione relativa a Moodle, ed ai suoi potenziali problemi, è piuttosto scarna ed incompleta, talvolta non aggiornata all'ultima versione disponibile del software. Questo provoca un innalzamento del livello di competenza e capacità richieste per l'uso e la gestione della piattaforma, poiché non esiste una documentazione, essenziale e di semplice consultazione e comprensione, per la risoluzione dei problemi più frequenti.

Tale situazione è piuttosto comune nel campo dei progetti open source che, spesso, essendo guidati dalla buona volontà della comunità di utenti, non dispongono dei servizi di assistenza tecnica che sono invece resi disponibili da parte delle aziende *for-profit*.

Per concludere, l'esperienza ottenuta attraverso questo progetto può essere considerata di alto valore personale e formativo; essa ha innalzato non solo le conoscenze e competenze tecniche dell'autore, ma lo ha anche reso più consapevole rispetto alla complessità di costituire e gestire una piattaforma.

Tale esperienza, se opportunamente accompagnata dal sostegno di una manualistica adeguata e di un supporto tecnico di base, può rivelarsi una utile fonte di arricchimento anche all'interno di percorsi didattici tradizionali nella cornice dell'istruzione superiore.

4. Risultati

4.1. Statistiche descrittive univariate

4.1.1 Caratteristiche generali del campione

Il campione preso in esame, sebbene non statisticamente significativo, è composto da docenti e personale di ricerca di due università: si tratta, nello specifico, di una università italiana pubblica, con circa 60.000 studenti e di antica tradizione, e di una giovane università spagnola privata, con circa 80.000 studenti e che opera prevalentemente on-line (Raffaghelli et al., 2021, p. 8).

Il tasso di risposta al questionario, somministrato on-line, è stato rispettivamente del 9% e del 17% (ibidem), con 374 e 800 risposte. Eliminando i questionari che presentavano risposte incomplete, sono rimasti 547 record, di cui 210 appartenenti al caso italiano e 337 a quello spagnolo (ivi).

In termini assoluti, il campione è stato analizzato per:

- Genere: Uomini (280), Donne (261), Non dichiarato (6);
- Et : Meno di 25 anni (1), 25-34 (40), 35-44 (170), 45-54 (215), Pi  di 55 (131);
- Nazione: Italia (210), Spagna (337);
- Area disciplinare: Scienze umanistiche e Linguistica (56), Scienze della vita (110), Scienze naturali e formali (57), Scienze sociali (227), Tecnologia (97);
- Esperienza d'insegnamento: Meno di 3 anni (73), 3-4 (56), 5-10 (79), 10-15 (96), Pi  di 15 (243);
- Esperienza di ricerca: Nessuna esperienza (80), Meno di 3 anni (46), 3-5 (44), 6-10 (80), 11-15 (70), Pi  di 15 (227).

4.1.2 Statistiche descrittive univariate per le variabili quantitative

Prima di calcolare le statistiche descrittive per ogni singola scala del questionario, sono stati identificati ed eliminati gli *outlier* presenti nella distribuzione dei punteggi, riducendo cos  i record a 494. Di seguito, la Tabella 2 riassume le principali statistiche analizzate: Media, Deviazione Standard (DS), Minimo (Min), Q1 (Primo Quartile), Mediana (Med.), Terzo Quartile (Q3), Massimo (Max), Scarto Medio Assoluto (SMA), Scarto Interquartile (SI), il Campo di Variazione (CV), lo Standard Error (SE) e, infine, gli indici di simmetria: Skewness e Curtosi.

<i>Statistiche descrittive univariate degli item del questionario</i>													
	Media	DS	Med.	Q1	Q3	Min	Max	S.M.A.	S.I.	C.V.	Skewness	Curtosi	S.E.
EDMQ1	1,65	1,55	1,00	0,00	3,00	0,00	5,00	1,48	3,00	0,94	0,81	-0,46	0,07
EDMQ2	1,89	1,60	2,00	1,00	3,00	0,00	5,00	1,48	2,00	0,85	0,54	-0,86	0,07
EDMQ3	2,73	1,72	3,00	1,00	4,00	0,00	5,00	2,97	3,00	0,63	-0,14	-1,27	0,08
EDMQ4	2,56	1,77	3,00	1,00	4,00	0,00	5,00	2,97	3,00	0,69	-0,01	-1,38	0,08
EDMQ5	2,75	1,77	3,00	1,00	4,00	0,00	5,00	2,97	3,00	0,64	-0,16	-1,36	0,08
EDMQ6	2,28	1,71	2,00	1,00	4,00	0,00	5,00	1,48	3,00	0,75	0,18	-1,28	0,08
EDMQ7	1,55	1,51	1,00	0,00	3,00	0,00	5,00	1,48	2,75	0,98	0,85	-0,33	0,07
EDMQ8	1,09	1,21	1,00	0,00	2,00	0,00	5,00	1,48	2,00	1,12	1,29	1,21	0,05
DER1	2,16	1,67	2,00	1,00	4,00	0,00	5,00	1,48	3,00	0,77	0,35	-1,14	0,08
DER2	2,34	1,66	2,00	1,00	4,00	0,00	5,00	1,48	3,00	0,71	0,13	-1,19	0,07
DER3	2,21	1,67	2,00	1,00	4,00	0,00	5,00	1,48	3,00	0,75	0,25	-1,22	0,07
DER4	2,03	1,64	2,00	1,00	3,00	0,00	5,00	1,48	2,00	0,81	0,39	-1,10	0,07
SED1	1,69	1,56	1,00	0,00	3,00	0,00	5,00	1,48	3,00	0,92	0,72	-0,63	0,07
SED2	1,74	1,57	1,00	0,00	3,00	0,00	5,00	1,48	3,00	0,90	0,69	-0,60	0,07
SED3	1,16	1,29	1,00	0,00	2,00	0,00	5,00	1,48	2,00	1,11	1,31	1,13	0,06
DTL1	2,21	1,70	2,00	1,00	4,00	0,00	5,00	1,48	3,00	0,77	0,26	-1,23	0,08
DTL2	1,31	1,39	1,00	0,00	2,00	0,00	5,00	1,48	2,00	1,07	1,13	0,47	0,06
DTL3	2,00	1,72	2,00	0,00	3,00	0,00	5,00	1,48	3,00	0,86	0,43	-1,10	0,08
DTL4	1,75	1,58	1,00	0,00	3,00	0,00	5,00	1,48	3,00	0,90	0,69	-0,65	0,07
DTL5	1,58	1,60	1,00	0,00	2,00	0,00	5,00	1,48	2,00	1,02	0,88	-0,40	0,07
DA1	2,65	1,67	3,00	1,00	4,00	0,00	5,00	1,48	3,00	0,63	-0,08	-1,31	0,08
DA2	2,72	1,66	3,00	1,00	4,00	0,00	5,00	1,48	3,00	0,61	-0,16	-1,23	0,07
DA3	2,92	1,69	3,00	2,00	4,00	0,00	5,00	1,48	2,00	0,58	-0,27	-1,22	0,08
DA4	2,80	1,67	3,00	1,00	4,00	0,00	5,00	1,48	3,00	0,59	-0,18	-1,23	0,07
DA5	2,39	1,64	2,00	1,00	4,00	0,00	5,00	1,48	3,00	0,68	0,15	-1,18	0,07
DA6	2,36	1,71	2,00	1,00	4,00	0,00	5,00	1,48	3,00	0,73	0,17	-1,31	0,08
DA7	1,76	1,63	1,00	0,00	3,00	0,00	5,00	1,48	3,00	0,93	0,69	-0,75	0,07
DA8	1,78	1,64	1,00	0,00	3,00	0,00	5,00	1,48	3,00	0,93	0,70	-0,74	0,07
DA9	1,46	1,42	1,00	0,00	2,00	0,00	5,00	1,48	2,00	0,98	1,02	0,14	0,06
DA10	1,56	1,61	1,00	0,00	3,00	0,00	5,00	1,48	2,75	1,03	0,90	-0,45	0,07
SDL1	2,46	1,75	2,00	1,00	4,00	0,00	5,00	2,97	3,00	0,71	0,07	-1,39	0,08
SDL2	2,33	1,72	2,00	1,00	4,00	0,00	5,00	2,97	3,00	0,74	0,13	-1,34	0,08
SDL3	1,42	1,45	1,00	0,00	2,00	0,00	5,00	1,48	2,00	1,02	1,02	0,03	0,07
SDL4	1,56	1,48	1,00	0,00	2,00	0,00	5,00	1,48	2,00	0,95	0,84	-0,28	0,07
SDL5	1,75	1,56	1,00	0,00	3,00	0,00	5,00	1,48	3,00	0,89	0,62	-0,76	0,07

Tabella 2: Principali statistiche descrittive degli item del questionario

Ciò che si evince da questa prima estrapolazione è come le pratiche legate all'uso dei dati, almeno all'interno del campione analizzato, siano ancora poco diffuse: il punteggio medio più basso – pari a 1,09 – lo si riscontra nella variabile EDMQ8, relativa all'item “*Ho estratto e utilizzato dati, provenienti da social media a cui i miei studenti partecipano liberamente, al fine di valutare l'efficacia dell'insegnamento* ”; mentre il punteggio più elevato – pari a 2,92 – è relativo alla variabile DA3 (“*Ho utilizzato dati provenienti da attività di verifica per dare feedback*”). Ciò significa che i comportamenti legati alla *Data Literacy* vengono messi in pratica con una frequenza che oscilla tra il “quasi mai” e il “meno della metà delle volte”.

A scopo di sintesi, nella tabella sottostante (Tabella 3) vengono invece illustrate le principali statistiche descrittive per variabili raggruppate:

Variabili	<i>n</i>	Media	S.D.	Med.	SMA	Min	Max	S.I.	Skewness	Curtosi	S.E.
EDMQ	4.804	2,10	1,72	2,00	1,48	0,00	5,00	5,00	0,38	-1,17	0,02
DER	2.404	2,21	1,67	2,00	1,48	0,00	5,00	5,00	0,27	-1,20	0,03
SED	1.803	1,55	1,50	1,00	1,48	0,00	5,00	5,00	0,86	-0,30	0,04
DTL	3.005	1,77	1,62	1,00	1,48	0,00	5,00	5,00	0,67	-0,74	0,03
DA	6.010	2,28	1,72	2,00	1,48	0,00	5,00	5,00	0,24	-1,28	0,02
SDL	3.005	1,93	1,66	1,00	1,48	0,00	5,00	5,00	0,51	-1,02	0,03

Tabella 3: Statistiche descrittive per scale del questionario

Come si può notare, le medie complessive per gruppi di variabili si collocano all'interno di un *range* compreso tra 1,77 (DTL), e 2,28 (DA), evidenziando come, in termini generali, le attività basate sui dati siano piuttosto rare all'interno del campione, concentrandosi su attività didattiche, di valutazione, o di supporto allo sviluppo istituzionale.

I successivi paragrafi introducono le analisi statistiche inferenziali che sono state eseguite, e i relativi risultati. Innanzitutto, viene illustrato l'uso del Cronbach's α ai fini della verifica dell'affidabilità delle scale del questionario; ad esso segue l'analisi della varianza ANOVA e, ad essa, la MANOVA. Il capitolo, quindi, si chiude con le analisi fattoriali esplorativa e confermativa.

4.2. Analisi statistica inferenziale

4.2.1. Misure di affidabilità: il *Cronbach's α*

Per misurare l'affidabilità delle misure del questionario, è stato eseguito il test del *Cronbach's α* , utilizzando la funzione “*alpha*” del pacchetto “*psych*”. Il valore risultante è stato $\alpha = 0.95$; dal momento che la regola generale per ritenere affidabile una misura, almeno ad un livello sufficiente, è di avere un valore di α maggiore o uguale a 0.7, si può concludere che il questionario ha un'elevata consistenza interna ed altrettanto elevata affidabilità.

4.2.2. Premesse per l'analisi della varianza ANOVA

L'ANOVA (analisi della varianza) è una generalizzazione del test-t che, a differenza di quest'ultimo, permette di eseguire confronti tra più di due gruppi d'interesse. Il suo scopo è quello di individuare i possibili effetti di variabili indipendenti categoriali su variabili risposta di tipo quantitativo. Considerata la natura dei dati a disposizione, ovvero le caratteristiche dei soggetti (età, nazione di residenza, ambito disciplinare, esperienza d'insegnamento e di ricerca) come fattori indipendenti, e le risposte agli item su scala Likert a cinque punti come variabili dipendenti, si è deciso di utilizzare un test ANOVA; quest'ultimo è stato eseguito *entro i soggetti*.

Le premesse da rispettare per poter eseguire correttamente un'ANOVA sono:

- 1) Assenza di outlier,
- 2) Dati distribuiti secondo una gaussiana univariata (curva normale),
- 3) Omoschedasticità: varianze uguali all'interno dei gruppi.

Sono stati quindi eseguiti i test specifici per verificare il rispetto di tali assunzioni: è stato usato il metodo dei Box-plot per identificare i valori anomali, il test di Shapiro-Wilk di normalità, ed infine il test di Levene per l'omogeneità delle varianze.

In primo luogo, sono stati individuati gli outlier univariati, attraverso la funzione di R “*is_outlier*”; in questo caso, ne sono stati rilevati solamente 3, poi esclusi dal dataset. Successivamente, è stato eseguito il test di normalità univariata, il quale ha dato un esito statisticamente significativo (*p-value* inferiore a 0,001 in ogni variabile esaminata): si è pertanto stabilito che i dati non seguono una distribuzione gaussiana.

Tenendo conto di quest'ultimo aspetto, si è quindi proceduto ad eseguire un test non parametrico per l'analisi dell'omogeneità delle varianze all'interno dei gruppi, il test di Levene: tale verifica ha

mostrato come, in cinque gruppi su sei, le varianze fossero diverse; sono, perciò, eteroschedastici. Al contrario, l'unico gruppo omoschedastico è quello relativo all'età dei partecipanti.

Tenendo presente quanto emerso dalla verifica degli assunti, ovvero che ben due premesse su tre siano state violate, si è scelto di procedere con l'ANOVA nel caso del gruppo "Età", e con una Welch-ANOVA negli altri.

4.2.3. ANOVA per gruppo "Genere"

È stato eseguito un Welch-ANOVA test a una via per verificare l'esistenza di una relazione tra la variabile "Genere" – caratterizzata dai due livelli "Maschio" e "Femmina" – e le risposte date al questionario: nello specifico, si è voluta testare l'ipotesi che le varianze nelle risposte, tra i due gruppi, fossero differenti. Si è scelto il test Welch-ANOVA poiché tale gruppo è eteroschedastico. È stato quindi utilizzato il seguente script di R:

```
aov_dati_gender <- dati_ragg %>% welch_anova_test(value ~ Gender)
print(aov_dati_gender, n = 35)
tab_aov_gender <- data.frame(aov_dati_gender)
View(tab_aov_gender)
```

L'output risultante ha evidenziato varianze diverse per alcune variabili appartenenti al gruppo "Data for Assessment", "Educational Management and Quality", e "(promoting) Students' Data Literacy"; nello specifico, gli item in cui sono state riscontrate varianze differenti e statisticamente significative erano:

- "Uso di dati ottenuti da verifiche progressive per il monitoraggio dell'apprendimento (DA1)";
- "Uso di dati ottenuti da verifiche progressive per il monitoraggio didattico (DA2)";
- "Uso dei dati ottenuti da verifiche per dare feedback sommativi agli studenti (DA3)";
- "Uso dei dati ottenuti dalla valutazione di un corso per dare feedback formativi agli studenti (DA4)";
- "Uso di analitiche di apprendimento per la progettazione didattica (EDMQ5)";
- "Promozione dell'abilità degli studenti per valutare criticamente l'affidabilità e credibilità dei dati incorporati in una fonte informativa (SDL1)".

I risultati sono riassunti nelle tabelle che seguono (Tabella 4 e 5):

Variabile dipendente	Media Uomini	D.S. Uomini	Media Femmine	D.S. Femmine
DA1	2,8	1,7	2,5	1,6
DA2	2,9	1,7	2,5	1,6
DA3	3,1	1,7	2,7	1,6
DA4	3,0	1,7	2,6	1,6
EDMQ5	3,0	1,7	2,5	1,8
SDL1	2,6	1,8	2,3	1,7

Tabella 4: Statistiche bivariate per gruppo "Genere"

Variabile dipendente	F	P ($\leq 0,05$)
DA1	5,87	0,016
DA2	5,72	0,017
DA3	5,83	0,016
DA4	5,42	0,020
EDMQ5	7,79	0,006
SDL1	4,91	0,027

Tabella 5: ANOVA per gruppo "Genere"

Come si evince dalle tabelle, per quanto concerne le variabili che hanno mostrato differenze significative all'interno dei gruppi, gli uomini sembrano compiere attività basate sui dati con maggiore frequenza rispetto alle donne, concentrandosi però su processi di valutazione degli studenti (DA1, DA2, DA3, DA4), di monitoraggio dell'apprendimento e dell'efficacia dell'insegnamento, e per fornire feedback; più frequente, all'interno del gruppo degli uomini, è anche il ricorso alle analitiche di apprendimento ai fini del miglioramento della didattica e della progettazione formativa (EDMQ5), nonché dell'impiego dei dati per una riflessione critica, assieme agli studenti, sulla loro attendibilità e affidabilità in relazione alle fonti d'origine (SDL1).

Per meglio rappresentare le differenze tra i gruppi, sono stati elaborati i box plot attraverso il codice:

```
ggboxplot(dati_two_universities_so, x = "Gender",
          y = c("DA1", "DA2", "DA3", "DA4", "EDMQ5", "SDL1"),
          merge = TRUE, palette = "rainbow", fill = "grey")
```

L'output è qui rappresentato (Figura 1):

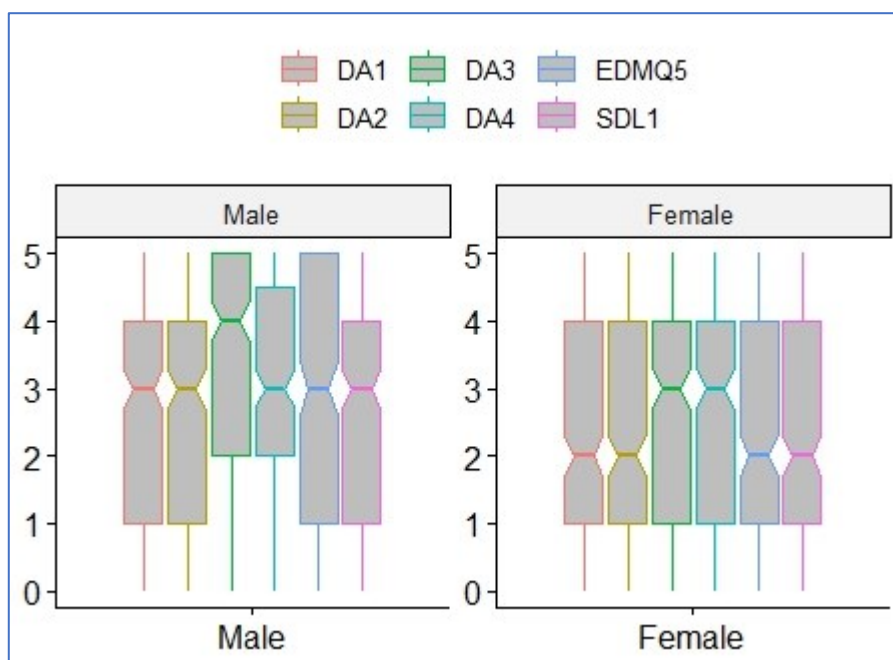


Figura 1: Box-plot affiancati per gruppo "Genere"

4.2.4. ANOVA per gruppo "Età"

La seconda variabile indipendente – "Età dei rispondenti" – è stata analizzata attraverso un test ANOVA in quanto omoschedastico. La scala relativa variava tra "Meno di 25", "25-34", "35-44", "45-54", e "Più di 55". Di nuovo, è stato eseguito il test:

```
aov_dati_age <- dati_ragg %>% anova_test(value ~ Age)
print(aov_dati_age, n = 35)
tab_aov_age <- data.frame(aov_dati_age)
View(tab_aov_age)
```

I risultati hanno mostrato che alcune variabili, appartenenti ai gruppi "Data for Assessment", "Data as Educational Resource", "Educational Management and Quality" e "Students' Data Literacy", avevano varianze diverse. Nello specifico, gli item interessati erano:

- "Uso dei dati dalle piattaforme di apprendimento per monitorare/valutare la didattica (DA6)";
- "Uso di sistemi digitali automatizzati per analizzare e valutare i risultati degli studenti (per esempio, quiz online) (DA8)";
- "Uso di dati di report internazionali/nazionali come risorsa didattica (DER1)";
- "Riutilizzo di Dati Aperti (Open Data) provenienti dalla ricerca o dalla pubblica amministrazione come risorsa didattica (DER2)";
- "Riutilizzo di Dati Aperti (Open Data) provenienti dalla propria ricerca come risorsa didattica (DER3)";

- “Riflessione su aspetti etici relative all’integrazione dei dati aperti come risorse didattiche (DER4)”;
- “Uso di dati da Report Nazionali per lo sviluppo istituzionale e la pianificazione (EDMQ1)”;
- “Uso di dati da Report Istituzionali per lo sviluppo istituzionale e la pianificazione (EDMQ2)”;
- “Uso di dati dalla valutazione del proprio corso per lo sviluppo istituzionale e la pianificazione (EDMQ3)”;
- “Uso di dati del proprio corso per la progettazione curricolare (EDMQ4)”;
- “Promozione di skills tra gli studenti per integrare dati in modo significativo entro report narrativi o presentazioni (SDL2)”.

I risultati dell’analisi sono qui riassunti (tabelle 6 e 7):

	DA6	DA8	DER1	DER2	DER3	DER4	EDMQ1	EDMQ2	EDMQ3	EDMQ4	SDL2
Media “25-34”	3,1	2,1	2,8	2,9	2,7	2,3	1,7	2,0	2,5	2,4	3,0
D.S. “25-34”	1,5	1,5	1,9	1,7	1,7	1,7	1,6	1,6	1,8	1,7	1,6
Media “35-44”	2,3	1,6	1,9	2,1	2,0	1,7	1,4	1,6	2,6	2,3	2,4
D.S. “35-44”	1,8	1,7	1,6	1,7	1,7	1,6	1,5	1,6	1,8	1,7	1,8
Media “45-54”	2,5	2,0	2,1	2,3	2,2	2,1	1,7	1,9	2,6	2,5	2,3
D.S. “45-54”	1,7	1,7	1,7	1,7	1,6	1,6	1,6	1,6	1,7	1,8	1,7
Media “Più di 55”	2,0	1,5	2,3	2,5	2,4	2,3	1,9	2,3	3,2	3,1	2,1
D.S. “Più di 55”	1,7	1,5	1,6	1,5	1,6	1,6	1,5	1,6	1,6	1,7	1,7

Tabella 6: Statistiche bivariante per gruppo “Età”

Variabile dipendente	F	P ($\leq 0,05$)
DA6	4,54	0,004
DA8	3,04	0,029
DER1	2,83	0,038
DER2	3,18	0,024
DER3	3,09	0,027
DER4	3,99	0,008
EDMQ1	3,04	0,029
EDMQ2	5,07	0,002
EDMQ3	3,54	0,015
EDMQ4	5,56	0,000934
SDL2	2,63	0,05

Tabella 7: ANOVA per gruppo "Età"

Per quanto riguarda le differenze nelle *data practices* in relazione all'età, si può notare come esse siano legate all'uso dei dati per il monitoraggio dell'insegnamento e per attribuire punteggi agli studenti (DA6, DA8), come risorsa didattica *anche* ai fini di una riflessione etica sugli stessi (DER1, DER2, DER3, DER4), per la gestione e la qualità dell'insegnamento, nell'ottica della progettazione educativa e la pianificazione istituzionale (EDMQ1, EDMQ2, EDMQ3, EDMQ4) e, infine, promuovendone l'uso per elaborare rappresentazioni grafiche o presentazioni all'interno di narrative capaci di attribuire loro un senso (SDL2).

Un numero rilevante di variabili (sette su undici), tuttavia, si dimostra debolmente significativo ($p \geq 0,1$, e $p \leq 0,5$); pertanto, l'effetto dell'età sulle pratiche connesse ai dati appare limitato.

Pertanto, sono stati realizzati i box-plot per rappresentare i risultati emersi (Figura 2), utilizzando:

```
windows()
boxplot_age <- ggboxplot(dati_two_universities_so,
  x = "Age",
  y = c("DA6", "DA8", "DER1", "DER2", "DER3", "DER4", "EDMQ1",
    "EDMQ2", "EDMQ3", "EDMQ4", "SDL2"),
  merge = TRUE, palette = "rainbow", fill = "grey",
  repel = TRUE, xlab = FALSE, ylab = FALSE)
facet(boxplot_age, facet.by = "Age", scales = "free")
```

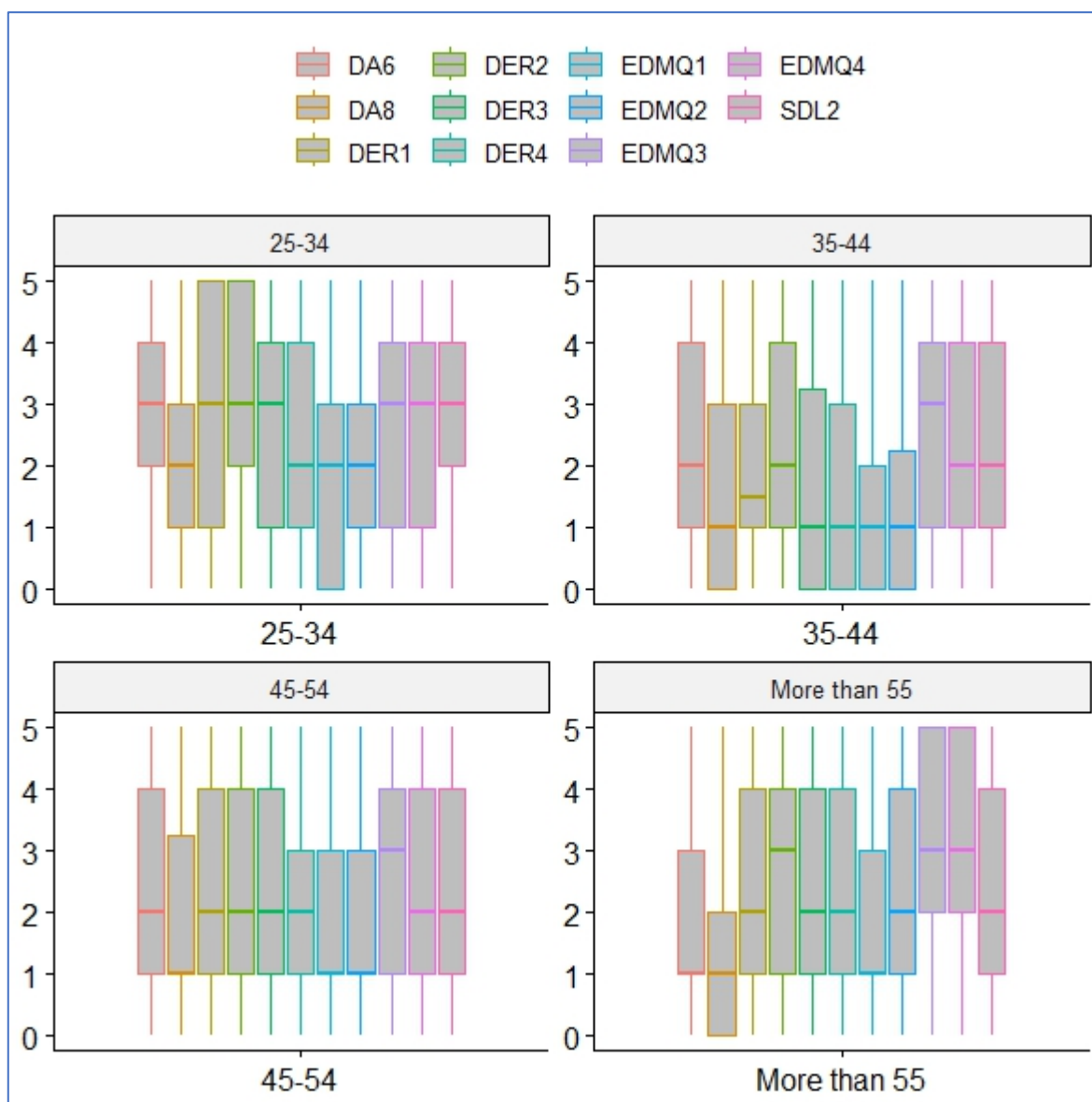


Figura 2: Box-plot affiancati per gruppo "Età"

4.2.5. ANOVA per gruppo "Nazione"

Un'altra variabile indipendente di sicuro interesse era quella relativa alla Nazione di residenza dei rispondenti – Spagna o Italia. Pertanto, è stato eseguito il test:

```
aov_dati_country <- dati_ragg %>% welch_anova_test(value ~ Country)
print(aov_dati_country, n = 35)
tab_aov_country <- data.frame(aov_dati_country)
View(tab_aov_country)
```

I risultati hanno mostrato marcate differenze tra i due gruppi in un elevato numero di variabili, di cui almeno una per gruppo. Nello specifico, si è trattato di DA1, DA2, DA6, DA7, DA9, DER2, DER3, DTL2, DTL5, EDMQ2, EDMQ3, EDMQ4, EDMQ5, EDMQ6, SDL1, SDL5, SED2.

Gli item (escludendo quelli già citati) erano quindi:

- “Uso di “dashboards” presenti nelle piattaforme di apprendimento per informare la mia didattica (DA7)”;
- “Uso di dati dalle piattaforme di apprendimento per riflettere insieme agli studenti sulla qualità della didattica (DA9)”;
- “Integrazione di tecniche estrattive di dati (massivi) come parte delle attività e compiti degli studenti (DTL2)”;
- “Integrazione di dashboards (pannelli di controllo visuali) per supportare le abilità autoregolative degli studenti (DTL5)”;
- “Uso di analitiche di apprendimento per analizzare l’efficacia della propria didattica (EDMQ6)”;
- “Discussione congiunta con gli studenti sull’impatto dell’uso dei dati nella società in generale (SDL5)”;
- “Condivisione di dati provenienti dalla piattaforma di apprendimento (analitiche di apprendimento) per supportare la riflessione degli studenti sul progresso dei propri apprendimenti nel contesto istituzionale o sociale (SED2)”.

Di seguito le tabelle contenenti i risultati (Tabelle 8 e 9):

Variabile dipendente	Media Italia	D.S. Italia	Media Spagna	D.S. Spagna
DA1	2,9	1,7	2,5	1,7
DA2	3,1	1,6	2,5	1,6
DA6	1,7	1,5	2,8	1,7
DA7	1,2	1,2	2,1	1,7
DA9	1,1	1,1	1,7	1,5
DER2	2,6	1,7	2,2	1,7
DER3	2,5	1,6	2,1	1,7
DTL2	1,1	1,2	1,4	1,5
DTL5	1,8	1,7	1,5	1,5
EDMQ2	2,4	1,6	1,6	1,5
EDMQ3	3,4	1,6	2,3	1,7
EDMQ4	3,7	1,5	1,9	1,6
EDMQ5	3,8	1,4	2,1	1,7
EDMQ6	1,7	1,6	2,6	1,6
SDL1	2,9	1,7	2,2	1,7
SDL5	1,9	1,6	1,6	1,6
SED2	2,1	1,6	1,5	1,5

Tabella 8: Statistiche bivariate per gruppo “Nazione”

Variabile dipendente	F	P ($\leq 0,05$)
DA1	9,43	2^{-3}
DA2	15,31	$1,08^{-4}$
DA6	50,49	$5,08^{-12}$
DA7	45,69	$4,01^{-11}$
DA9	18,15	$2,46^{-5}$
DER2	5,37	$2,1^{-2}$
DER3	7,14	8^{-3}
DTL2	7,50	6^{-3}
DTL5	4,49	$3,5^{-2}$
EDMQ2	37,55	$2,17^{-9}$
EDMQ3	45,58	$5,15^{-11}$
EDMQ4	165,67	$3,37^{-32}$
EDMQ5	140,79	$2,33^{-28}$
EDMQ6	36,18	$4,10^{-9}$
SDL1	24,20	$1,29^{-6}$
SDL5	3,97	$4,7^{-2}$
SED2	18,92	$1,76^{-5}$

Tabella 9: ANOVA per gruppo "Nazione"

Sono state riscontrate differenze notevoli tra i due gruppi in relazione all'università di appartenenza: come si può constatare, all'interno del caso italiano le medie sono spesso superiori a quelle del contesto spagnolo, con l'eccezione delle variabili DA6, DA7, DA9, DTL2, ed EDMQ6; ad indicare che, all'interno delle due realtà, le differenze si presentano in rapporto alle pratiche di valutazione attraverso i dati, in particolare per giudicare l'efficacia dell'insegnamento – anche con il coinvolgimento degli studenti (DA6, DA7, DA9); all'uso dei dati per il miglioramento delle competenze di *data literacy* degli studenti stessi, integrando nei compiti tecniche di *data extraction* (DTL2) e, infine, impiegando le informazioni ottenute dalle analitiche d'apprendimento per riflettere sull'efficacia della propria azione docente (EDMQ6).

All'interno del caso italiano, invece, si dimostrano più frequenti le pratiche legate al monitoraggio dell'apprendimento e alla valutazione dell'insegnamento (DA1, DA2), all'utilizzo dei dati aperti come risorsa educativa (DER2, DER3), per promuovere le capacità degli studenti di auto-regolarsi (DTL5), all'impiego delle informazioni per la pianificazione istituzionale e la progettazione formativa (EDMQ2, EDMQ3, EDMQ4, EDMQ5), per promuovere le abilità degli studenti di compiere valutazioni critiche sulla credibilità e affidabilità dei dati, nonché sulle implicazioni etiche connesse al loro uso (SDL1, SDL5) e, per concludere, anche per favorire la riflessione degli studenti sul loro apprendimento (SED2).

Tutte le variabili hanno elevata significatività, suggerendo un impatto notevole della variabile "Nazione" sulle *data practices* messe in atto in un determinato contesto.

Per visualizzare le differenze sono stati nuovamente realizzati i box-plot (Figura 3):

```

windows()
boxplot_country <- ggboxplot(dati_two_universities_so,
                             x = "Country",
                             y = c("DA1", "DA2", "DA6", "DA7", "DA9", "DER2", "DER3",
                                   "DTL2", "DTL5", "EDMQ2", "EDMQ3", "EDMQ4", "EDMQ5",
                                   "EDMQ6", "SDL1", "SDL5", "SED2"),
                             merge = TRUE, palette = "rainbow", fill = "grey",
                             repel = TRUE, xlab = FALSE, ylab = FALSE)
facet(boxplot_country, facet.by = "Country", scales = "free", ncol = 1, nrow = 2)

```

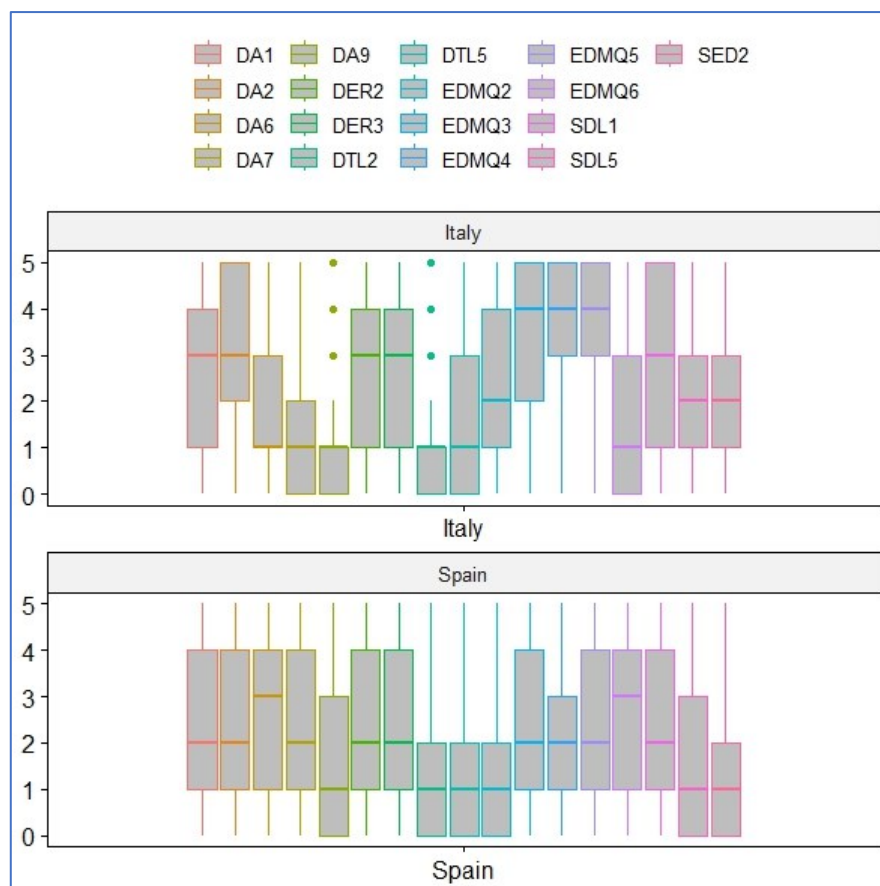


Figura 3: Box-plot affiancati per gruppo "Nazione"

4.2.6. ANOVA per gruppo "Area disciplinare"

È stata esaminata la relazione tra Area disciplinare dei rispondenti e gli item del questionario, anche in questo caso attraverso il Welch-ANOVA: sono emerse relazioni significative con variabili del gruppo "Data for Assessment", "Data as Educational Resource", "Data for Education Management and Quality", "Students' Data Literacy", e "Students' Empowerment through Data". Gli item interessati (sempre escludendo quelli già citati), in questo caso, erano:

- "Promozione della discussione e riflessione tra gli studenti sui vincoli sociali relativi all'uso dei dati e le applicazioni digitali nell'IA come processo socio-tecnico (SDL3)"

- “Promozione dell’analisi etica relativa alla raccolta e utilizzo dei propri dati come parte del processo di apprendimento (SDL4)”
- “Condivisione di report istituzionali per supportare la riflessione degli studenti sul progresso dei propri apprendimenti nel contesto istituzionale o sociale (SED1)”

Il test è stato eseguito tramite:

```
aov_dati_sf <- dati_ragg %>% welch_anova_test(value ~ Subjectfield_agg)
print(aov_dati_sf, n = 35)
tab_aov_sf <- data.frame(aov_dati_sf)
View(tab_aov_sf)
```

Le tabelle (Tabella 10 e 11) riassumono i risultati:

	DA6	DA8	DER1	DER2	EDMQ4	EDMQ5	SDL1	SDL3	SDL4	SDL5	SED1
Media “Scienze U. & Linguistica”	2,6	1,4	1,8	2,0	2,6	2,7	2,5	1,5	1,8	1,9	1,3
D.S. “Scienze U. & Linguistica”	1,8	1,6	1,4	1,5	1,7	1,7	1,8	1,5	1,6	1,7	1,2
Media “Scienze della vita”	2,2	1,9	2,0	2,2	2,8	2,9	2,8	1,5	1,5	1,9	1,8
D.S. “Scienze della vita”	1,6	1,7	1,7	1,7	1,8	1,8	1,6	1,4	1,4	1,5	1,6
Media “Scienze Formali & Naturali”	1,8	2,2	2,0	2,1	3,4	3,3	2,6	1,1	1,3	1,4	1,2
D.S. “Scienze Formali & Naturali”	1,6	1,7	1,5	1,6	1,4	1,6	1,8	1,1	1,2	1,4	1,1
Media “Scienze tecnologiche”	2,3	1,5	1,9	2,1	2,5	2,7	2,0	1,0	1,2	1,3	1,4
D.S. “Scienze tecnologiche”	1,8	1,4	1,7	1,6	1,8	1,8	1,7	1,1	1,3	1,4	1,5
Media “Scienze Sociali”	2,6	1,8	2,5	2,6	2,3	2,6	2,5	1,6	1,7	1,9	2,0
D.S. “Scienze Sociali”	1,7	1,7	1,7	1,7	1,8	1,8	1,8	1,6	1,6	1,7	1,7

Tabella 10: Statistiche bivariate per gruppo “Area disciplinare”

Variabile dipendente	F	P ($\leq 0,05$)
DA6	2,8	0,028
DA8	2,59	0,039
DER1	3,8	0,006
DER2	2,61	0,037
EDMQ4	6,86	0,000039
EDMQ5	2,6	0,038
SDL1	2,75	0,03
SDL3	3,79	0,006
SDL4	3,59	0,008
SDL5	4,35	0,002
SED1	4,84	0,000998

Tabella 11: ANOVA per gruppo “Area disciplinare”

Le differenze statisticamente significative, connesse alle pratiche di *data literacy* in base all’area disciplinare, riguardavano l’impiego dei dati per attività di *assessment* riguardanti il monitoraggio

dell'insegnamento e l'attribuzione di punteggi ai lavori degli studenti (DA6, DA8), l'utilizzo dei dati aperti come risorsa educativa (DER1, DER2), all'impiego dei dati – anche provenienti dai *learning analytics* – ai fini della progettazione formativa (EDMQ4, EDMQ5), al loro impiego per favorire la riflessione degli studenti sulle implicazioni sociali ed etiche legate all'innovazione tecnologica, nonché per stimolare il ragionamento critico rispetto all'affidabilità dei dati e al loro uso nel processo di apprendimento (SDL1, SDL3, SDL4, SDL5) e, infine, per supportare riflessioni da parte degli studenti sui loro apprendimenti, in relazione al contesto istituzionale e sociale (SED1).

In cinque casi su undici, le variabili si sono dimostrate come poco significative da un punto di vista statistico, con un valore di p compreso tra 0,028 e 0,038; ciò potrebbe denotare una debole influenza della variabile “Area disciplinare” sulle *data practices*.

Anche in questo caso, sono stati realizzati i box-plot, rappresentati in Figura 4:

```
windows()
boxplot_sf <- ggboxplot(dati_two_universities_so,
  x = "Subjectfield_agg",
  y = c("DA1", "DA2", "DA6", "DA7", "DA9", "DER2",
    "DER3", "DTL2", "DTL5", "EDMQ2", "EDMQ3", "EDMQ4",
    "EDMQ5", "EDMQ6", "SDL1", "SDL5", "SED2"),
  merge = TRUE, palette = "rainbow", fill = "grey",
  repel = TRUE, xlab = FALSE, ylab = FALSE)
facet(boxplot_sf, facet.by = "Subjectfield_agg", scales = "free", ncol = 2, nrow = 3)
```

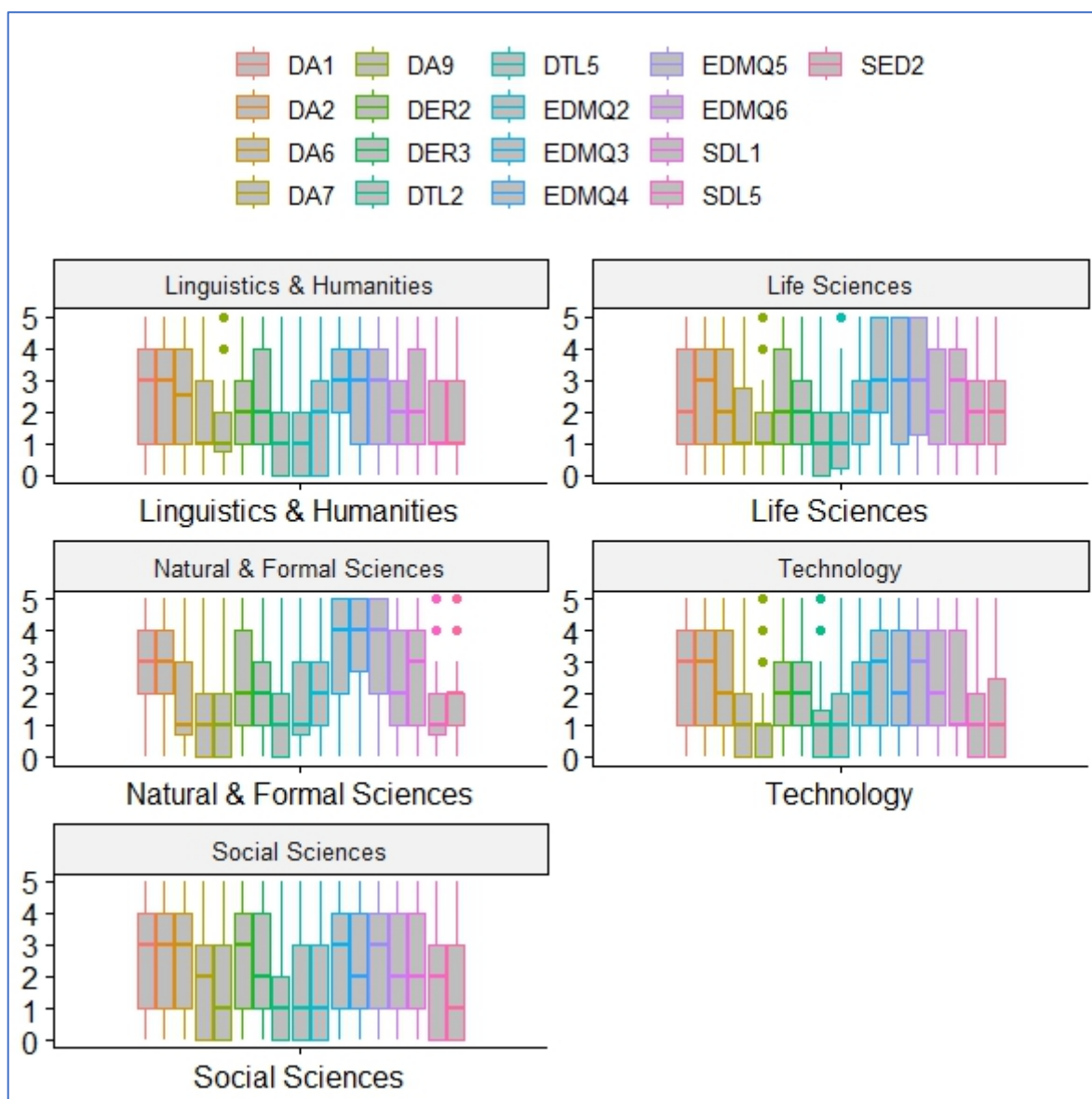



Figura 4: Box-plot affiancati per gruppo “Area disciplinare”

4.2.7. ANOVA per gruppo “Esperienza d’insegnamento”

È stata investigata (tramite Welch-ANOVA) la relazione tra anni d’esperienza di insegnamento e impiego di pratiche di *data literacy*. Quindi:

```
aov_dati_te <- dati_ragg %>% welch_anova_test(value ~ Teach.exp)
print(aov_dati_te, n = 35)
tab_aov_te <- data.frame(aov_dati_te)
View(tab_aov_te)
```

L’ANOVA ha evidenziato relazioni nei due gruppi “Data for Assessment” e “Data for Education Management and Quality”. Nello specifico, si trattava di DA7, EDMQ1, EDMQ2, EDMQ3, EDMQ4. Le tabelle 12 e 13 offrono una sintesi dell’analisi:

	DA7	EDMQ1	EDMQ2	EDMQ3	EDMQ4
Media “Meno di 3”	2,5	1,2	1,5	2,2	1,9
D.S. “Meno di 3”	1,8	1,5	1,5	1,8	1,7
Media “3-4”	1,6	1,6	1,7	2,6	2,1
D.S. “3-4”	1,7	1,6	1,6	1,8	1,7
Media “5-10”	1,7	1,3	1,5	2,7	2,5
D.S. “5-10”	1,4	1,4	1,4	1,7	1,7
Media “10-15”	1,7	2,7	2,5	2,7	2,1
D.S. “10-15”	1,6	1,7	1,8	1,8	1,8
Media “Più di 15”	1,7	1,9	2,2	3,0	2,8
D.S. “Più di 15”	1,6	1,6	1,7	1,7	1,8

Tabella 12: Statistiche bivariate per gruppo “Esperienza d’insegnamento”

Variabile dipendente	F	P ($\leq 0,05$)
DA7	2,96	0,021
EDMQ1	3,41	0,01
EDMQ2	4,10	0,003
EDMQ3	2,73	0,031
EDMQ4	4,97	0,000811

Tabella 13: ANOVA per gruppo “Esperienza d’insegnamento”

Per quanto concerne l’esperienza di insegnamento e le pratiche di *data literacy*, si osservano poche differenze all’interno dei gruppi; queste sono legate, perlopiù, all’uso delle informazioni ricavate dai *learning analytics* per il miglioramento della didattica (DA7), e all’impiego dei dati ottenuti da report nazionali o istituzionali, o provenienti dalla valutazione dei propri corsi, per la progettazione formativa e la pianificazione istituzionale (EDMQ1, EDMQ2, EDMQ3, EDMQ4).

Vista l’esiguità delle variabili significative (cinque), di cui tre (DA7, EDMQ1, ed EDMQ3) con basso livello di significatività, si può supporre una scarsa, se non forse nulla, influenza della variabile “Esperienza d’insegnamento” sulle pratiche di *data literacy*.

Per concludere, sono stati realizzati i box-plot relativi (Figura 5):

```

windows()
boxplot_te <- ggboxplot(dati_two_universities_so,
  x = "Teach.exp",
  y = c("DA7", "EDMQ1", "EDMQ2", "EDMQ3", "EDMQ4"),
  merge = TRUE, palette = "rainbow", fill = "grey",
  repel = TRUE, xlab = FALSE, ylab = FALSE)
facet(boxplot_te, facet.by = "Teach.exp", scales = "free", ncol = 2, nrow = 3)

```

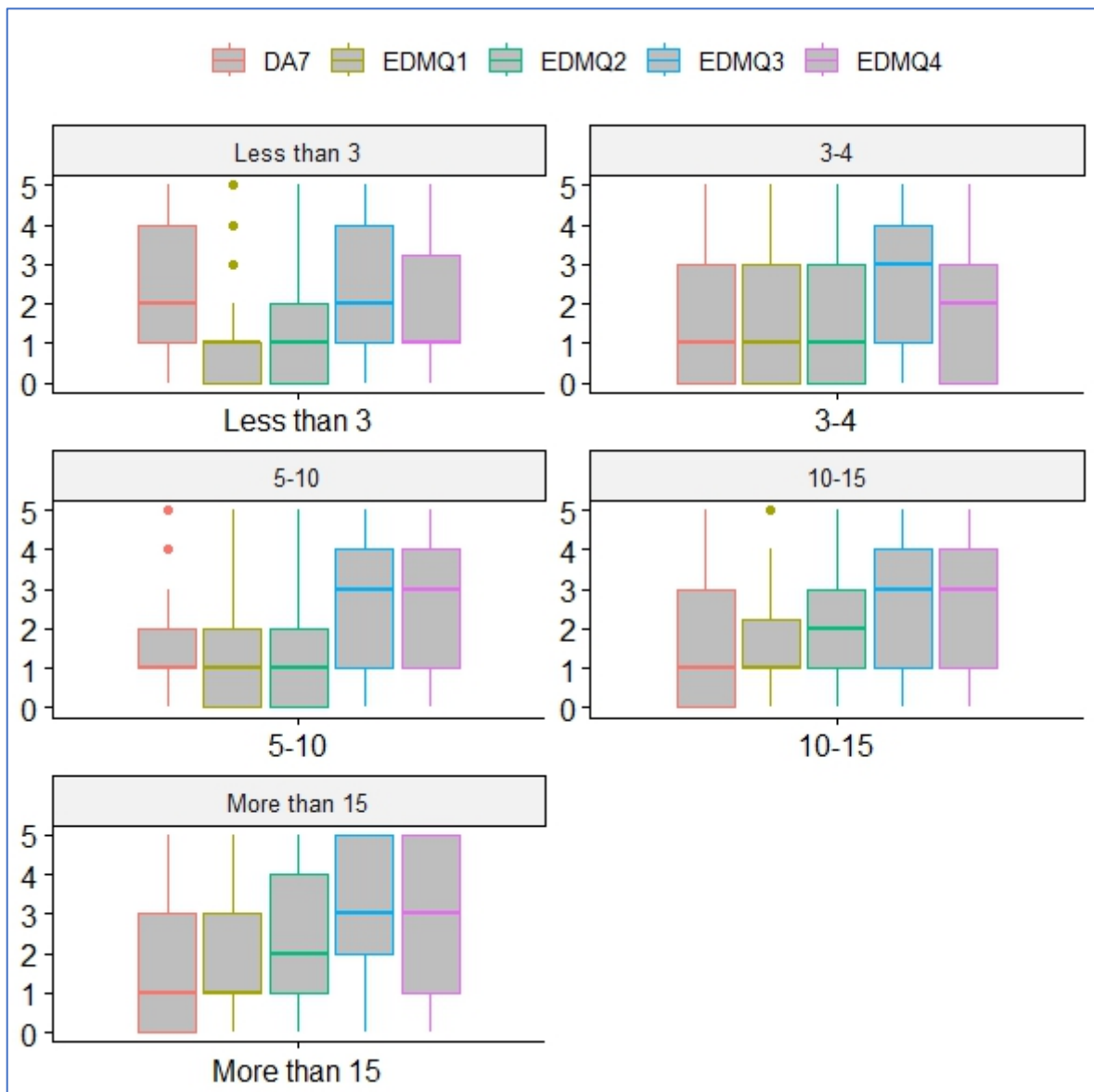


Figura 5: Box-plot affiancati per gruppo "Esperienza d'insegnamento"

4.2.8. ANOVA per gruppo “Esperienza di Ricerca”

Infine, si sono volute indagare le possibili relazioni tra gli anni di esperienza di ricerca e le variabili. Come in precedenza, è stato eseguito il Welch-ANOVA:

```
aov_dati_re <- dati_ragg %>% welch_anova_test(value ~ Research.exp)
print(aov_dati_re, n = 35)
tab_aov_re <- data.frame(aov_dati_re)
View(tab_aov_re)
```

Anche in questo caso sono state individuate relazioni statisticamente significative con variabili all'interno dei gruppi “Data for Assessment” e “Data for Education Management and Quality”, nonché “Data as Educational Resource”. Gli item interessati erano: DA6, DA7, DA9, DER2, DER3, EDMQ2, EDMQ3, EDMQ4, EDMQ5.

Le tabelle 14 e 15 contengono i risultati:

	DA6	DA7	DA9	DER2	DER3	EDMQ2	EDMQ3	EDMQ4	EDMQ5
Media “Nessuna esperienza”	3,0	2,3	1,7	1,7	1,4	1,3	2,2	1,5	1,8
D.S. “Nessuna esperienza”	1,7	1,8	1,6	1,6	1,7	1,5	1,8	1,6	1,7
Media “Meno di 3”	2,8	2,3	1,8	2,4	2,2	1,7	2,7	2,4	2,6
D.S. “Meno di 3”	1,6	1,6	1,4	1,5	1,6	1,5	1,5	1,6	1,6
Media “3-5”	3,0	2,4	1,8	2,4	2,4	1,8	2,3	2,2	2,2
D.S. “3-5”	1,7	1,8	1,6	1,6	1,6	1,6	1,6	1,7	1,8
Media “6-10”	2,2	1,7	1,4	2,5	2,6	1,8	2,5	2,4	2,7
D.S. “6-10”	1,7	1,7	1,5	1,7	1,7	1,7	1,7	1,8	1,8
Media “11-15”	2,3	1,6	1,5	2,6	2,3	1,9	2,8	2,7	3,3
D.S. “11-15”	1,8	1,6	1,4	1,8	1,7	1,6	1,7	1,6	1,6
Media “Più di 15”	2,3	1,6	1,5	2,6	2,3	1,9	2,8	2,7	3,3
D.S. “Più di 15”	1,8	1,6	1,4	1,8	1,7	1,6	1,7	1,6	1,6

Tabella 14: Statistiche bivariate per gruppo “Esperienza di ricerca”

Variabile dipendente	F	P ($\leq 0,05$)
DA6	5,57	$9,50^{-5}$
DA7	6,54	$1,55^{-5}$
DA9	2,38	$4,1^{-2}$
DER2	2,8	$1,9^{-2}$
DER3	4,09	2^{-3}
EDMQ2	4,18	1^{-3}
EDMQ3	3,17	9^{-3}
EDMQ4	10,46	$1,11^{-8}$
EDMQ5	7,87	$1,26^{-6}$

Tabella 15: ANOVA per gruppo "Esperienza di ricerca"

Seguendo un *pattern* già osservato per altre variabili, le differenze si osservano per quanto concerne le pratiche di valutazione basate sui dati, nello specifico per monitorare l'apprendimento, introdurre modifiche alla didattica, e come base per ragionare con gli studenti sulla qualità dell'insegnamento (DA6, DA7, DA9), come risorsa per l'insegnamento, sia da fonti istituzionali che dalla propria attività di ricerca (DER2, DER3), e per la progettazione formativa e la pianificazione istituzionale (EDMQ2, EDMQ3, EDMQ4, EDMQ5).

Osservando le statistiche, si può constatare come tutte abbiano un *p-value* minore di 0,1, indicando una elevata significatività. Per quanto concerne le pratiche di uso dei dati considerate, l'esperienza di ricerca sembra avere un ruolo rilevante.

I box-plot sono rappresentati di seguito (Figura 6):

```
windows()
boxplot_re <- ggboxplot(dati_two_universities_so,
  x = "Research.exp",
  y = c("DA6", "DA7", "DA9", "DER2", "DER3", "EDMQ2",
    "EDMQ3", "EDMQ4", "EDMQ5"),
  merge = TRUE, palette = "rainbow", fill = "grey",
  repel = TRUE, xlab = FALSE, ylab = FALSE)
facet(boxplot_re, facet.by = "Research.exp", scales = "free", ncol = 2, nrow = 3)
```

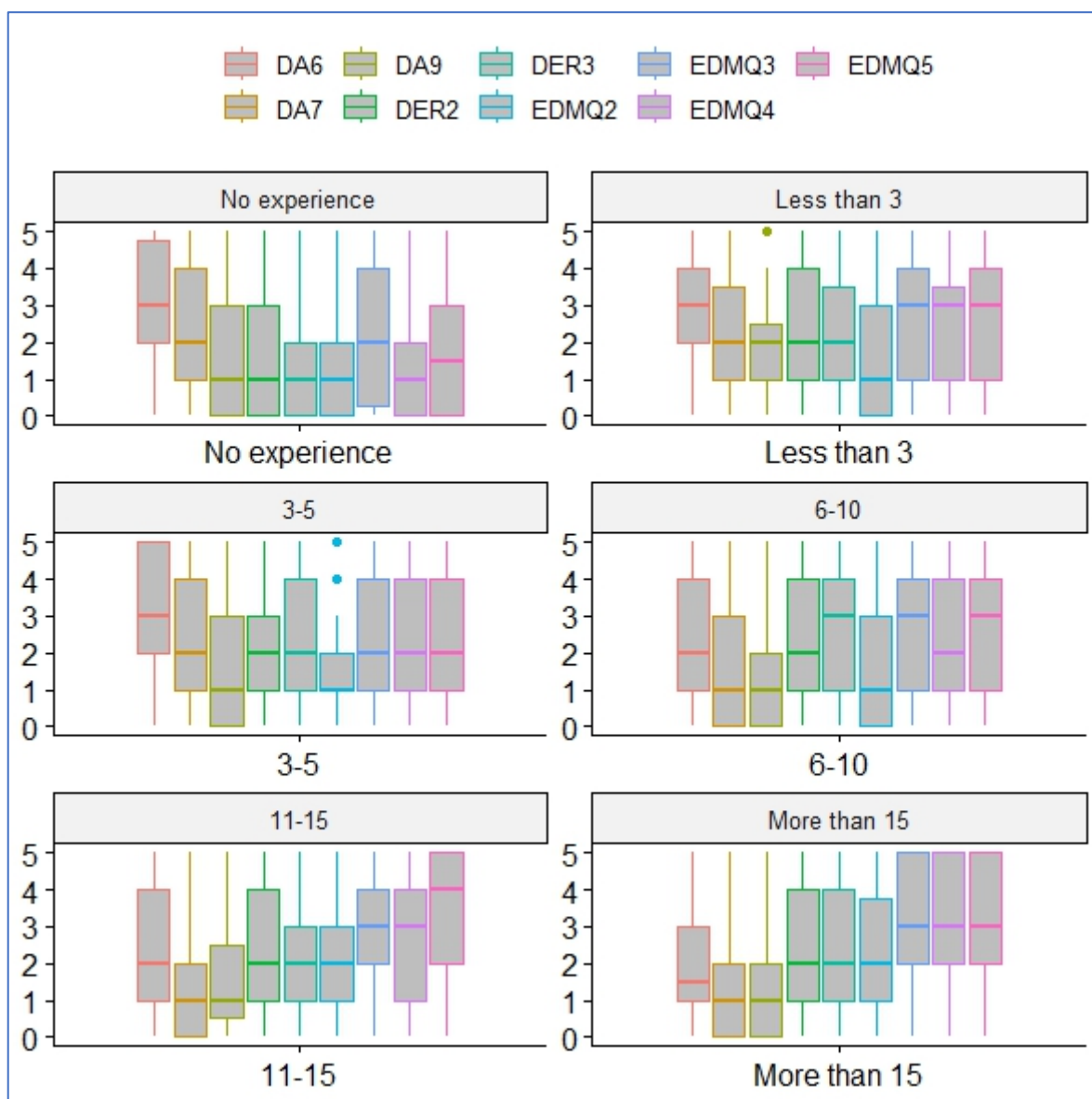


Figura 6: Box-plot affiancati per gruppo "Esperienza di ricerca"

4.3. Test multivariato di analisi della varianza: MANOVA

4.3.1. Premesse per il test MANOVA

Per testare l'associazione multivariata tra le variabili – in questo caso, il grado di associazione tra le diverse modalità delle variabili indipendenti e le risposte date al questionario – sono stati eseguiti dei test MANOVA sull'intero dataset a disposizione. Prima di fare ciò, tuttavia, è stato necessario verificare che le assunzioni del test fossero rispettate. Nel caso della MANOVA, e similmente alla ANOVA, le premesse per una corretta applicazione del test sono:

- 1) Assenza di outlier, sia univariati che multivariati;
- 2) Variabili distribuite secondo la curva normale, sia univariata che multivariata;
- 3) Varianze omogenee tra le variabili;
- 4) Multicollinearità: nessuna correlazione maggiore di 0.9.

Al fine di verificare tali assunti sono pertanto stati eseguiti i test corrispondenti: in primo luogo, sono stati rilevati ed esclusi gli outlier, sia univariati che multivariati: nel primo caso è stato utilizzato il metodo del Box-plot, mentre nel secondo la Mahalanobis distance; successivamente, è stato eseguito il test di Shapiro-Wilk per testare la normalità, sia univariata che multivariata, dei punteggi; il test non-parametrico di Levene per l'omogeneità delle varianze; infine, è stata realizzata la matrice di correlazione per verificare l'assunzione di multicollinearità.

4.3.1.1. Test per la presenza di outlier

In primo luogo, è stata controllata la presenza di *outlier* sia univariati che multivariati: verificare, ed eventualmente eliminare, punteggi anomali dal dataset è essenziale per ottenere statistiche affidabili e, più nello specifico, per assicurarsi di eseguire correttamente la MANOVA, la quale è vulnerabile a tali valori.

Allo stesso modo, anche gli assunti di normalità possono essere violati in presenza di molti, o anche solo alcuni, outlier: pertanto, utilizzando la funzione “*is_outlier*” per ogni variabile del dataset completo (N = 601), si è esaminata se tale premessa fosse, o meno, rispettata.

Il test ha rilevato, per le variabili EDMQ7, EDMQ8, e DER1, l'effettiva esistenza di valori anomali, sebbene in numero molto ristretto (rispettivamente 1, 2 e 1).

Successivamente, applicando la *Mahalanobis distance*, si sono ricercati outlier multivariati. Utilizzando sempre l'intero dataset, raccogliendo i punteggi a seconda delle variabili da analizzare (Genere, Età, Nazione, Area Disciplinare, Esperienza d'insegnamento ed Esperienza di ricerca),

sono stati identificati numerose anomalie: si è reso quindi necessario escludere 51 osservazioni dal dataset, sul quale sono poi state eseguite le successive analisi.

4.3.1.2. Test di normalità Shapiro-Wilk

Una volta elaborato il dataset “depurato” dalle anomalie, si è proseguito con le verifiche delle assunzioni di normalità, utilizzando il test di Shapiro-Wilk. L'output è visualizzato di seguito (Tabella 16):

Variabile	W	p
DA1	0,90	< ,001
DA10	0,82	< ,001
DA2	0,91	< ,001
DA3	0,89	< ,001
DA4	0,90	< ,001
DA5	0,91	< ,001
DA6	0,90	< ,001
DA7	0,86	< ,001
DA8	0,86	< ,001
DA9	0,84	< ,001
DER1	0,89	< ,001
DER2	0,91	< ,001
DER3	0,90	< ,001
DER4	0,89	< ,001
DTL1	0,90	< ,001
DTL2	0,82	< ,001
DTL3	0,88	< ,001
DTL4	0,87	< ,001
DTL5	0,84	< ,001
EDMQ1	0,85	< ,001
EDMQ2	0,89	< ,001
EDMQ3	0,90	< ,001
EDMQ4	0,89	< ,001
EDMQ5	0,89	< ,001
EDMQ6	0,90	< ,001
EDMQ7	0,85	< ,001
EDMQ8	0,80	< ,001
SDL1	0,89	< ,001
SDL2	0,90	< ,001
SDL3	0,82	< ,001
SDL4	0,86	< ,001
SDL5	0,88	< ,001
SED1	0,86	< ,001
SED2	0,87	< ,001
SED3	0,79	< ,001

Tabella 16: Test di Shapiro-Wilk

Dal momento che il p-value è significativo, si è accettata l'ipotesi alternativa che i punteggi non seguano la distribuzione normale univariata.

Successivamente, è stato svolto anche il test di normalità multivariata, il cui esito è stato nuovamente significativo: si è ottenuto un valore di W pari a 0,841, e un p-value di 7,73⁻²². Come in precedenza, si è accettata l'ipotesi alternativa che i punteggi non seguano una distribuzione normale, in questo caso multivariata.

4.3.1.3. Test non-parametrico di Levene per l'omoschedasticità

Dal momento che i punteggi non erano distribuiti su una curva gaussiana, si è scelto di procedere con il test non-parametrico di Levene per l'omogeneità delle varianze: trattandosi di un test centrato sulla mediana, anziché sulla media, è più resistente alle violazioni alla normalità.

Di conseguenza, sono state raccolte le variabili dipendenti in raggruppamenti creati sulla base delle diverse variabili indipendenti presenti nel dataset.

In primo luogo, per il gruppo "Genere", le variabili che hanno ottenuto un *p-value* significativo sono: DER1, DER2, DER3, DER4, SED2, SED3, DTL3, DA5, DA6, SDL3.

Per il gruppo "Nazione", invece, hanno mostrato valori statisticamente significativi le variabili: DA7, DA9, DTL2, EDMQ5, DA6, ed EDMQ8.

Per il gruppo "Area disciplinare": DA7, EDMQ4, SDL3, SDL4, SDL5, SED1.

All'interno del gruppo "Esperienza d'insegnamento", invece, sono stati riscontrati valori significativi del consuntivo-test nel solo caso della variabile DA7; per quanto concerne il gruppo "Esperienza di ricerca": DA7, DA9, EDMQ7, SDL1.

Infine, per il raggruppamento relativo all'età, non sono state rilevate varianze diverse: all'interno di questo gruppo, pertanto, le variabili sono omoschedastiche; gli altri gruppi, al contrario, sono eteroschedastici.

4.3.1.4. Test di multicollinearità

L'ultimo passo necessario per testare le ipotesi alla base di ANOVA e MANOVA è stato quello di elaborare la matrice di correlazione, per accertarsi che nessuna correlazione fosse troppo elevata (ovvero, r superiore a 0,9). Pertanto, utilizzando il comando "cor" di R, è stata realizzata la matrice di correlazione, di cui si riporta in calce la rappresentazione grafica (Figura 7):

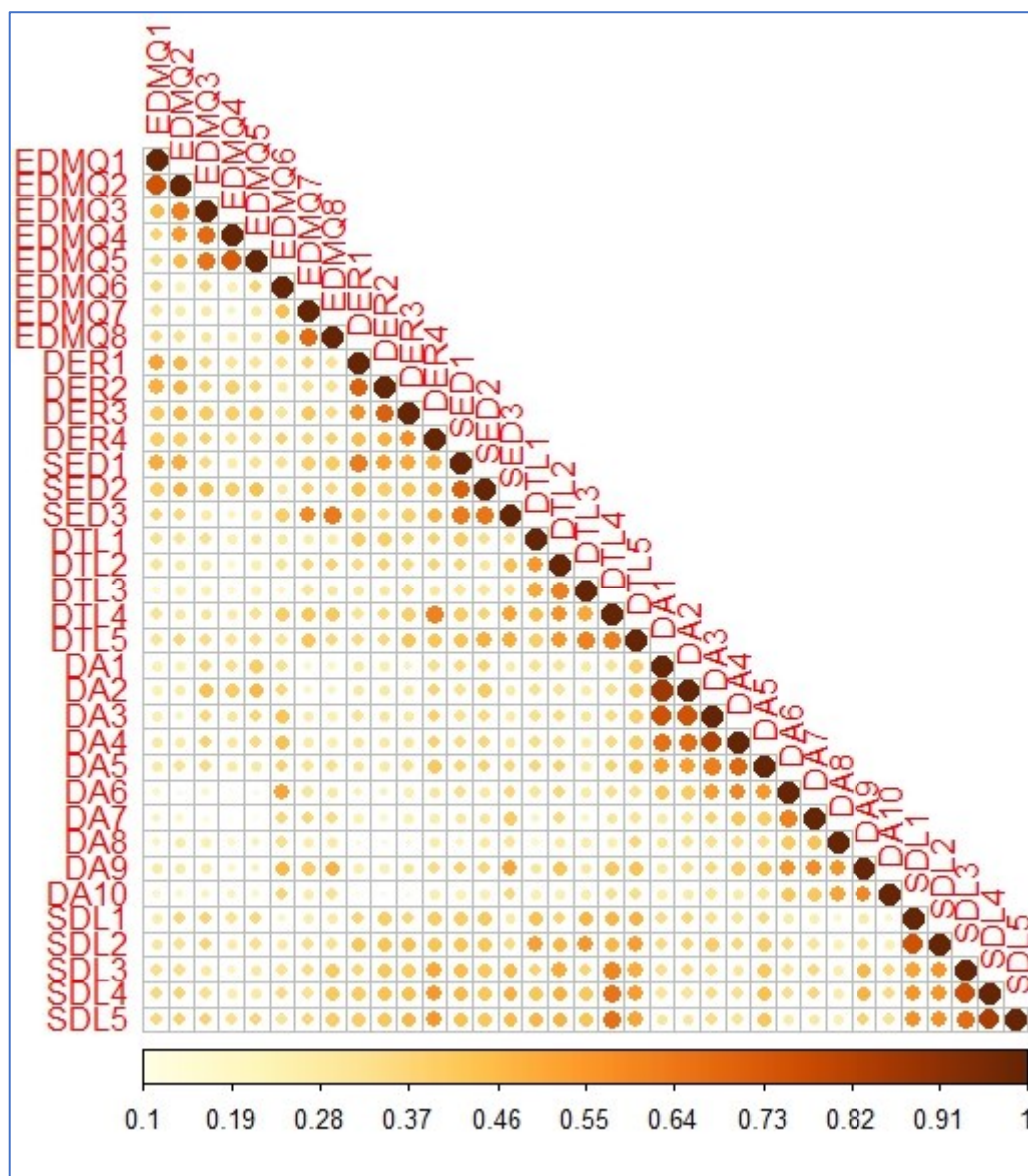


Figura 7: Grafico della matrice delle correlazioni

Come si evince dal grafico, nessun valore di R supera la soglia-limite di 0,9; l'assunto di multicollinearità è, quindi, rispettato.

4.3.2. Analisi multivariata della varianza: MANOVA

Tenendo in considerazione quanto emerso – dati non-normali, per lo più eteroschedastici, con la presenza di outlier sia univariati che multivariati – si è scelto di procedere ugualmente con la MANOVA, poiché:

- 1) come sottolineato in precedenza, i valori anomali, a cui sia l'ANOVA che la MANOVA sono sensibili, sono stati rimossi dal dataset;
- 2) l'assunto di normalità può essere violato, in quanto in presenza di campioni con elevata numerosità (in questo caso, $N = 494$), la MANOVA risulta comunque resistente ed in grado di fornire statistiche affidabili;
- 3) sebbene eteroschedastici, i gruppi presentano differenze nelle varianze molto ridotte.

Si è quindi passati all'esecuzione della MANOVA, analizzando la relazione tra gruppi e risposte al questionario e utilizzando un livello di significatività p di 0,05.

4.3.2.1. MANOVA per gruppo “Genere”

La prima relazione esaminata è stata quella tra “Genere” e risposte, attraverso il codice:

```
MANOVA_Gender_allvars_so <- manova(cbind(EDMQ1, EDMQ2, EDMQ3, EDMQ4,
    EDMQ5, EDMQ6, EDMQ7, EDMQ8,
    DER1, DER2, DER3, DER4,
    SED1, SED2, SED3,
    DTL1, DTL2, DTL3, DTL4, DTL5,
    DA1, DA2, DA3, DA4, DA5,
    DA6, DA7, DA8, DA9, DA10,
    SDL1, SDL2, SDL3, SDL4, SDL5)
    ~ Gender, dati_MANOVA_so)
summary(MANOVA_Gender_allvars_so)
summary.aov(MANOVA_Gender_allvars_so)
```

L'output risultante è stato il seguente (Tabella 17):

	G.d.L.	Pillai	F appros.	n G.d.L.	Den. G.d.L.	Pr(>F)
Genere	1	0,068423	0,96113	35	458	0,5349
Residui	492					

Tabella 17: MANOVA per gruppo “Genere”

Il test ha dato un esito statisticamente non significativo ($p\text{-value} > 0,05$, in questo caso uguale a 0,53). Approfondendo l'analisi (comando “*summary.aov*”), invece, è emerso che le variabili

EDMQ5, DA1, DA2, DA3, DA4, ed SDL1, presentano differenze statisticamente significative tra i generi.

In calce la sintesi dell'analisi (Tabella 18):

Variabile dipendente	Variabile Indipendente	G.d.L.	Somma dei quadrati	Media dei quadrati	F-value	Pr(>F)
EDMQ5	Genere	1	24,1	24,1	7,79	0,005
	Residui	492	1522,3	3,09		
DA1	Genere	1	16,22	16,22	5,87	0,016
	Residui	492	1360,20	2,76		
DA2	Genere	1	15,59	15,59	5,72	0,017
	Residui	492	1341,42	2,73		
DA3	Genere	1	16,41	16,41	5,83	0,016
	Residui	492	1384,35	2,81		
DA4	Genere	1	14,91	14,91	5,42	0,020
	Residui	492	1353,05	2,75		
SDL1	Genere	1	15,0	15,0	4,91	0,027
	Residui	492	1501,7	3,05		

Tabella 18: MANOVA per gruppo "Genere"

Il test MANOVA conferma quanto osservato nell'ANOVA ad una via: le variabili statisticamente significative sono le stesse. Tuttavia, è da sottolineare come il *p-value* ottenuto sia, quasi sempre, scarsamente significativo ($P \geq 0,1$), eccetto che per l'item EDMQ5 ($p = 0,005$).

Come suggerito anche dalla non significatività del test, si può inferire un debole effetto del genere sulle pratiche di *data literacy*.

4.3.2.2. MANOVA per gruppo "Età"

Il passaggio successivo ha riguardato l'analisi delle relazioni tra Età e risposte.

```
MANOVA_Age_allvars_so <- manova(cbind(EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5,
EDMQ6, EDMQ7, EDMQ8,
DER1, DER2, DER3, DER4,
SED1, SED2, SED3,
DTL1, DTL2, DTL3, DTL4, DTL5,
DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,
SDL1, SDL2, SDL3, SDL4, SDL5) ~ Age, dati_MANOVA_so)
summary(MANOVA_Age_allvars_so)
summary.aov(MANOVA_Age_allvars_so)
```

Di seguito l'output del test (Tabella 19):

	G.d.L.	Pillai	F appros.	n G.d.L.	Den. G.d.L.	Pr(>F)
Età	3	0,028975	1,399	105	1374	0,006
Residui	490					

Tabella 19: MANOVA per gruppo "Età"

Diversamente dalla MANOVA relativo al Genere, il test concernente l'Età ha dato un esito statisticamente significativo (p -value uguale a 0,006). Sembra pertanto sussistere un rapporto tra l'età dei rispondenti e i comportamenti di *data literacy*. Nello specifico (Tabella 20):

<i>Variabile dipendente</i>	<i>Variabile Indipendente</i>	<i>G.d.L.</i>	<i>Somma dei quadrati</i>	<i>Media dei quadrati</i>	<i>F-value</i>	<i>Pr(>F)</i>
<i>EDMQ1</i>	<i>Età</i>	3	21,68	7,23	3,04	0,03
	<i>Residui</i>	490	1164,13	2,38		
<i>EDMQ2</i>	<i>Età</i>	3	38,13	12,71	5,07	0,002
	<i>Residui</i>	490	1227,97	2,50		
<i>EDMQ3</i>	<i>Età</i>	3	30,83	10,28	3,54	0,015
	<i>Residui</i>	490	1423,28	2,90		
<i>EDMQ4</i>	<i>Età</i>	3	51,05	17,02	5,56	0,0009
	<i>Residui</i>	490	1522,3	3,09		
<i>DER1</i>	<i>Età</i>	3	23,53	7,84	2,83	0,04
	<i>Residui</i>	490	1356,19	2,77		
<i>DER2</i>	<i>Età</i>	3	25,95	8,65	3,18	0,02
	<i>Residui</i>	490	1331,55	2,72		
<i>DER3</i>	<i>Età</i>	3	25,43	8,48	3,09	0,03
	<i>Residui</i>	490	1360,20	2,76		
<i>DER4</i>	<i>Età</i>	3	31,47	10,49	3,99	0,008
	<i>Residui</i>	490	1288,13	2,63		
<i>DA6</i>	<i>Età</i>	3	39,17	13,06	4,53	0,004
	<i>Residui</i>	490	1410,70	2,879		
<i>DA8</i>	<i>Età</i>	3	24,37	8,12	3,04	0,03
	<i>Residui</i>	490	1384,35	2,81		
<i>SDL2</i>	<i>Età</i>	3	23,25	7,75	2,63	0,049
	<i>Residui</i>	490	1443,63	2,95		

Tabella 20: MANOVA per gruppo "Età"

Anche in questo caso, la MANOVA conferma quanto già constatato con l'ANOVA; a ciò si aggiunge però l'evidenza di una relazione tra l'età e le competenze di *data literacy* – sebbene tale relazione non sia sempre forte. Sette variabili, sulle undici considerate, sono debolmente significative, avendo un valore della statistica p compreso tra 0,1 e 0,5.

In conclusione, l'età sembra essere un fattore influente sulle pratiche legate ai dati, ma essa appare, in ogni caso, modesta.

4.3.2.3. MANOVA per gruppo "Nazione"

È stata poi esaminata la relazione tra la Nazione dei rispondenti e i comportamenti di Alfabetizzazione ai dati.

```
MANOVA_Country_allvars_so <- manova(cbind(EDMQ1, EDMQ2, EDMQ3, EDMQ4,
EDMQ5, EDMQ6, EDMQ7, EDMQ8,
```

DER1, DER2, DER3, DER4,
 SED1, SED2, SED3,
 DTL1, DTL2, DTL3, DTL4, DTL5,
 DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,
 SDL1, SDL2, SDL3, SDL4, SDL5) ~ Country, dati_MANOVA_so)

summary(MANOVA_Country_allvars_so)
 summary.aov(MANOVA_Country_allvars_so)

La Tabella 21 rappresenta i risultati:

	G.d.L.	Pillai	F appros.	n G.d.L.	Den. G.d.L.	Pr(>F)
Età	1	0,59695	19,381	35	458	<2,2 ⁻¹⁶
Residui	492					

Tabella 21: MANOVA per gruppo "Nazione"

Come si evince dalla tabella, il test ha dato un esito altamente significativo (*p-value* inferiore a 2,2⁻¹⁶): il rapporto tra la nazionalità dei rispondenti, e i comportamenti di competenza digitale, appare decisamente rilevante. In particolare, è emerso quanto segue (Tabella 22):

Variabile dipendente	Variabile Indipendente	G.d.L.	Somma dei quadrati	Media dei quadrati	F-value	Pr(>F)
EDMQ2	Nazione	1	90,31	90,31	37,79	1,631 ⁻⁹
	Residui	492	1175,79	2,39		
EDMQ3	Nazione	1	122,37	122,37	45,21	4,922 ⁻¹¹
	Residui	492	1227,97	2,50		
EDMQ4	Nazione	1	379,42	379,42	159,26	<2,2 ⁻¹⁶
	Residui	492	1172,13	2,38		
EDMQ5	Nazione	1	325,89	325,89	131,37	<2,2 ⁻¹⁶
	Residui	492	1220,49	2,48		
EDMQ6	Nazione	1	98,13	98,13	36,14	3,575 ⁻⁹
	Residui	492	1335,76	2,715		
DER2	Nazione	1	14,65	14,65	5,37	0,021
	Residui	492	1342,85	2,73		
DER3	Nazione	1	19,51	19,51	7,11	0,0079
	Residui	492	1360,20	2,76		
SED2	Nazione	1	46,36	46,36	19,64	1,155 ⁻⁵
	Residui	492	1161,43	2,36		
DTL2	Nazione	1	12,96	12,96	6,74	0,0097
	Residui	492	946,27	1,92		
DTL5	Nazione	1	12,1	12,1	4,75	0,03
	Residui	492	1254,50	2,55		
DA1	Nazione	1	25,82	25,82	9,41	0,0023
	Residui	492	1350,59	2,74		
DA2	Nazione	1	41,05	41,05	15,35	0,0001
	Residui	492	1315,96	2,67		
DA6	Nazione	1	129,37	129,37	48,20	1,221 ⁻¹¹
	Residui	492	1320,49	2,68		
DA7	Nazione	1	96,65	96,65	39,01	9,148 ⁻¹⁰

	Residui	492	1219,16	2,48		
DA9	Nazione	1	30,92	30,92	15,72	8,427 ⁻⁵
	Residui	492	967,69	1,97		
SDL1	Nazione	1	72,07	72,07	24,54	1 ⁻⁶
	Residui	492	1444,62	2,94		
SDL5	Nazione	1	9,6	9,6	3,96	0,047
	Residui	492	1192,3	2,42		

Tabella 22: MANOVA per gruppo "Nazione"

Ciò che è utile sottolineare – oltre alla significatività statistica degli item ottenuta sia con la MANOVA che con l'ANOVA – è che la nazione di residenza influenza notevolmente le pratiche d'uso dei dati: come emerge chiaramente dal test, si riscontra un l'elevato livello di significatività delle variabili considerate, eccetto che per DER2 ($p = 0,021$), DTL5 ($p = 0,03$), ed SDL5 ($p = 0,047$).

4.3.2.4. MANOVA per gruppo "Area disciplinare"

È stata analizzata anche la relazione tra Alfabetizzazione digitale e l'area disciplinare di appartenenza dei rispondenti.

```
MANOVA_Subjectfield_agg_allvars_so <- manova(cbind(EDMQ1, EDMQ2, EDMQ3, EDMQ4,
  EDMQ5, EDMQ6, EDMQ7, EDMQ8,
  DER1, DER2, DER3, DER4,
  SEDI, SED2, SED3,
  DTL1, DTL2, DTL3, DTL4, DTL5,
  DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,
  SDL1, SDL2, SDL3, SDL4, SDL5) ~ Subjectfield_agg,
  dati_MANOVA_so)
summary(MANOVA_Subjectfield_agg_allvars_so)
summary.aov(MANOVA_Subjectfield_agg_allvars_so)
```

Di cui i risultati, riassunti in Tabella 23:

	G.d.L.	Pillai	F appros.	n G.d.L.	Den. G.d.L.	Pr(>F)
Area disciplinare	4	0,39101	1,4178	140	1832	0,001375
Residui	489					

Tabella 23: MANOVA per gruppo "Area disciplinare"

Anche in questo caso, è stata individuata una relazione statisticamente significativa (p uguale a 0,001375). Specificamente (Tabella 24):

<i>Variabile dipendente</i>	<i>Variabile Indipendente</i>	<i>G.d.L.</i>	<i>Somma dei quadrati</i>	<i>Media dei quadrati</i>	<i>F-value</i>	<i>Pr(>F)</i>
<i>EDMQ4</i>	<i>Area disciplinare</i>	4	66,07	16,52	5,44	0,00027
	<i>Residui</i>	489	1485,48	3,04		
<i>EDMQ5</i>	<i>Area disciplinare</i>	4	30,19	7,55	2,43	0,04652
	<i>Residui</i>	489	1516,18	3,10		
<i>DER1</i>	<i>Area disciplinare</i>	4	40,13	10,03	3,66	0,006
	<i>Residui</i>	489	1339,59	2,74		
<i>DER2</i>	<i>Area disciplinare</i>	4	28,58	7,14	2,63	0,034
	<i>Residui</i>	489	1328,92	2,72		
<i>SED1</i>	<i>Area disciplinare</i>	4	38,61	9,65	4,07	0,003
	<i>Residui</i>	489	1159,75	2,37		
<i>DA6</i>	<i>Area disciplinare</i>	4	30,73	7,68	2,65	0,03
	<i>Residui</i>	489	1419,13	2,90		
<i>DA8</i>	<i>Area disciplinare</i>	4	25,62	6,41	2,39	0,0496
	<i>Residui</i>	489	1307,88	2,67		
<i>SDL1</i>	<i>Area disciplinare</i>	4	31,43	7,86	2,59	0,036
	<i>Residui</i>	489	1485,26	3,04		
<i>SDL3</i>	<i>Area disciplinare</i>	4	25,31	6,33	3,08	0,016
	<i>Residui</i>	489	1004,79	2,05		
<i>SDL4</i>	<i>Area disciplinare</i>	4	27,7	6,92	3,24	0,012
	<i>Residui</i>	489	1046	2,14		
<i>SDL5</i>	<i>Area disciplinare</i>	4	36,67	9,17	3,85	0,0043
	<i>Residui</i>	489	1165,20	2,38		

Tabella 24: MANOVA per gruppo "Area disciplinare"

La MANOVA conferma le relazioni osservate tramite l'ANOVA; tuttavia, numerose variabili (sette su undici) appaiono scarsamente significative.

I rapporti tra area disciplinare e uso dei dati del proprio corso per la progettazione curricolare (EDMQ4), l'uso dei dati, da report nazionali/internazionali, come risorsa educativa (DER1), la condivisione di report istituzionali ai fini della riflessione degli studenti sul proprio apprendimento

(SED1), e la discussione con gli studenti sull'impatto dei dati nella società (SDL5), si rivelano invece più forti.

Nell'insieme, si osserva un'influenza dell'Area disciplinare sui comportamenti di *data literacy*, ma anche in questo caso l'effetto appare limitato.

4.3.2.5. MANOVA per gruppo “Esperienza d’insegnamento”

La penultima variabile analizzata è stata l'Esperienza d'insegnamento:

```
MANOVA_Teach.exp_allvars_so <- manova(cbind(EDMQ1, EDMQ2, EDMQ3, EDMQ4,
EDMQ5, EDMQ6, EDMQ7, EDMQ8,
      DER1, DER2, DER3, DER4,
      SED1, SED2, SED3,
      DTL1, DTL2, DTL3, DTL4, DTL5,
      DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,
      SDL1, SDL2, SDL3, SDL4, SDL5) ~ Teach.exp, dati_MANOVA_so)
summary(MANOVA_Teach.exp_allvars_so)
summary.aov(MANOVA_Teach.exp_allvars_so)
```

Di cui (Tabella 25):

	G.d.L.	Pillai	F appros.	n G.d.L	Den. G.d.L.	Pr(>F)
Esperienza d'insegnamento	4	0,3211	1,1422	140	1832	0,13
Residui	489					

Tabella 25: MANOVA per gruppo “Esperienza d’insegnamento”

Anche in questo caso ci si è trovati di fronte ad un *p-value* superiore a 0,05 ($p = 0,13$), dimostrando l'insussistenza di relazioni tra l'esperienza di insegnamento e le variabili investigate. Tra i singoli item, invece, sono state rilevate le seguenti statistiche, riassunte in Tabella 26:

Variabile dipendente	Variabile Indipendente	G.d.L.	Somma dei quadrati	Media dei quadrati	F-value	Pr(>F)
EDMQ1	Esperienza d'insegnamento	4	30,57	7,64	3,23	0,01
	Residui	489	1155,24	2,36		
EDMQ2	Esperienza d'insegnamento	4	39,51	9,88	3,94	0,004
	Residui	489	1226,58	2,51		
EDMQ3	Esperienza d'insegnamento	4	33,17	8,29	2,85	0,02
	Residui	489	1420,94	2,90		
EDMQ4	Esperienza d'insegnamento	4	60,73	15,18	4,98	0,0006
	Residui	489	1490,82	3,05		

<i>DA7</i>	Esperienza d'insegnamento	4	37,39	9,35	3,58	0,0069
	Residui	489	1278,42	2,61		

Tabella 26: MANOVA per gruppo "Esperienza d'insegnamento"

Anche in questo caso, si conferma quanto ricavato dall'ANOVA. Oltre a ciò, è necessario evidenziare, da un lato, l'esiguità delle variabili che superano la soglia di significatività e, dall'altro, come queste mostrino (eccetto EDMQ1 ed EDMQ2), una relazione abbastanza forte con l'anzianità di insegnamento. L'effetto di questa variabile sulle pratiche d'uso dei dati può essere considerato, in ultima analisi, molto limitato.

4.3.2.6. MANOVA per gruppo "Esperienza di ricerca"

Infine, l'ultima relazione investigata è stata quella tra Esperienza di ricerca e comportamenti di *data literacy*.

```
MANOVA_Research.exp_allvars_so <- manova(cbind(EDMQ1, EDMQ2, EDMQ3, EDMQ4,
EDMQ5, EDMQ6, EDMQ7, EDMQ8,
DER1, DER2, DER3, DER4,
SED1, SED2, SED3,
DTL1, DTL2, DTL3, DTL4, DTL5,
DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,
SDL1, SDL2, SDL3, SDL4, SDL5) ~ Research.exp,
dati_MANOVA_so)
summary(MANOVA_Research.exp_allvars_so)
summary.aov(MANOVA_Research.exp_allvars_so)
```

Del quale l'output è stato (Tabella 27):

	G.d.L.	Pillai	F appros.	n G.d.L.	Den. G.d.L.	Pr(>F)
Esperienza di ricerca	5	0,55427	1,6315	175	2290	9,361 ⁻⁷
Residui	488					

Tabella 27: MANOVA per gruppo "Esperienza di ricerca"

La MANOVA si è rivelata essere altamente significativa ($P = 9,361^{-7}$). Tra le variabili dipendenti, invece, sono risultate statisticamente legate all'esperienza di ricerca (Tabella 28):

Variabile dipendente	Variabile Indipendente	G.d.L.	Somma dei quadrati	Media dei quadrati	F-value	Pr(>F)
<i>EDMQ2</i>	Esperienza di ricerca	5	49,3	9,86	3,95	0,0016
	Residui	488	1216,8	2,49		
<i>EDMQ3</i>	Esperienza di ricerca	5	48,95	9,79	3,40	0,005
	Residui	488	1405,16	2,88		

<i>EDMQ4</i>	<i>Esperienza di ricerca</i>	5	145,52	29,10	10,10	3,226 ⁻⁹
	<i>Residui</i>	488	1406,04	2,88		
<i>EDMQ5</i>	<i>Esperienza di ricerca</i>	5	114,45	22,89	7,80	4,459 ⁻⁷
	<i>Residui</i>	488	1431,92	2,93		
<i>DER2</i>	<i>Esperienza di ricerca</i>	5	37,95	7,59	2,80	0,01638
	<i>Residui</i>	488	1319,54	2,70		
<i>DER3</i>	<i>Esperienza di ricerca</i>	5	57,76	11,55	4,30	7,761 ⁻⁴
	<i>Residui</i>	488	1311,49	2,69		
<i>DA6</i>	<i>Esperienza di ricerca</i>	5	77,63	15,53	5,52	5,865 ⁻⁵
	<i>Residui</i>	488	1372,23	2,81		
<i>DA7</i>	<i>Esperienza di ricerca</i>	5	87,45	17,49	6,95	2,778 ⁻⁶
	<i>Residui</i>	488	1228,36	2,52		
<i>DA9</i>	<i>Esperienza di ricerca</i>	5	23,40	4,68	2,34	0,04
	<i>Residui</i>	488	975,21	2		

Tabella 28: MANOVA per gruppo "Esperienza di ricerca"

Per concludere, la MANOVA conferma quanto visto in precedenza nell'ANOVA; delle variabili esaminate, solamente due appaiono scarsamente significative (DER2, con $p = 0,02$, e DA9, con $p = 0,04$).

L'esperienza di ricerca, come suggerito dal test, sembra influire sui comportamenti legati ai dati, sebbene anche in quest'ultimo in caso in maniera ridotta.

4.3.2.7. Dimensione dell'effetto (“*Eta squared*”)

È stata calcolata la dimensione dell'effetto (*effect size*) tramite la funzione “*eta_squared*” del pacchetto “*effectsize*”, per misurare l'effettiva influenza delle variabili indipendenti sulle variabili dipendenti: essa si è generalmente dimostrata moderata (valore di η^2 compreso tra 0,06 e 0,14; Chiorri, 2014, p. 436), fatta eccezione per la variabile “Nazione” ($\eta^2 = 0,60$), in cui il valore trovato dimostra un'elevata incidenza del fattore sul comportamento.

I risultati sono riassunti nella tabella seguente (Tabella 29):

Variabile	η^2	Intervallo di confidenza (0.95)	
		Limite inferiore	Limite superiore
Nazione	0,60	0,54	1,00
Esperienza di ricerca	0,11	0,12	1,00
Età	0,10	0,01	1,00
Area disciplinare	0,10	0,01	1,00
Esperienza d'insegnamento	0,08	0	1,00
Genere	0,07	0	1,00

Tabella 29: Dimensione dell'effetto delle variabili indipendenti

La variabile “Nazione” risulta avere una incidenza molto elevata e decisamente superiore a tutte le altre ($\eta^2 = 0,6$); l'Esperienza di ricerca, infatti – la seconda variabile per dimensione dell'effetto – sembra avere un impatto molto più contenuto sui comportamenti di *data literacy* ($\eta^2 = 0,11$).

Età e Area Disciplinare sembrano avere la stessa influenza rispetto all'alfabetizzazione ai dati, con valore di η^2 uguale a 0,1; mentre le ultime due variabili considerate – Esperienza d'insegnamento e Genere – risultano le meno rilevanti in questo senso, ottenendo un η^2 appena sopra la soglia minima per rientrare nel range “moderato”.

4.3.2.8. Test Post-Hoc di Games-Howell

Al fine di attestare l'effettiva esistenza delle relazioni individuate – nonché per approfondire la comprensione delle relazioni tra i diversi gruppi – è stato eseguito un confronto multiplo a coppie (*multiple pairwise comparison*) all'interno dei gruppi di variabili indipendenti, utilizzando il test post-hoc di Games-Howell.

Tale scelta si è resa necessaria poiché i dati violavano sia l'assunzione di omoschedasticità che quella di normalità: trattandosi di un test non-parametrico, quest'ultimo assunto non è richiesto, e può essere applicato anche nel caso di varianze e numerosità differenti nei campioni analizzati.

Pertanto, utilizzando la funzione “*games_howell_test*” del pacchetto “*rstatix*”, sono stati ottenuti i seguenti risultati:

- 1) Per quanto riguarda i confronti all'interno del gruppo “Genere”, le seguenti variabili sono state associate ad un p-value significativo: DA1, DA2, DA3, DA4, EDMQ5, SDL1;
- 2) Per il gruppo “Età”: DA6, DA8, DER4, EDMQ1, EDMQ2, EDMQ3, EDMQ4, SDL2;
- 3) Per il gruppo “Nazione”: DA1, DA2, DA6, DA7, DA9, DER2, DER3, DTL2, DTL5, EDMQ2, EDMQ3, EDMQ4, EDMQ5, EDMQ6, SDL1, SDL5, SED2;
- 4) Per il gruppo “Area disciplinare”: DA6, DER1, EDMQ4, EDMQ5, SDL1, SDL3, SDL4, SDL5, SED1;
- 5) Per la variabile “Esperienza d'insegnamento”: DA7, EDMQ1, EDMQ2, EDMQ3, EDMQ4;
- 6) Per la variabile “Esperienza di ricerca”: DA6, DA7, DER2, DER3, EDMQ3, EDMQ4, EDMQ5.

All'interno delle tabelle vengono specificate le statistiche di maggior rilevanza per ogni variabile significativa, e la coppia ad essa associata.

1. Confronti all'interno del gruppo “Genere”

Variabile	Gruppo 1	Gruppo 2	Media delle differenze	Intervallo di confidenza per la media		P
				Lim. Inf.	Lim Sup.	
DA1	Uomini	Donne	-0,36	-0,66	-0,07	0,016
DA2	Uomini	Donne	-0,36	-0,65	-0,06	0,017
DA3	Uomini	Donne	-0,36	-0,66	-0,069	0,016
DA4	Uomini	Donne	-0,35	-0,64	-0,06	0,020
EDMQ5	Uomini	Donne	-0,44	-0,75	-0,13	0,006
SDL1	Uomini	Donne	-0,35	-0,66	-0,04	0,027

Tabella 30: Comparazione multipla a coppie per gruppo “Genere”

I risultati del test per il gruppo “Genere” confermano quanto già osservato con le precedenti analisi della varianza; le variabili con *p-value* statisticamente significativo sono, infatti, le stesse dell'ANOVA e del MANOVA.

Pertanto, trova ulteriore riscontro il dato – già osservato in precedenza – della maggiore frequenza di comportamenti di *data literacy* all'interno del gruppo degli uomini, con particolare riferimento all'uso dei dati per la valutazione dell'apprendimento (DA1), per il monitoraggio didattico (DA2), per dare feedback sia sommativi (DA3), che formativi (DA4), l'uso delle analitiche di apprendimento ai fini della progettazione curricolare (EDMQ5) e, infine, per promuovere le capacità di elaborare *data narratives* (SDL2).

Tuttavia, sono da considerare due aspetti che emergono dall'analisi:

- Il *range* della media delle differenze è molto contenuto, variando tra -0,35 e -0,44; la differenza osservata tra i gruppi è, quindi, limitata;
- I valori di significatività statistica – benché ben al di sotto della soglia, fissata a 0,05 – si collocano al di sopra dell'1% (con l'eccezione della variabile EDMQ5, la quale ha un valore di *p* uguale a 0,006), mostrando quindi una bassa significatività.

Queste evidenze portano a confermare l'idea che la variabile “Genere” abbia importanza limitata rispetto all'alfabetizzazione ai dati.

2. Confronti all'interno del gruppo “Età”

Variabile	Gruppo 1	Gruppo 2	Media delle differenze	Intervallo di confidenza per la media		P
				Lim. Inf.	Lim Sup.	
DA6	45-54	Più di 55	-0,52	-1,03	-0,015	0,041
DA6	25-34	Più di 55	-1,07	-1,85	-0,29	0,003
DA8	45-54	Più di 55	-0,48	-0,96	-0,002	0,049
DER4	35-44	Più di 55	0,65	0,13	1,16	0,007
EDMQ1	35-44	Più di 55	0,54	0,06	1,02	0,021
EDMQ2	35-44	Più di 55	0,75	0,25	1,24	0,0007
EDMQ3	35-44	Più di 55	0,60	0,08	1,12	0,018
EDMQ3	45-54	Più di 55	0,52	0,028	1,01	0,034
EDMQ4	35-44	Più di 55	0,84	0,30	1,38	0,0004
EDMQ4	45-54	Più di 55	0,60	0,075	1,12	0,018
SDL2	25-34	Più di 55	-0,90	-1,71	-0,09	0,023

Tabella 31: Comparazione multipla a coppie per gruppo “Età”

La prima considerazione riguardo a quanto emerso dai confronti multipli a coppie è relativa al numero di variabili che raggiungono, con questo test, il livello di significatività: in controtendenza

rispetto a quanto già osservato, infatti, le variabili DER1, DER2, e DER3, non raggiungono la soglia; tra i gruppi non sembrano esserci differenze rilevanti in merito all'uso dei dati come risorsa educativa, eccetto per il loro impiego nel promuovere una riflessione negli studenti sugli aspetti etici dell'insegnamento con dati aperti (DER4).

Queste variabili, pertanto, potrebbero aver raggiunto in precedenza un livello di significatività statistica a causa di alcuni outlier rimasti all'interno del dataset, nonostante esso sia stato "pulito", oppure a causa di una distribuzione dei punteggi che si discostava eccessivamente dalla gaussiana.

I confronti all'interno del gruppo "Età" hanno, per il resto, messo in luce un quadro piuttosto articolato: in primo luogo, per la classe d'età "45-54", rispetto a quella "over-55", si osserva una maggiore frequenza (-0,52) nell'uso dei dati provenienti da *learning analytics* per l'autovalutazione della didattica (DA6), e nell'utilizzo di sistemi automatizzati per la valutazione dell'apprendimento (DA8, -0,48); relazione inversa si nota, invece, nell'uso di dati provenienti dalla valutazione del proprio corso per lo sviluppo istituzionale e la pianificazione (EDMQ3, +0,52), e uso di dati del proprio corso per la progettazione curricolare (EDMQ4, +0,6).

Per quanto concerne la classe d'età "35-44" – sempre confrontata con la "over-55" – si può constatare come quest'ultima faccia un impiego molto più massiccio di pratiche basate sui dati rispetto alla prima; in tutti i confronti, infatti, la media delle differenze denota una minor frequenza di *data practices* nel gruppo più giovane e, nello specifico, nel promuovere la riflessione etica negli studenti sull'uso dei dati aperti nella didattica (DER4, +0,65), nell'uso di dati provenienti da report nazionali per lo sviluppo istituzionale e la pianificazione (EDMQ1, +0,54), nell'uso di dati da report istituzionali per lo sviluppo istituzionale e la pianificazione (EDMQ2, +0,75), nell'uso di dati provenienti dalla valutazione del proprio corso per lo sviluppo istituzionale e la pianificazione (EDMQ3, +0,6) e, infine, nell'uso di dati del proprio corso per la progettazione curricolare (EDMQ4, +0,84).

L'ultimo confronto ha riguardato le classi "25-34" e "over-55": in questo caso, le uniche variabili ad aver raggiunto la significatività statistica sono state DA6 (Uso di dati da *learning analytics* per il monitoraggio/valutazione della didattica), e SDL2 (Promozione delle abilità di narrazione con e attraverso i dati); queste hanno messo in luce una maggiore propensione del gruppo più giovane ad usare le analitiche di apprendimento (-1,07), e a supportare lo sviluppo delle capacità degli studenti di elaborare narrazioni con, e attraverso, i dati (-0,9).

Anche in questo caso, si tratta di fare alcune precisazioni:

- Il *p-value* della differenza delle medie, tra "45-54" e "Più di 55", varia tra un minimo di 0,02 (EDMQ3) ed un massimo di 0,05 (DA8) – in quest'ultimo caso, la variabile DA8

mostra un valore di probabilità addirittura pari alla soglia di significatività: questo porta a ritenere che non vi siano discrepanze rilevanti tra i gruppi in termini di *data literacy*;

- I confronti tra le classi d'età “35-44” e “over-55” hanno generalmente raggiunto punteggi di significatività statistica abbastanza elevati, con l'eccezione degli item EDMQ1 ed EDMQ3 – entrambi con $p = 0,02$. Ciò suggerisce che vi siano delle differenze sufficientemente rilevanti tra i due gruppi nelle pratiche basate sui dati, come viene mostrato anche dai valori delle medie delle differenze, le quali variano tra 0,54 e 0,84;
- Infine, per quanto concerne il confronto tra “25-34” e “over-55”, solo le due variabili DA6 e SDL2 raggiungono la significatività, sebbene con valori nettamente diversi (0,003 nel primo caso, e 0,023 nel secondo). Ciò potrebbe suggerire una discrepanza non molto marcata nell'uso dei dati per promuovere le capacità di elaborazione di *data narratives* negli studenti, mentre appare rilevante nell'uso di analitiche di apprendimento per il monitoraggio didattico e dell'apprendimento.

Sulla base di queste considerazioni, si può concludere che possano sussistere discrepanze degne di nota in termini di *data practices* tra la classe d'età “35-44” e “Più di 55”; che queste siano, al contrario, ridotte, tra i gruppi “45-54” e “Più di 55”; infine, sembra che solo la variabile SDL2 possa essere annoverata tra quelle significative del confronto tra “25-34” e “Più di 55”.

L'impatto della variabile “Età” sulla *data literacy* appare, anche in questo caso, modesto.

3. Confronti all'interno del gruppo “Nazione”

Variabile	Gruppo 1	Gruppo 2	Media delle differenze	Intervallo di confidenza per la media		P
				Lim. Inf.	Lim Sup.	
DA1	Italia	Spagna	-0,47	-0,77	-0,17	0,002
DA2	Italia	Spagna	-0,59	-0,89	-0,30	0,000108
DA6	Italia	Spagna	1,05	0,76	1,35	0
DA7	Italia	Spagna	0,91	0,65	1,18	9,8 ⁻¹¹
DA9	Italia	Spagna	0,52	0,28	0,75	2,46 ⁻⁵
DER2	Italia	Spagna	-0,36	-0,66	-0,05	0,021
DER3	Italia	Spagna	-0,41	-0,71	-0,11	0,008
DTL2	Italia	Spagna	0,33	0,09	0,57	0,006
DTL5	Italia	Spagna	-0,32	-0,62	-0,02	0,035
EDMQ2	Italia	Spagna	-0,88	-1,16	-0,60	2,16 ⁻⁹
EDMQ3	Italia	Spagna	-1,03	-1,32	-0,73	3,05 ⁻¹¹
EDMQ4	Italia	Spagna	-1,81	-2,08	-1,53	0
EDMQ5	Italia	Spagna	-1,67	-1,95	-1,40	0
EDMQ6	Italia	Spagna	0,92	0,62	1,22	4,08 ⁻⁹
SDL1	Italia	Spagna	-0,79	-1,10	-0,47	1,29 ⁻⁶
SDL5	Italia	Spagna	-0,29	-0,57	-0,0038	0,047

SED2	Italia	Spagna	-0,63	-0,92	-0,35	1,76 ⁻⁵
-------------	--------	--------	-------	-------	-------	--------------------

Tabella 32: Comparazione multipla a coppie per gruppo "Nazione"

Un primo dato preliminare rispetto ai confronti a coppie all'interno della categoria "Nazione" mostra che, nell'insieme, viene confermato quanto già constatato in precedenza grazie alle analisi della varianza.

Per quanto concerne il contesto italiano, quindi, si osserva una media delle differenze maggiore rispetto al caso spagnolo per quanto concerne l'uso di dati per la valutazione dell'apprendimento (DA1) e il monitoraggio didattico (DA2), l'uso di dati aperti come risorsa educativa, provenienti dalla ricerca o dalla pubblica amministrazione (DER2), oppure dalla propria attività di ricerca (DER3), all'uso dei dati per implementare *dashboards* ai fini della promozione delle abilità autoregolate degli studenti (DTL5), per la gestione e la qualità della didattica, impiegando report istituzionali (EDMQ2), dati provenienti dalla valutazione del proprio corso (EDMQ3), nell'utilizzare dati del proprio corso per la progettazione curricolare (EDMQ4), nell'impiego di analitiche d'apprendimento per la progettazione didattica (EDMQ5), per lo sviluppo dell'alfabetizzazione ai dati degli studenti, nei termini della valutazione critica dell'affidabilità e credibilità di una fonte informativa (SDL1), e della riflessione rispetto all'impatto dei dati nella società (SDL5) e, infine, nella condivisione di dati provenienti dalle piattaforme d'apprendimento per supportare gli studenti nella riflessione riguardo al loro apprendimento nel contesto istituzionale e sociale (SED2).

Nel contesto spagnolo, invece, risulta più frequente il ricorso ai dati ai fini dell'*assessment*, in particolare impiegando dati da *learning analytics* per il monitoraggio e la valutazione della didattica (DA6), nell'uso di *dashboards* per informare la propria didattica (DA7), nell'impiego dei dati ottenuti dalle piattaforme d'apprendimento per riflettere, assieme agli studenti, sulla qualità della didattica (DA9), nell'integrazione di tecniche massive di *data extraction* nei compiti assegnati agli studenti (DTL2) e, per concludere, nell'impiegare le analitiche d'apprendimento per indagare l'efficacia della propria didattica (EDMQ6).

In relazione a quanto emerso, si può constatare che:

- I livelli di significatività delle variabili sono, generalmente, molto prossimi allo zero, eccetto che per DER2 ($p = 0,02$), DTL5 ($p = 0,035$), e SDL5 ($p = 0,047$), che si rivelano scarsamente significative;
- Le differenze rispetto alle competenze di *data literacy* appaiono marcate: escludendo gli item meno significativi, il *range* della media delle differenze tra contesto italiano e spagnolo varia tra -0,47 e -1,81; il caso spagnolo, invece, si distingue per un *range* inferiore, variando quest'ultimo tra 0,33 e 1,05. I docenti appartenenti all'università

italiana, pertanto, mostrano una frequenza maggiore nell'attuazione delle pratiche di *data literacy* rispetto ai loro omologhi dell'università iberica, e ciò accade per la maggior parte degli item discussi (12 vs. 5); in quelli restanti, le diversità appaiono meno rilevanti.

Nel complesso, le analisi sembrano confermare l'elevata incidenza della variabile "Nazione" sull'alfabetizzazione ai dati dei docenti universitari.

4. Confronti all'interno del gruppo "Area disciplinare"

Variabile	Gruppo 1	Gruppo 2	Media delle differenze	Intervallo di confidenza per la media		P
				Lim. Inf.	Lim. Sup.	
DA6	Scienze Naturali e Formali	Scienze sociali	0,76	0,06	1,46	0,026
DER1	Discipline Umanistiche & Linguistica	Scienze sociali	0,74	0,10	1,38	0,015
EDMQ4	Scienze Naturali e Formali	Scienze tecnologiche	-0,98	-1,75	-0,22	0,005
EDMQ4	Scienze Naturali e Formali	Scienze sociali	-1,19	-1,84	-0,54	1,6 ⁻⁵
EDMQ5	Scienze Naturali e Formali	Scienze sociali	-0,79	-1,51	-0,07	0,025
SDL1	Scienze della vita	Scienze tecnologiche	-0,80	-1,47	-0,13	0,011
SDL3	Scienze tecnologiche	Scienze sociali	0,56	0,10	1,01	0,008
SDL4	Scienze tecnologiche	Scienze sociali	0,57	0,09	1,05	0,01
SDL5	Scienze della vita	Scienze tecnologiche	-0,63	-1,20	-0,06	0,023
SDL5	Scienze tecnologiche	Scienze sociali	0,68	0,16	1,19	0,004
SED1	Discipline Umanistiche e Linguistica	Scienze sociali	0,62	0,03	1,21	0,033
SED1	Scienze Naturali e Formali	Scienze sociali	0,71	0,18	1,23	0,003
SED1	Scienze tecnologiche	Scienze sociali	0,55	0,012	1,10	0,043

Tabella 33: Comparazione multipla a coppie per gruppo "Area disciplinare"

Dopo aver eseguito i confronti all'interno della categoria "Area disciplinare", è stato osservato che – diversamente dalle analisi della varianza – le variabili DA8 e DER2 non hanno superato la soglia di significatività; anche in questo caso, il risultato ottenuto in precedenza potrebbe essere stato distorto da *outliers* residui, o da una distribuzione che si discostava eccessivamente dalla normale.

Addentrando nell'analisi, emerge che i docenti appartenenti all'area "Scienze Naturali e Formali" mostrano maggiore propensione nell'uso dei dati ricavati dal proprio corso ai fini della progettazione curricolare (EDMQ4) rispetto ai colleghi di "Scienze tecnologiche" (-0,98) e "Scienze Sociali" (-1,19), così come nell'uso di analitiche di apprendimento per la progettazione didattica (EDMQ5) rispetto all'area delle scienze sociali (-0,79).

All'interno di quest'ultima categoria, invece, i docenti tendono ad usare con maggiore frequenza le informazioni ricavate dalle analitiche d'apprendimento per il monitoraggio e valutazione dell'insegnamento (DA6, +0,76), nonché nella pratica di condividere report istituzionali ai fini della riflessione degli studenti sul loro apprendimento, all'interno del contesto istituzionale e sociale (SED1, +0,71), rispetto ai colleghi appartenenti a “Scienze Naturali e Formali”.

All'interno dell'area delle scienze sociali, inoltre, la media delle differenze è superiore a quella di “Discipline Umanistiche e Linguistica” rispetto all'impiego di report nazionali e internazionali come risorsa didattica (DER1, +0,74), così come per l'item SED1 (+0,62).

Una maggiore competenza in *data literacy* dei docenti di scienze sociali sembra emergere anche rispetto al gruppo delle “Scienze tecnologiche”, in particolare per quanto concerne la promozione della riflessione negli studenti in relazione ai vincoli sociali all'uso dei dati e alle IA (SDL3, +0,56), alle implicazioni etiche nella raccolta ed uso dei dati nell'apprendimento (SDL4, +0,57), e all'impatto dei dati sulla società in generale (SDL5, +0,68); anche nell'item SED1, l'area delle “Scienze sociali” ha media delle differenze maggiore rispetto a “Scienze tecnologiche” (+0,55).

Soffermandosi brevemente su queste ultime osservazioni, si può notare come l'impiego dei dati ai fini di una riflessione critica degli studenti, in relazione al proprio apprendimento, all'etica dei dati, nonché alle implicazioni dell'innovazione digitale, risulta più carente proprio nell'area disciplinare maggiormente “tecnica” e che, verosimilmente, contribuirà allo sviluppo e alla successiva integrazione delle future tecnologie digitali. Anche nel campo delle “Scienze della vita” si può constatare una maggiore abitudine alla riflessione e discussione critica assieme agli studenti (SDL1, -0,8, e SDL5, -0,63), rispetto all'area tecnologica.

In sintesi:

- I docenti di “Scienze sociali” mostrano maggior abitudine alle pratiche – in special modo critiche e riflessive – di *data literacy*; tuttavia, rispetto a loro il gruppo di “Scienze Naturali e Formali” appare più propenso nell'uso dei dati per la gestione e la qualità della didattica;
- All'interno dell'area tecnologica si riscontra una tendenza minore, rispetto ai gruppi di “Scienze sociali” e “Scienze della vita”, a mettere in atto pratiche volte a stimolare il dialogo e la riflessione critica degli studenti in relazione alle implicazioni della tecnologia stessa;
- Nell'insieme, le variabili hanno valori di significatività non elevati; nella fattispecie, gli item DA6 (Scienze Naturali e Formali vs. Scienze Sociali), DER1 (Discipline Umanistiche e Linguistica vs Scienze sociali), EDMQ5 (Scienze Naturali e Formali vs. Scienze Sociali), SDL1 ed SDL5 (Scienze della vita vs. Scienze tecnologiche), e SED1

(Scienze tecnologiche vs. Scienze sociali), hanno un *p-value* compreso tra 0,1 e 0,5, mostrandosi poco significative;

- Infine, con riferimento alle variabili maggiormente significative, le medie delle differenze appaiono consistenti per EDMQ4 nel confronto tra Scienze Naturali e Formali e Scienze tecnologiche (-0,98), così come nella comparazione con Scienze sociali (-1,19); per SDL3 e SDL5, nel raffronto tra Scienze tecnologiche e Scienze sociali (rispettivamente 0,56 e 0,68), e infine tra Scienze Naturali e Formali e Scienze sociali (-0,71) per SED1.

In conclusione, i valori di probabilità, e il ristretto numero di comparazioni significative, mettono ulteriormente in luce una limitata influenza della variabile “Area disciplinare” sulle competenze digitali dei docenti.

5. Confronti all'interno del gruppo “Esperienza d'insegnamento”

Variabile	Gruppo 1	Gruppo 2	Media delle differenze	Intervallo di confidenza per la media		P
				Lim. Inf.	Lim Sup.	
DA7	Meno di 3	5-10	-0,89	-1,67	-0,11	0,017
DA7	Meno di 3	Più di 15	-0,80	-1,50	-0,11	0,015
EDMQ1	Meno di 3	Più di 15	0,65	0,06	1,24	0,022
EDMQ2	Meno di 3	Più di 15	0,68	0,07	1,29	0,022
EDMQ2	5-10	Più di 15	0,67	0,11	1,22	0,01
EDMQ3	Meno di 3	Più di 15	0,78	0,09	1,47	0,018
EDMQ4	Meno di 3	10-15	0,80	0,03	1,58	0,04
EDMQ4	Meno di 3	Più di 15	0,95	0,26	1,63	0,002
EDMQ4	3-4	Più di 15	0,78	0,05	1,52	0,031

Tabella 34: Comparazione multipla a coppie per gruppo “Esperienza d'insegnamento”

In questo caso, i risultati delle analisi della varianza vengono confermati dal test a coppie appaiate; le variabili significative sono le stesse tra le diverse analisi.

All'interno del gruppo “Esperienza d'insegnamento” sembra delinearci un effetto dell'anzianità di servizio almeno per quanto concerne le pratiche di uso dei dati per la gestione e la qualità della didattica, impiegando report nazionali (EDMQ1), report istituzionali (EDMQ2), informazioni ricavate dalla valutazione del proprio corso (EDMQ3) e, più in generale, dati provenienti dal proprio corso (EDMQ4).

Il gruppo di docenti con più di 15 anni d'esperienza, infatti, supera sistematicamente i meno esperti (con meno di 3 anni d'esperienza) in ognuno di questi item; inoltre, il gruppo con maggior esperienza mette in atto le pratiche di pianificazione istituzionale tramite report istituzionali (EDMQ2) in misura maggiore rispetto al gruppo “5-10” (+0,67), nonché impiega più spesso dati

provenienti dal proprio corso per la progettazione curricolare (EDMQ4) in confronto alla categoria “3-4” (+0,78). Riguardo quest’ultima variabile, la classe “10-15” sembra mettere in atto queste pratiche più spesso rispetto ai meno esperti (+0,95).

Infine, si può notare come il gruppo con meno esperienza (“Meno di 3”) sembri utilizzare, con maggior propensione rispetto sia alla classe “5-10” che alla “Più di 15”, le informazioni ottenute dalle *dashboards* d’apprendimento per informare la propria didattica, con valori rispettivamente di -0,89 e -0,80 (DA7).

Ulteriori osservazioni rispetto all’analisi:

- I valori di significatività statistica sono – con l’eccezione di EDMQ4, nel confronto “Meno di 3 vs. Più di 15” ($p\text{-value} = 0,002$) – sempre compresi tra 0,1 e 0,5, evidenziano un effetto debole dell’anzianità di servizio sulle pratiche di *data literacy*;
- Rispetto alla dimensione delle medie delle differenze, esse sembrano piuttosto elevate, collocandosi in un *range* compreso tra -0,89 e +0,95: le discrepanze osservate appaiono, quindi, rilevanti.

Nell’insieme, l’impatto dell’esperienza d’insegnamento sulle pratiche basate sui dati – considerando i bassi livelli dei $p\text{-value}$, e l’esiguità dei confronti che raggiungono la significatività – appare molto limitato, se non potenzialmente nullo.

6. Confronti all’interno del gruppo “Esperienza di ricerca”

Variabile	Gruppo 1	Gruppo 2	Media delle differenze	Intervallo di confidenza per la media		P
				Lim. Inf.	Lim Sup.	
DA6	Nessuna esperienza	Più di 15	-0,96	-1,62	-0,31	0,000571
DA6	3-5	Più di 15	-0,97	-1,81	-0,12	0,016
DA7	Nessuna esperienza	Più di 15	-0,98	-1,67	-0,30	0,000895
DA7	Meno di 3	Più di 15	-0,97	-1,75	-0,19	0,007
DA7	3-5	Più di 15	-1,01	-1,89	-0,13	0,016
DER2	Nessuna esperienza	11-15	0,92	0,08	1,76	0,024
DER2	Nessuna esperienza	Più di 15	0,73	0,09	1,36	0,015
DER3	Nessuna esperienza	3-5	0,97	0,05	1,89	0,032
DER3	Nessuna esperienza	6-10	1,15	0,35	1,95	0,000835
DER3	Nessuna esperienza	11-15	0,86	0,03	1,68	0,037
DER3	Nessuna esperienza	Più di 15	0,86	0,21	1,51	0,003
EDMQ2	Nessuna esperienza	Più di 15	0,90	0,31	1,49	0,000297
EDMQ3	Nessuna esperienza	Più di 15	0,82	0,11	1,52	0,013
EDMQ4	Nessuna esperienza	6-10	0,84	0,03	1,65	0,038
EDMQ4	Nessuna esperienza	11-15	1,14	0,35	1,93	0,000817
EDMQ4	Nessuna esperienza	Più di 15	1,56	0,92	2,20	1,03 ⁻⁹

EDMQ4	3-5	Più di 15	0,89	0,04	1,75	0,037
EDMQ4	6-10	Più di 15	0,72	0,01	1,43	0,043
EDMQ5	Nessuna esperienza	11-15	1,42	0,60	2,24	2,75 ⁻⁵
EDMQ5	Nessuna esperienza	Più di 15	1,26	0,60	1,93	2,84 ⁻⁶

Tabella 35: Comparazione multipla a coppie per gruppo "Esperienza di ricerca"

La prima considerazione rilevante riguarda l'assenza delle variabili DA9 ed EDMQ2 dai confronti significativi; pertanto, come osservato in precedenza, potrebbe essersi trattato di risultati, ottenuti dalle analisi della varianza, distorti da valori anomali o da distribuzioni troppo lontane dalla curva normale standard.

Per quanto concerne l'analisi del test a coppie appaiate, si può constatare come la categoria con maggiore esperienza ("Più di 15"), rispetto ai gruppi con "Nessuna esperienza" e "3-5", mostri una minor propensione nell'impiego dei dati da piattaforme d'apprendimento per monitorare/valutare la didattica (DA6), con i punteggi rispettivamente di -0,96 e -0,97; questo gruppo, inoltre, sembra fare meno uso dei dati da *dashboards* per la propria didattica (DA7) in confronto alle categorie "Nessuna esperienza", "Meno di 3", e "3-5", con valori, nell'ordine, di -0,98, -0,97, e -1,01.

Il gruppo con la maggior esperienza di ricerca, invece, sembra avere una maggiore propensione nell'utilizzo di dati aperti come risorsa didattica, provenienti dalla ricerca o dalla pubblica amministrazione (DER2), o dalla propria attività di ricerca (DER3), rispetto al gruppo senza esperienza, con +0,73 e +0,86.

La classe più esperta, inoltre, sembra fare maggior uso dei dati ai fini della gestione istituzionale e della qualità della didattica, impiegando dati da report istituzionali (EDMQ2), informazioni ricavate dalla valutazione del proprio corso (EDMQ3), dati provenienti dal proprio corso (EDMQ4) e utilizzando dati da *learning analytics* (EDMQ5), in rapporto al gruppo senza esperienza; le medie delle differenze sono, nell'ordine, di +0,9, +0,82, +1,56, e +1,26. Il gruppo con la maggior esperienza, inoltre, sembra fare uso di dati dal proprio corso ai fini della progettazione curricolare (EDMQ4) con maggior frequenza sia rispetto alla classe "3-5", che "6-10" (+0,84 e +1,14).

Il gruppo "Senza esperienza" appare debole anche nel confronto con la classe "11-15" (+0,92) nell'uso dei dati, provenienti dalla ricerca o dalla P.A., come risorsa educativa (DER2), così come nell'impiego di dati propri (DER3): le classi "3-5", "6-10", e "11-15", in questo senso, sembrano farne maggior uso (medie delle differenze rispettivamente di +0,97, +1,15, e +0,86).

Per quanto concerne l'uso di dati del proprio corso per la progettazione curricolare (EDMQ4), le classi "6-10" e "11-15" sembrano farne maggior ricorso rispetto al gruppo "Senza esperienza" (+0,84 e +1,14); invece, nel confronto tra la categoria "11-15" e "Senza esperienza", sull'impiego di informazioni provenienti dalle piattaforme d'apprendimento per la progettazione didattica (EDMQ5), i primi mostrano utilizzarli in maniera più massiccia (+1,42).

Le considerazioni conclusive portano a ritenere che:

- L'effetto generale della variabile "Esperienza di ricerca" sulla *data literacy* sia maggiore rispetto a quello dell'esperienza d'insegnamento; oltre alla presenza di un numero maggiore di coppie significative, infatti, si può constatare che numerosi item abbiano raggiunto un'elevata significatività statistica.
- Le discrepanze nell'uso dei dati, sia come risorsa educativa che come fonte informativa ai fini della gestione e la qualità della didattica, tra i gruppi più esperti e quello con meno esperienza di ricerca, potrebbero non essere così marcate: i *p-value* in molti di questi confronti sono compresi tra 0,1 e 0,5, dimostrandosi poco significativi. Discorso analogo può essere fatto per le coppie formate dal gruppo con la maggiore esperienza e nessuna esperienza, in relazione all'item EDMQ3 ($p = 0,13$), e dalle coppie tra "Più di 15" e "3-5", e "6-10", in rapporto all'item EDMQ4 (*p-value* uguale, rispettivamente, a 0,037 e 0,043);
- Le differenze in termini di pratiche di *data literacy*, in rapporto agli anni di esperienza di ricerca, appaiono rilevanti: i docenti senza esperienza di ricerca sembrano più propensi degli altri ad utilizzare le informazioni ricavate dalle piattaforme a fini di *assessment*, mentre al crescere dell'esperienza si osserva una maggiore abitudine all'uso di dati con finalità di *management* istituzionale e della qualità della didattica.

Queste ultime riflessioni portano a ritenere l'effetto della variabile "Esperienza d'insegnamento" come maggiormente rilevante, ai fini delle pratiche di *data literacy*, rispetto all'esperienza d'insegnamento; tale impatto risulta, comunque, limitato a casi specifici.

4.4. Analisi Fattoriale

Una volta identificate ed esaminate le possibili relazioni tra variabili dipendenti e indipendenti, si è proseguito con l'esecuzione delle analisi necessarie per individuare i fattori alla base delle risposte date al questionario.

È stata quindi eseguita un "Analisi Fattoriale Esplorativa" (AFE), per ottenere le indicazioni sul numero sufficiente di costrutti per elaborare un modello statisticamente valido e teoricamente solido dei fabbisogni professionali e formativi dei docenti universitari, nonché – attraverso una "Analisi Fattoriale Confermativa" (AFC) – convalidare, se possibile, la fondatezza delle scale del questionario.

I vantaggi offerti dall'AFE, rispetto ad altri metodi di analisi dei dati (con particolare riferimento all'Analisi delle Componenti Principali), sono:

- 1) Permette di analizzare solo la varianza comune alle variabili;
- 2) Può produrre valori di comunalità stimate superiori a 1, con la conseguenza che la varianza spiegata è uguale o maggiore al 100% (*Heywood Case*): ciò è utile a fini diagnostici, in quanto tale anomalia è indice di un modello con cattivo adattamento ai dati, o forti violazioni degli assunti alla base dell'AFE. Il ricercatore dispone pertanto di un indicatore che permette, eventualmente, di ricorrere ad altre tecniche di analisi;
- 3) Poiché tiene conto degli errori casuali, essa rappresenta in maniera più realistica e fedele la struttura delle correlazioni;
- 4) Utilizzando il metodo della Massima Verosimiglianza ("*Maximum Likelihood*"), dà modo al ricercatore di abbandonare un modello con scarsa solidità e validità (Giannini e Pannocchia, 2006, pp. 33-36).

Tale tecnica è diversa dall'Analisi Fattoriale Confermativa (AFC), la quale rappresenta invece il necessario completamento di quella Esplorativa: se, da un lato, l'AFE permette di individuare il numero di fattori partendo dai dati (approccio *data-driven*), l'AFC dà modo di testare il modello che emerge da tale analisi, riducendo l'incertezza legata al numero possibile di modelli concorrenti da testare (*ibidem*, p. 36).

4.4.1. Analisi Fattoriale Esplorativa (AFE) e Confermativa (AFC)

Come già accennato, uno dei motivi per compiere un'Analisi Esplorativa è quella di scoprire i costrutti alla base delle correlazioni osservate all'interno di un insieme di variabili: più nello

specifico, si tratta di un insieme di metodi statistici volti a “*studiare, riassumere e semplificare le relazioni in gruppo di variabili*” (Giannini e Pannocchia, 2006, p. 13), con lo scopo di “*ridurre l’informazione contenuta in una serie di dati, individuando uno o più fattori o dimensioni latenti in grado di rendere conto di ciò che accomuna un insieme di variabili*” (ibidem).

In ragione di ciò, pochi fattori, anche se non direttamente osservabili, possono essere utilizzati per elaborare un modello esplicativo di un determinato fenomeno; in questo caso, i comportamenti legati alla competenza di *data literacy*.

Sempre Giannini e Pannocchia (ivi), propongono un riepilogo delle fasi richieste dall’AFE:

- 1) Selezione delle variabili,
- 2) Calcolo della matrice delle correlazioni tra le variabili,
- 3) Estrazione dei fattori,
- 4) Rotazione dei fattori,
- 5) Interpretazione della matrice dei fattori ruotati.

Di seguito, si discutono i diversi passaggi in relazione alla presente analisi.

1. Selezione delle variabili

Per quanto concerne la selezione delle variabili, esse devono rispettare in primo luogo due parametri (ivi):

- è necessario che *sovra-determinino* i fattori di almeno 4 o 5 volte, e che
- tali fattori posseggano *comunalità* (quantità di varianza spiegata dai fattori comuni) elevate, ed in ogni caso mai inferiori a 0.30 (ivi).

Rispettando tali parametri, è lecito utilizzare anche campioni piccoli di cento soggetti (MacCallum et al. (1999), citato da Giannini e Pannocchia (2006), p. 27); qualora tali assunti siano rispettati solo in parte, sono necessari campioni di almeno 200 soggetti, mentre in caso di forti violazioni anche campioni con elevata numerosità ($N > 400$) possono produrre un’AFE non valida.

Nella situazione presa in esame, i 6 fattori (scale del questionario) corrispondevano a 35 variabili, superando la soglia minima di 1:4 e raggiungendo un valore (quasi) uguale a 1:6; l’ampiezza del dataset oggetto dell’analisi superava, inoltre, quello minimo richiesto di almeno due volte ($N = 494$), garantendo il rispetto dei requisiti di sovra-determinazione e numerosità campionaria.

Oltre a queste considerazioni, è stato eseguito il Test di Adeguatezza Campionaria (Test di Kaiser-Meyer-Olkin) per ottenere un indice quantitativo dell’idoneità del campione all’AFE:

pertanto, utilizzando la funzione “*KMO*” del pacchetto “*psych*”, è stato ottenuto una misura di adeguatezza campionaria generale pari a 0,95. Al di sopra dello 0,9, la conformità del campione all’analisi fattoriale viene considerata eccellente: per i singoli item, inoltre, si è ricavata una SMA (*Sample Measure Adequacy*) che variava da un minimo di 0,91 ad un massimo di 0,97.

Analizzando le comunalità si è trovato quanto segue (Tabella 36):

Variabile	Comunalità
EDMQ1	0.58310156
EDMQ2	0.64230799
EDMQ3	0.66586578
EDMQ4	0.75904692
EDMQ5	0.77557355
EDMQ6	0.49536862
EDMQ7	0.76052675
EDMQ8	0.73089353
DER1	0.68653422
DER2	0.65856477
DER3	0.61670468
DER4	0.56240293
SED1	0.78314821
SED2	0.76937405
SED3	0.81821644
DTL1	0.51137703
DTL2	0.63886913
DTL3	0.69152821
DTL4	0.69886777
DTL5	0.63382587
DA1	0.83791255
DA2	0.86690497
DA3	0.83522313
DA4	0.77819476
DA5	0.5754563
DA6	0.65072591
DA7	0.58369696
DA8	0.58742731
DA9	0.76641294
DA10	0.55453373
SDL1	0.59178927
SDL2	0.62917361
SDL3	0.74792312
SDL4	0.90993413
SDL5	0.86333954

Tabella 36: Comunalità calcolate

Come si evince dai risultati qui esposti, nessuna comunalità trovata è risultata inferiore al valore-soglia di 0,30: il secondo requisito è stato, pertanto, rispettato.

2. Matrice delle correlazioni

Calcolata la matrice delle correlazioni – compito eseguibile con una certa semplicità attraverso strumenti informatici quali software statistici o fogli di calcolo – è necessario eseguire dei test per controllare che essa sia fattorizzabile: è stato eseguito (funzioni “*cortest.bartlett*” del pacchetto “*psych*”) il Test di Sfericità di Bartlett, il quale ha restituito un valore di Chi-quadrato pari a 15337

per 595 gradi di libertà, con un *p-value* significativo ($P = 5,9^{-14}$). Come confermato dal test, la matrice delle correlazioni è idonea alla fattorizzazione.

3. Estrazione dei fattori

Compiuto il passaggio precedente, si deve scegliere quale metodo di estrazione dei fattori utilizzare. I principali (Giannini e Pannocchia, 2006, pp. 39-42) sono:

- Il “Metodo dei Fattori Principali” (AFP),
- Il “Metodo dei Minimi Quadrati” (MQ),
- Infine, il “Metodo della Massima Verosimiglianza” (ML).

Ognuna di queste tecniche possiede vantaggi e svantaggi. Il Metodo dei Fattori Principali, in primo luogo, non presuppone l'assunzione di normalità multivariata, con il limite, tuttavia, di richiedere il calcolo delle stime di comunalità; per tale problema non esiste una soluzione univoca, e tale aspetto ha la conseguenza fondamentale di inserire un certo grado di arbitrarietà ed indeterminatezza nella soluzione trovata.

Per ovviare a tali limiti è stato proposto il Metodo dei Minimi Quadrati il quale, non richiedendo stime di comunalità, risulta più affidabile, ma offre anche risultati diversi e non sovrapponibili a quelli ottenuti con il Metodo dei Fattori Principali: impiegando solo parte della matrice delle correlazioni, ovvero quella al di fuori della diagonale matriciale, tali differenze diventano gradualmente più marcate nel momento in cui la stessa si riduce di dimensione.

Il Metodo della Massima Verosimiglianza presenta, infine, notevoli vantaggi: dispone di test di adattamento (*fit*) del modello ai dati; permette, inoltre, di testare la significatività statistica delle saturazioni fattoriali e delle correlazioni tra fattori, e di calcolare gli intervalli di fiducia per tali parametri. Le stime della ML risultano generalmente non distorte, poiché indipendenti dalle scale sulle quali sono state misurate le variabili e, nello specifico, esse sono proporzionali alle deviazioni standard delle variabili stesse. Tuttavia, qualora l'assunzione di normalità multivariata non sia rispettata, tale metodo può dar luogo a risultati distorti.

Per campioni di numerosità elevata ($N \geq 400$), i tre metodi elencati tendono a fornire risultati simili; pertanto, si è deciso – nonostante la violazione dell'assunto di normalità multivariata – di utilizzare il metodo della massima verosimiglianza.

Scelto il metodo, pertanto, è necessario stabilire *quanti* fattori estrarre: per fare ciò esistono diverse tecniche, tra cui:

- Il Criterio degli Autovalori > 1 ,

- Lo “*Scree Test*”,
- L'Analisi Parallela.

Il criterio di Kaiser-Guttman degli Autovalori > 1 consiste nel calcolo degli autovalori della matrice di correlazione completa, selezionando per l'estrazione il numero di fattori pari alla numerosità degli autovalori che hanno valori maggiori di uno (Giannini e Pannocchia, 2006, p. 44).

Gli autovalori e gli autovettori (vettori associati ai primi), in breve, trovano applicazione nel sintetizzare “*l'informazione relativa alla varianza delle variabili e alla correlazione tra di esse*” (ibidem, p. 32). Tuttavia, poiché tale tecnica risulta incompatibile con la ML, dando luogo a risultati distorti (ivi, p. 44), si è optato per la sua esclusione dall'analisi.

Anche lo “*Scree Test*” di Cattell è basato sugli autovalori. Esso trova fondamento nell'idea che, qualora le somme dei quadrati delle saturazioni smettano di diminuire, rimanendo tendenzialmente costanti, è possibile decidere di interrompere l'estrazione dei fattori (ivi, p. 45): la regola di decisione si basa sul grafico detto *scree plot*, nel quale sull'asse delle ordinate si possono trovare gli autovalori della matrice di correlazione e sull'asse delle ascisse il numero di fattori ad essi relativi; si tratta quindi di identificare il punto di flesso in cui c'è il cambio di pendenza e la curva diviene sostanzialmente piatta (ivi).

Tale metodo, tuttavia, risente di una certa *ambiguità*, in quanto il cambio di pendenza può non essere così marcato e repentino da permettere un'identificazione univoca e sicura del numero di fattori da utilizzare.

Per ovviare a tale limite, è stata quindi elaborata l'Analisi Parallela, che impiega un confronto tra gli autovalori della matrice campionaria, e i dati elaborati casualmente tramite programmi informatici: si sceglie un numero di fattori comuni uguale al numero degli autovalori effettivamente superiori rispetto a quelli artificiali (ivi, p. 46). Anche questa tecnica risente di una certa arbitrarietà; pertanto, risulta opportuno affidarsi a più tecniche per ridurre l'incertezza e cercare di stabilire la quantità ottimale di fattori da estrarre.

In questo senso, attraverso la funzione “*fa.parallel*” del pacchetto “*psych*”, vengono integrate i due metodi dello Scree Test e dell'Analisi Parallela; l'output ottenuto (Figura 8) è stato il seguente:

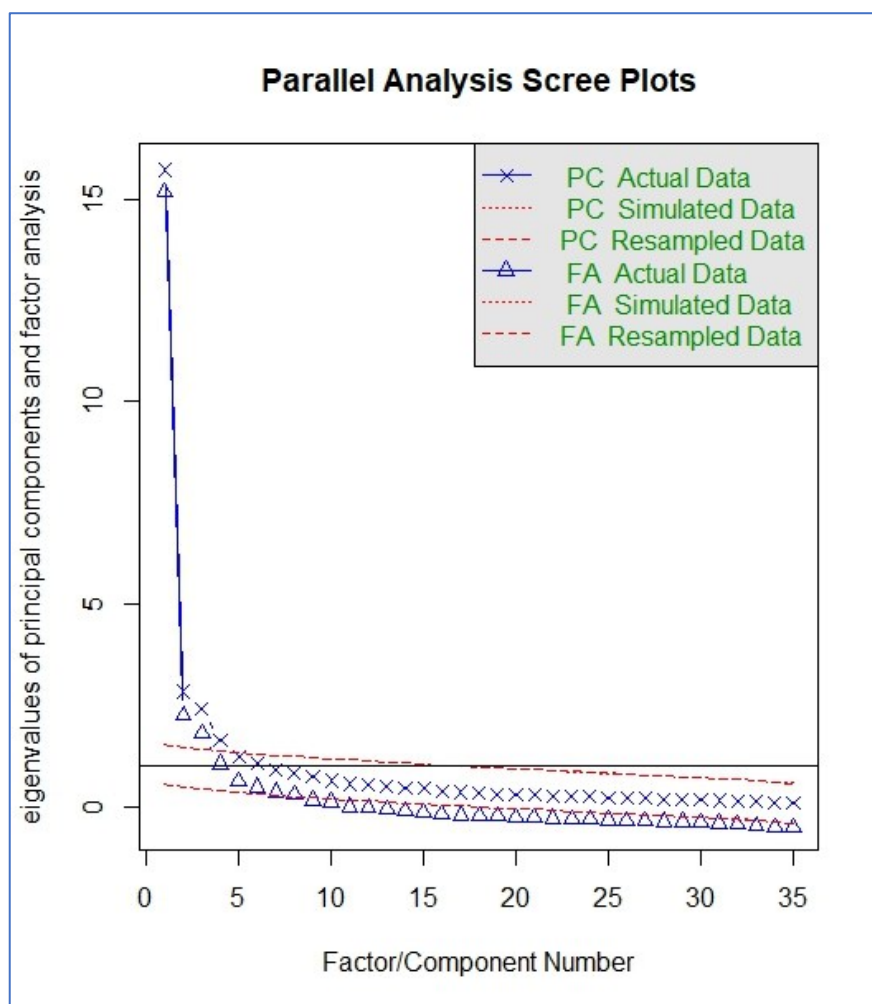


Figura 8: Scree-plot

Il risultato dell'analisi suggerisce un numero di fattori da estrarre pari a 8.

4. Rotazione dei fattori

Successivamente, al fine di semplificare l'interpretazione dei risultati, è necessario eseguire la rotazione dei fattori; essa consiste nello spostare i fattori all'interno dello spazio fattoriale in una posizione che ne renda più agevole la comprensione. Ciò è reso possibile dal fatto che, cambiando le saturazioni fattoriali, la quantità di varianza spiegata da ogni singolo fattore diventa più omogenea tra i diversi fattori (ivi, p. 51).

In primo luogo, quindi, bisogna distinguere tra rotazioni ortogonali ed oblique: le prime pongono come vincolo il fatto che i fattori devono essere tra loro non correlati, e sono generalmente più semplici da eseguire; tuttavia, sono spesso inadeguate per analizzare costrutti psicologici poiché essi sono spesso composti da fattori tra loro correlati. Impiegare una rotazione ortogonale con fattori correlati porta a risultati distorti (ivi, p. 62).

La rotazione obliqua, al contrario, può essere impiegata con fattori correlati; comporta elaborazioni più complesse, ma soluzioni tendenzialmente più affidabili e coerenti, poiché non

introduce alterazioni nel caso i fattori siano tra loro indipendenti e, stimando quanto le componenti co-variano tra loro, porta ad esiti più credibili (ivi).

Escludendo, in prima battuta, le modalità di rotazione di tipo obliquo – più complesse da manipolare – ci si è rivolti alle tecniche ortogonali, tra cui figurano metodi quali *Varimax* e *Quartimax*, maggiormente adatte agli scopi della ricerca:

- *Varimax* massimizza le saturazioni più elevate e minimizza quelle più basse, consentendo di distinguere meglio un fattore dall'altro;
- *Quartimax* concentra la maggior quantità di varianza per ogni variabile su un fattore unico, permettendo di identificare un fattore unico alla base del modello.

Si è optato per una rotazione “*Quartimax*”, poiché alla ricerca di un singolo fattore generale. Pertanto, attraverso il codice:

```
EFA_quartimax <- fa(r = dati_alldepvars_so,
  nfactors = 8,
  rotate = "quartimax",
  fm = "ml")
```

Sono state calcolate le saturazioni per ogni fattore. L'output dell'Analisi Fattoriale è sintetizzato in Tabella 37 (è stata stabilita, come soglia di *cut-off*, |0,2|):

	ML1	ML2	ML3	ML6	ML5	ML4	ML7	ML8
EDMQ1	0.55		0.26		0.42			
EDMQ2	0.56		0.44		0.33			
EDMQ3	0.47		0.62					
EDMQ4	0.39		0.76					
EDMQ5	0.41	0.22	0.74					
EDMQ6	0.48	0.29		0.25		0.29		
EDMQ7	0.59					0.63		
EDMQ8	0.60					0.60		
DER1	0.64				0.50			
DER2	0.63				0.46			
DER3	0.66				0.35			
DER4	0.71							
SED1	0.73				0.27			0.41
SED2	0.67		0.20					0.50
SED3	0.74					0.34		0.37
DTL1	0.64						0.25	
DTL2	0.72						0.32	
DTL3	0.66						0.49	
DTL4	0.81							
DTL5	0.73						0.25	
DA1	0.40	0.81						

DA2	0.42	0.80						
DA3	0.43	0.80						
DA4	0.45	0.74						
DA5	0.55	0.50						
DA6	0.48	0.47		0.41				
DA7	0.48	0.28		0.48				
DA8	0.39	0.27		0.60				
DA9	0.63			0.57				
DA10	0.47			0.55				
SDL1	0.71							
SDL2	0.75							
SDL3	0.84							
SDL4	0.89						-0.25	
SDL5	0.87							

Tabella 37: Saturazioni fattoriali

Come si evince dalla tabella, accanto ad un fattore principale, ML1, si possono identificare altri sette fattori, sostanzialmente distinti l'uno dall'altro. Questo dato preliminare può essere assunto come una ulteriore conferma della bontà dell'impostazione del questionario, poiché i fattori elencati seguono quasi perfettamente le scale proposte al suo interno.

Una volta ottenuto questo risultato, si è passati all'Analisi Fattoriale Confermativa, per verificare l'effettiva tenuta del modello identificato, attraverso l'uso degli indici di bontà dell'adattamento offerti dalla tecnica della Massima Verosimiglianza.

5. Analisi Fattoriale Confermativa

Come già accennato, si è scelto di eseguire un'Analisi Fattoriale Confermativa sull'intero set di dati a disposizione (N = 494), con l'intenzione di verificare l'effettivo adattamento del modello alle osservazioni. Quindi, attraverso il codice seguente, si è codificato il modello:

```
var_model <-'
ML2 =~ DA1 + DA2 + DA3 + DA4 + DA5 + DA6 + DA9
ML3 =~ EDMQ1 + EDMQ2 + EDMQ3 + EDMQ4 + EDMQ5 + DER1 + SED2 + DA1 + DA2
+ DA6
ML6 =~ DA4 + DA5 + DA6 + DA7 + DA8 + DA9 + DA10
ML5 =~ EDMQ1 + EDMQ2 + DER1 + DER2 + DER3 + SED1
ML4 =~ EDMQ1 + EDMQ7 + EDMQ8 + DER3 + SED3 + DA5 + DA8
ML7 =~ DA5 + DTL2 + DTL3
ML8 =~ SED1 + SED2 + SED3 + DA5 + DA9
ML1 =~ ML2 + ML3 + ML4 + ML5 + ML6 + ML7 + ML8 + DA5 + DA9
EDMQ2 ~~ EDMQ1
EDMQ4 ~~ EDMQ5
EDMQ5 ~~ DA7
DER1 ~~ DER2
DER2 ~~ DER3
```



```
DA1 ~~ DA2
DA8 ~~ DA10
DA6 ~~ DA7
DA2 ~~ EDMQ2
ML2 ~~ ML6
ML2 ~~ ML3
ML3 ~~ ML4
ML6 ~~ ML4
ML6 ~~ ML7
ML3 ~~ ML5'
```

Mentre, con i passaggi esposti in calce:

```
fit_var <- cfa(var_model, data = dati_alldevars_so,
              ordered = names(dati_alldevars_so))
```

```
summary(fit_var, standardized = TRUE, fit.measures = T)
```

```
mod_ind <- modindices(fit_var)
head(mod_ind[order(mod_ind$mi, decreasing=TRUE),], 10)
```

Sono state sia calcolate le statistiche relative al modello, sia gli indici utili per modificarlo nella direzione di una *fit* (adattamento) ancora migliore. Gli esiti sono qui esposti (Tabella 38):

	Modello dell'utente	Modello di Baseline
Comparative Fit Index (CFI)	1.000	0.992
Indice di Tucker-Lewis (TLI)	1.000	0.989
Root Mean Square Error of Approximation:		
RMSEA	0.000	0.044
Intervallo di confidenza al 90%:		
Limite inferiore	0.000	0.038
Limite superiore	0.000	0.043
Standardized Root Mean Square Residual:		
SRMR	0.024	0.024

Tabella 38: Misure di bontà di adattamento

Si può quindi commentare il risultato complessivo dell'analisi fattoriale, in primo luogo attribuendo un'etichetta alle scale individuate:

- Fattore 1 (ML1): Generale comprensione delle pratiche basate sui dati, o *Data-driven decision making*;
- Fattore 2 (ML2): Pratiche di *assessment* basate sui dati analogici, o *Data-driven Assessment*;

- Fattore 3 (ML3): Pratiche di progettazione didattica e/o *management* istituzionale (*Data-driven Evaluation of Educational Impact*);
- Fattore 6 (ML6): Pratiche di *assessment* basate su dati digitali, o automatici (*Data-driven Educational management*);
- Fattore 5 (ML5): Uso dei dati come contenuto didattico (*Data as Educational Resource*);
- Fattore 4 (ML4): Utilizzo partecipato dei dati;
- Fattore 7 (ML7): Uso dei dati per il processo decisionale docente (*Pedagogical Data Literacy*);
- Fattore 8 (ML8): Promozione della *data literacy* degli studenti (*Students' Data Literacy*).

In relazione ai punteggi di *fit* ottenuti e alle saturazioni fattoriali calcolate, si può concludere che il modello identificato possa essere considerato forte, con l'eccezione del Fattore 7 ("Uso dei dati per il processo decisionale docente"), il quale mostra saturazioni ridotte.

La quantità di varianza spiegata da ogni fattore è sintetizzata in Tabella 39:

	ML1	ML2	ML3	ML6	ML5	ML4	ML7	ML8
Saturazioni	13.501	3.470	2.070	1.613	1.144	1.084	0.691	0.688
Varianza proporzionale	0.386	0.099	0.059	0.046	0.033	0.031	0.020	0.020
Varianza cumulata	0.386	0.485	0.544	0.590	0.623	0.654	0.673	0.693

Tabella 39: Varianza spiegata dai fattori

Rispetto alle varianze individuate, si può notare come il primo fattore riesca a giustificare, da solo, quasi il 39% della varianza spiegata; il fattore due, un ulteriore 10%.

I primi due fattori, quindi, sono responsabili di quasi la metà della varianza spiegata. In totale, i fattori individuati spiegano circa il 69% della varianza.

Ulteriori precisazioni rispetto al modello individuato:

- Il fattore generale risulta maggiormente legato alla presa di decisioni didattiche basate sui dati, all'analisi critica dei dati assieme agli studenti, e alla promozione dell'apprendimento basato sui dati negli studenti;
- Il fattore cinque, "Uso dei dati come contenuto didattico", mostra una relazione inversa con le variabili DA ed SDL, suggerendo un approccio *tecnicistico* all'insegnamento dei dati, il quale non supporta una riflessione critica sul loro impiego;
- Rispetto all'utilizzo dei dati per il processo decisionale docente – fattore sette – si può constatare come esso correli negativamente con la scala relativa allo sviluppo

dell'alfabetizzazione ai dati degli studenti (SDL), evidenziando un accentramento del potere nelle mani del docente.

Riguardo le relazioni individuate tra le variabili latenti, si può osservare come il fattore generale mostri correlazioni elevate e positive con gli altri fattori ad esso subordinati, all'interno di un *range* compreso tra 0,57 (ML2), e 0,82 (ML7); inoltre:

- Le correlazioni tra i fattori “dipendenti” si rivelano generalmente basse, se non assenti (il fattore otto, “Promozione della data literacy degli studenti”, non correla con nessun altro dei fattori “secondari”);
- Tra i fattori di second'ordine, ML3 (Progettazione didattica), ed ML5 (Dati come risorsa per l'apprendimento) mostrano una elevata correlazione (0,48), e lo stesso può dirsi per ML2 (Assessment basato sui dati analogici) ed ML6 (Assessment basato sui dati digitali), i quali hanno una correlazione significativa (0,41);
- Di intensità più modesta (0,29) sono le correlazioni tra ML2 ed ML3, così come quella tra ML4 (Utilizzo partecipato dei dati) ed ML6 (0,28), e quella tra ML6 ed ML7 (Uso dei dati per il processo decisionale docente) (0,21);
- La correlazione tra ML3 e ML4 è pressoché nulla (0,05).

Il grafico seguente rappresenta le relazioni tra le variabili latenti (Figura 9):

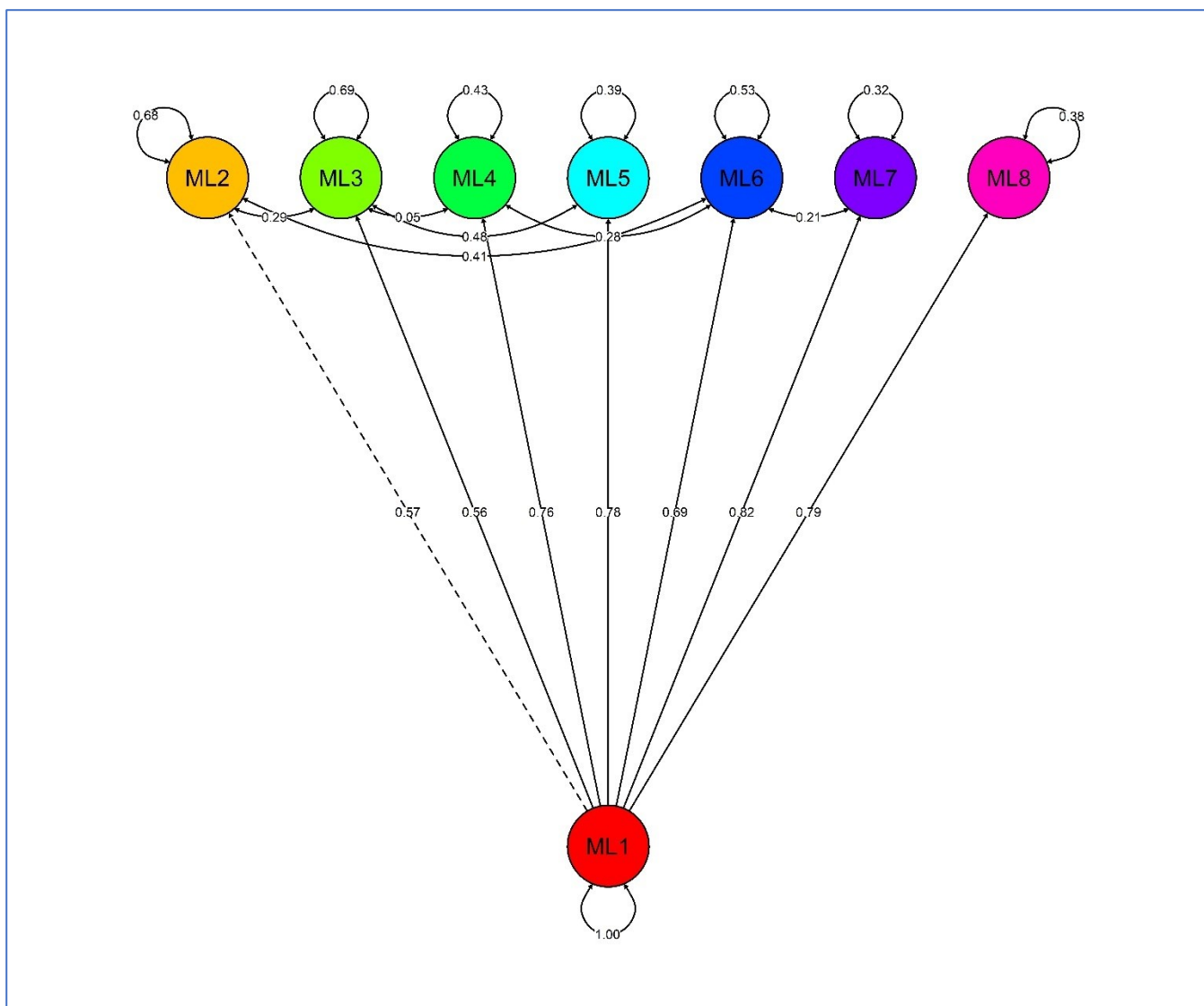


Figura 9: Relazioni tra le variabili latenti

Il *path diagram* inserito in calce, invece, sintetizza le informazioni sul modello complessivo (Figura 10):

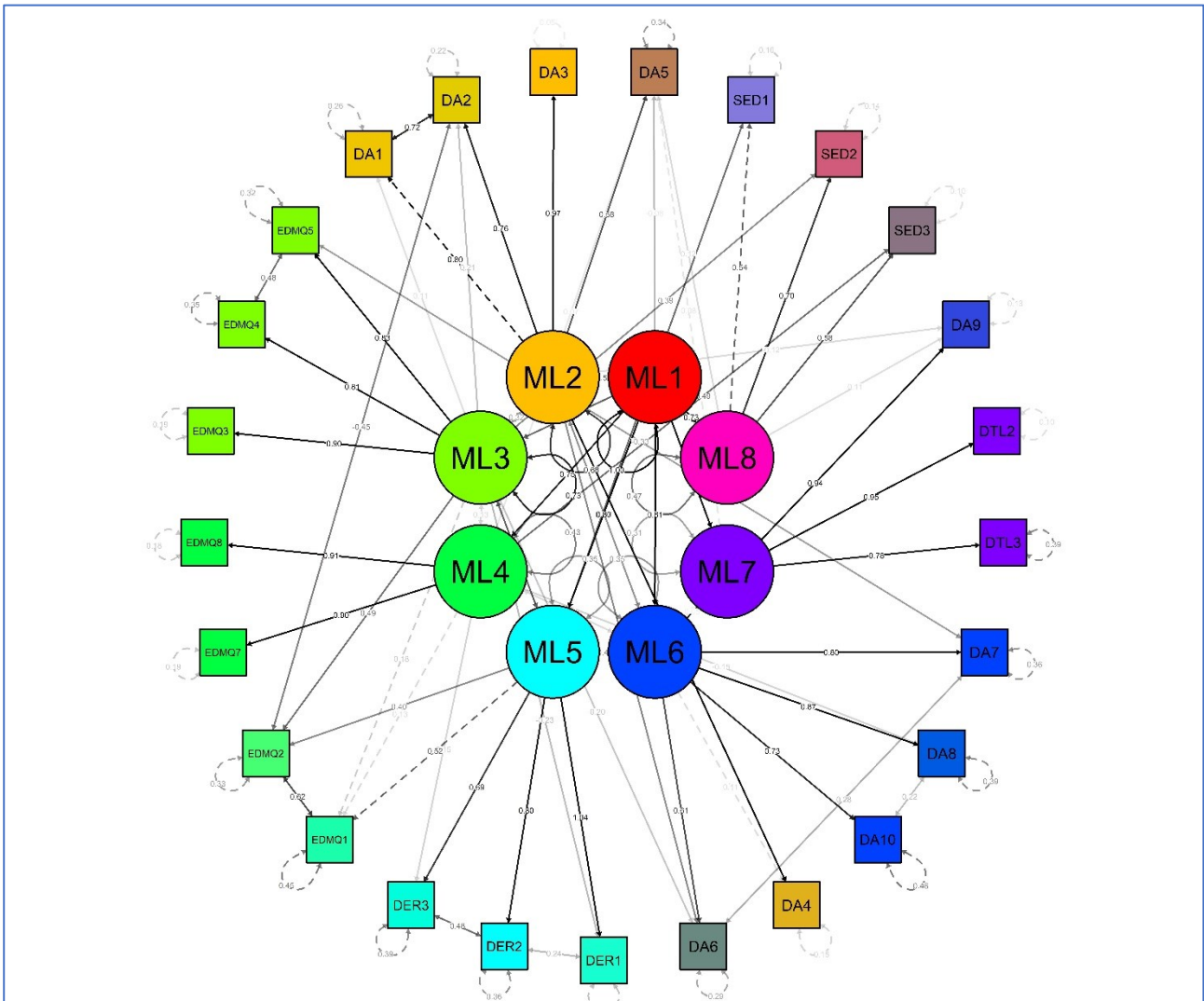


Figura 10: Path-diagram del modello elaborato

Il prossimo capitolo sarà dedicato alla discussione dei risultati ottenuti.

5. Discussione

Il grande volume di dati disponibili ha richiesto un'indagine varia ed approfondita al fine di individuare e chiarire le diverse relazioni tra le variabili e, sulla base di queste, delineare un modello fattoriale che possa essere utile per diagnosticare il fabbisogno formativo, rispetto alla *data literacy*, dei docenti universitari.

Da un punto di vista strettamente demografico, si può osservare come il campione sia formato da docenti appartenenti all'università spagnola per il 62% (337), mentre il restante 38% (210) al contesto italiano. Uomini e donne si distribuiscono in maniera pressoché equa (280 maschi e 261 femmine, pari rispettivamente al 52% e al 48% dei rispondenti, escludendo i casi non validi e i “Non dichiarato”), mentre la curva dell'età mostra uno spostamento verso i valori più elevati, in sintonia con l'andamento generale – osservato soprattutto nelle economie più avanzate – di un progressivo invecchiamento della popolazione. Non sorprende, quindi, che il 62% del campione (346 casi) abbia almeno 45 anni.

Da un punto di vista, invece, più strettamente accademico, si osserva che la maggioranza relativa dei docenti (il 41%, 227 casi) appartiene all'area delle scienze sociali, seguita dal gruppo di scienze della vita (il 20%, 110 casi); nell'insieme, la maggioranza dei rispondenti possiede più di 10 anni d'esperienza d'insegnamento (il 62%, 339 casi), nonché almeno undici anni di esperienza di ricerca (il 54%, 297 casi). Un numero rilevante, (80, il 15%) dichiara di non avere alcuna esperienza in questo campo.

Il quadro complessivo che emerge dall'analisi è quella di una diffusione ancora parziale delle pratiche legate ai dati all'interno del contesto accademico: il loro uso sembra legato alla valutazione (DA, $M = 2,28$; $DS = \pm 1,72$), come risorsa educativa (DER, $M = 2,21$; $DS = \pm 1,67$) e/o alla gestione istituzionale e della qualità della didattica (EDMQ, $M = 2,1$; $DS = \pm 1,72$).

Gli indici di forma calcolati per queste distribuzioni mostrano punteggi distribuiti con un'asimmetria positiva (Skewness uguale a 0,24, 0,27, 0,38) – quindi con una prevalenza di valori bassi – e una distribuzione *non-normale* dei punteggi, con una Curtosi negativa (rispettivamente di -1,28, -1,20, e -1,17), evidenziando come un grande numero di valori si concentri nelle “code” delle distribuzioni.

Analizzando le singole variabili, si può notare come nessuna di queste raggiunga valori medi collocati nella coda destra della distribuzione, ovvero all'interno del *range* più elevato: come osservato in precedenza, la pratica più diffusa è quella di uso dei dati da verifiche per dare feedback (DA3), con media di 2,92 e DS uguale a $\pm 1,69$; mentre la meno diffusa è l'estrazione di dati da social media per la valutazione dell'insegnamento (EDMQ8), con media pari a 1,09 e DS a $\pm 1,21$.

Gli item che possiedono i valori medi più elevati (nella fattispecie, con media maggiore di 2,5), e mostrano quindi la maggior diffusione all'interno del campione, sono, nell'ordine: DA3 (M = 2,92, DS = ±1,69), DA4 (M = 2,8, DS = ±1,67), DA2 (M = 2,72, DS = ±1,66), e DA1 (M = 2,65, DS = ±1,67); e EDMQ5 (M = 2,75, DS = ±1,77), EDMQ3 (M = 2,73, DS = ±1,72), ed EDMQ4 (M = 2,56, DS = ±1,77).

Per quanto concerne la scala relativa al *data-driven assessment*, gli item si riferiscono all' «uso dei dati ottenuti da verifiche per dare feedback sommativi agli studenti» (DA3), all' «uso dei dati ottenuti dalla valutazione di un corso per dare feedback formativi agli studenti» (DA4), all' «uso di dati ottenuti da verifiche progressive per il monitoraggio didattico» (DA2), e all'«uso di dati ottenuti da verifiche progressive per monitoraggio dell'apprendimento» (DA1). Per quanto riguarda, invece, la scala riferita al *data-driven educational management and quality*, gli item sono riferiti all' «uso di analitiche di apprendimento per la progettazione didattica» (EDMQ5), all' «uso di dati dalla valutazione del proprio corso per lo sviluppo istituzionale e la pianificazione» (EDMQ3) e, infine, all' «uso di dati del proprio corso per la progettazione curricolare» (EDMQ4).

Da ciò si può constatare come le pratiche di *data literacy* più diffuse siano quelle riferite alla valutazione, sia per fornire riscontri agli studenti che ai fini del monitoraggio della didattica e dell'apprendimento – sulla base delle informazioni ottenute da verifiche periodiche – e quelle implicate nella progettazione didattica, lo sviluppo istituzionale e la pianificazione, utilizzando informazioni provenienti, in maniera più o meno diretta, dai propri insegnamenti.

In questo senso, le analisi inferenziali sono state rivolte all'individuazione delle relazioni tra le variabili indipendenti, ovvero le caratteristiche dei soggetti, e le variabili dipendenti, cioè le sotto-scale del questionario.

L'analisi della varianza ANOVA eseguita ha messo in luce un effetto debole delle variabili “Genere” ed “Età” sulle *data practices* messe in atto dai docenti universitari, evidenziato dal basso livello di significatività degli item interessati. Anche l'esperienza d'insegnamento appare, in questo senso, poco rilevante, mentre l'area disciplinare e l'esperienza di ricerca si dimostrano maggiormente influenti.

Il fattore determinante in termini di *data practices* sembra essere la nazione di residenza, per la quale l'ANOVA ha messo in luce numerose relazioni rilevanti.

Tuttavia, al fine di comprendere più approfonditamente il nesso tra le variabili indipendenti e le risposte date agli item del questionario, è stato necessario eseguire dei test MANOVA: questi test hanno non solo confermato quanto già visto con l'ANOVA, ma anche evidenziato la *non-significatività* delle relazioni di “Genere” ($p = 0,54$) ed “Esperienza d'insegnamento” ($p = 0,13$) con le sotto-scale del questionario, mentre “Area disciplinare” ($p = 0,001$) ed “Età” ($p = 0,006$) si sono

dimostrate discretamente significative. Tra tutte, le variabili “Esperienza di ricerca” ($p = 9,361^{-7}$) e soprattutto “Nazione” hanno mostrato elevata significatività ($p < 2,2^{-16}$).

Le indicazioni emerse dai primi test suggeriscono una gerarchia sufficientemente chiara dei fattori di maggior impatto rispetto alle pratiche d'uso dei dati, in cui nazione ed esperienza di ricerca si rivelano le variabili più incisive, mentre l'età dei soggetti e l'appartenenza disciplinare risultano più marginali; infine, il genere e l'esperienza di insegnamento si dimostrano sostanzialmente ininfluenti.

Quanto appena esposto viene avvalorato dai calcoli sulla dimensione dell'effetto - eta quadro (η^2) - grazie ai quali è stato messo in luce un impatto estremamente elevato della variabile “Nazione” rispetto a tutte le altre ($\eta^2 = 0,6$), le quali rientrano nel *range* “moderato” (η^2 compreso tra 0,6 e 0,14).

La seconda variabile per dimensione dell'effetto – l'esperienza di ricerca – ottiene infatti un valore di *eta quadro* pari a 0,11; età e area disciplinare, un η^2 di 0,1; mentre l'esperienza di insegnamento ($\eta^2 = 0,08$) e il genere ($\eta^2 = 0,07$), come suggerito dalle diverse analisi della varianza, hanno un'importanza molto limitata.

Sulla scorta di queste prime evidenze, si può supporre che le politiche nazionali ed accademiche rispetto alle *data practices* portino a delle differenze notevoli in questo senso, così come la cultura organizzativa dell'istituzione; la dimensione dell'effetto, unita ai livelli di significatività osservati, dell'esperienza di ricerca, portano a considerare degno di nota il bagaglio di competenze del docente, per il quale l'abitudine a vagliare, consultare, e manipolare, dati di diversa origine e natura, risulta un'abilità capace di fare la differenza. Non deve sorprendere, quindi, che non sia l'esperienza d'insegnamento ad essere rilevante in termini di comportamenti d'uso dei dati, soprattutto in relazione alla didattica: le capacità che i docenti maturano nella pratica quotidiana, con tutta probabilità, modificano modi e contenuti dell'insegnamento, ovvero le modalità con cui gli insegnanti trasferiscono le loro conoscenze agli studenti.

Rispetto ai comportamenti specifici, le *comparazioni a coppie appaiate* (test di Games-Howell) hanno permesso di identificare le differenze all'interno dei diversi sottogruppi (ovvero, le modalità con cui si manifestano le variabili indipendenti).

Procedendo, quindi, in ordine di importanza – dal minore al maggiore – si conferma la scarsa significatività del genere sulla *data literacy*: benché sia emersa una differenza tra uomini e donne, essa si rivela ridotta, con un *range* della media delle differenze dei punteggi compreso tra -0,35 e -0,44, ed una sola sotto-scala (EDMQ5) con un *p-value* inferiore a 0,1 (in questo caso, $p = 0,006$).

Pertanto, anche se gli uomini dimostrano una maggiore propensione verso alcune pratiche d'uso dei dati – una su tutte, l'uso di analitiche di apprendimento per la progettazione didattica – è lecito

sospettare che tali differenze siano legate, perlopiù, al campionamento non-probabilistico utilizzato: un campione casuale e, possibilmente, più ampio, potrebbe evidenziare la completa insussistenza degli effetti del genere sulle *data practices*. È tuttavia necessario considerare che, al contrario – vista la sovra-rappresentazione del genere femminile nelle scienze sociali, e la sua corrispondente sotto-rappresentazione nelle discipline STEM (all'interno delle quali vi è un uso dei dati più intensivo) – potrebbe effettivamente sussistere una differenza di genere nelle pratiche d'uso dei dati: tale ipotesi merita, sicuramente, un maggiore approfondimento.

Discorso analogo merita l'anzianità di docenza: nel complesso, infatti, essa risulta ininfluenza, con un'unica variabile (EDMQ4, “*Uso dei dati del proprio corso per la progettazione curricolare*”), che risulta modestamente significativa ($p = 0,002$), all'interno del confronto tra docenti con meno di 3 anni d'esperienza e docenti con più di 15 anni d'anzianità di servizio.

Anche in questo caso, si può immaginare che i risultati ottenuti dipendano più dalle caratteristiche del campione, che da un reale effetto dell'esperienza d'insegnamento sulle pratiche connesse ai dati. Anche in questo caso è essenziale verificare che le differenze emerse – anziché un'anomalia legata al campionamento, alle caratteristiche dei dati, o ai metodi d'analisi impiegati – non rappresentino invece una reale discrepanza, in termini di *visione strategica*, tra docenti esperti e meno esperti. In questo senso, i primi sembrano possedere un orientamento, rispetto alla propria istituzione e alla progettazione didattica, maggiormente ampio, integrato e approfondito, in confronto ai docenti giovani. Pertanto, anche in questo caso, sarebbe più che opportuno appurare la fondatezza, o meno, di questa ipotesi.

Proseguendo nella disamina, “Età” ed “Area disciplinare” sembrano avere la stessa gravità rispetto alle pratiche digitali, sebbene con alcune differenze: l'età è risultata più rilevante rispetto alle modalità di *assessment* con i dati (DA) e per l'uso dei dati ai fini della gestione della qualità e della pianificazione istituzionale (EDMQ), mentre i confronti per settore di studio hanno indicato un effetto maggiore di questa variabile sulle pratiche connesse all'alfabetizzazione ai dati degli studenti (SDL) e al loro empowerment (SED).

Entrando nel dettaglio dei confronti più rilevanti, all'interno del gruppo “Età” sono state evidenziate differenze significative tra i docenti con età compresa tra 25 e 34 anni e il gruppo più anziano (“Più di 55 anni”), per cui i primi sembrano molto più propensi ad utilizzare i dati provenienti dalle piattaforme d'apprendimento per monitorare e valutare la didattica (DA6, $p = 0,003$); la media delle differenze degli over-55 è inferiore di 1,07 punti rispetto al gruppo più giovane.

Da segnalare la maggior attitudine del gruppo over-55 rispetto ai colleghi di età compresa tra i 35 e i 44 anni, alla riflessione etica sull'impiego dei dati aperti come risorse educative (DER4, $p =$

0,007), nonché all'uso di report istituzionali per lo sviluppo istituzionale e la pianificazione (EDMQ2, $p = 0,0007$), e all'impiego di dati del proprio corso per la progettazione curricolare (EDMQ4, $p = 0,0004$): le medie dei docenti più anziani sono superiori, rispettivamente, di 0,65, 0,75, e 0,84 punti rispetto al sottogruppo "35-44".

Invece, per quanto concerne le comparazioni per settore disciplinare, si è potuto constatare come i docenti del sottogruppo "Scienze Naturali e Formali" siano più inclini, rispetto ai colleghi delle discipline tecnologiche e di quelle sociali (aventi rispettivamente medie inferiori di 0,98 e 1,19 punti in confronto ai primi), ad impiegare informazioni del proprio corso per la progettazione curricolare (EDMQ4, $p = 0,005$). Gli insegnanti di "Scienze sociali", invece, risultano più propensi – in confronto ai colleghi di "Scienze tecnologiche" – a discutere assieme agli studenti riguardo i vincoli sociali all'uso dei dati e dell'IA (SDL3, $p = 0,008$), e sull'impatto dei dati sulla società nel suo insieme (SDL5, $p = 0,004$), avendo medie superiori di 0,56 e 0,68 punti. Nel paragone con "Scienze Naturali e Formali", essi si dimostrano più orientati, rispetto a questi ultimi, a condividere report istituzionali per supportare la riflessione degli studenti sui loro apprendimenti, all'interno del contesto istituzionale e sociale (SED1, $p = 0,003$), con una media superiore di 0,71 punti.

È da sottolineare come, proprio all'interno delle discipline tecnologiche, vi sia la minor propensione alle pratiche critico-riflessive rispetto all'uso e all'impatto che i dati e le innovazioni tecniche possono avere all'interno della società; ciò significa che coloro i quali, più probabilmente, saranno i fautori dei prossimi avanzamenti tecnologici, saranno anche i meno allenati al ragionare sulle implicazioni e le conseguenze di questi.

Le comparazioni per "Esperienza di ricerca", invece, sembrano evidenziare l'effetto della pratica di ricerca sui comportamenti di *data literacy* più frequentemente adottati: i docenti con più di 15 anni di esperienza, rispetto ai colleghi senza esperienza nel campo, sembrano meno inclini all'uso dei dati che provengono da piattaforme d'apprendimento allo scopo di monitorare e valutare la didattica (DA6, $p = 0,000571$), e ad utilizzare *dashboards* per informare la propria didattica (DA7, 0,000895); le medie delle differenze dei primi sono, infatti, inferiori di 0,96 e 0,98 punti. Tuttavia, quando viene preso in considerazione l'uso dei dati come risorsa educativa, la relazione si capovolge: gli insegnati con un'esperienza di ricerca compresa tra 6 e 10 anni, e con più di 15 anni, riutilizzano più spesso dati aperti, provenienti dalla propria attività di ricercatori, come risorse didattiche (DER3), in confronto ai colleghi senza esperienza; le medie dei due sottogruppi sono, in questo senso, superiori di 1,15 ($p = 0,000835$) e 0,86 punti ($p = 0,003$).

Una discrepanza ancora maggiore si può osservare nell'impiego dei dati per il *management* didattico-istituzionale: il gruppo dei più esperti mostra una marcata propensione, nel paragone con gli insegnanti senza esperienza di ricerca, all'uso di informazioni da report istituzionali per lo

sviluppo istituzionale e la pianificazione (EDMQ2, $p = 0,000297$), nel fare uso dei dati provenienti dal proprio corso per la progettazione curricolare (EDMQ4, $p = 1,03^{-9}$), e nell'utilizzo delle analitiche d'apprendimento per la progettazione didattica (EDMQ5, $p = 2,84^{-6}$); le medie delle differenze sono maggiori, infatti, di 0,9, 1,56, e 1,26 punti. Con riferimento alle ultime due variabili, anche il sottogruppo "11-15" dimostra una maggiore inclinazione verso tali pratiche rispetto ai docenti senza esperienza di ricerca: le medie delle differenze sono, infatti, superiori di 1,14 ($p = 0,000817$) e 1,42 punti ($p = 2,75^{-5}$), rispettivamente.

L'ultima variabile qui presa in esame, la nazione, è quella di maggior importanza. Essa ha mostrato le differenze esistenti, all'interno due contesti esaminati – università italiana e università spagnola – in termini di *data practices*: nel contesto italiano, infatti, gli insegnanti sembrano tendenzialmente più propensi a riutilizzare i dati della propria attività di ricerca nella didattica (DER3, +0,41; $p = 0,008$), ad impiegare report istituzionali (EDMQ2, +0,88; $p = 2,16^{-9}$), nonché dati dal proprio corso (EDMQ3, +1,03; $p = 3,05^{-11}$) per lo sviluppo accademico e la pianificazione, ad usare informazioni ricavate dai propri insegnamenti per la progettazione dei curricula (EDMQ4, +1,81; $p < 2,2^{-16}$), e a fare uso delle *learning analytics* per scopi di progettazione didattica (EDMQ5, +1,67; $p < 2,2^{-16}$). Differenze significative si segnalano anche nell'attitudine a promuovere le capacità degli studenti di vagliare e selezionare criticamente le informazioni (SDL1, +0,79; $p = 1,29^{-6}$), e a condividere con loro i dati ricavati da *learning analytics*, allo scopo di sostenere la riflessione sui loro apprendimenti nel contesto istituzionale e sociale (SED2, +0,63; $p = 1,76^{-5}$).

Se la realtà italiana presenta come maggiormente orientata all'uso dei dati per la gestione istituzionale e la qualità dell'insegnamento, nel contesto spagnolo si evidenzia soprattutto il maggior ricorso a pratiche di *assessment* attraverso i dati: in questo contesto è più frequente il ricorso alle analitiche d'apprendimento a fini di monitoraggio e valutazione della didattica (DA6, +1,05; $p < 2,2^{-16}$), all'utilizzo delle *dashboards* per informare la propria didattica (DA7, +0,91; $p = 9,8^{-11}$), e all'utilizzo delle informazioni ricavate dalle piattaforme per ragionare assieme agli studenti sulla qualità della didattica (DA9, +0,52; $2,46^{-5}$); superiore è anche l'uso dei *learning analytics* per analizzare la propria efficacia didattica (EDMQ6, +0,92; $p = 4,08^{-9}$).

Nell'insieme, i risultati sottolineano l'importanza del contesto istituzionale nell'orientare le pratiche d'uso dei dati e – sebbene con un impatto più ridotto – l'abitudine a trattare dati, nel proprio agire professionale quotidiano, in qualità di ricercatori.

Successivamente, il focus si è spostato sul consolidamento dello strumento utile all'analisi delle pratiche basate sui dati. A questo proposito, sono state svolte le analisi fattoriali esplorativa e confermativa. La prima – l'Analisi Fattoriale Esplorativa (AFE) – ha evidenziato il fatto che otto

fattori sarebbero stati sufficienti per spiegare i dati; i costrutti individuati sono sostanzialmente coincidenti con le scale teoriche già caratterizzate all'interno del questionario, ma con alcuni importanti elementi di differenziazione rispetto a queste, e che saranno oggetto di commento nella presente discussione.

Al fine di elaborare la matrice delle saturazioni ruotata, è stata scelta la rotazione “*quartimax*”. Tale scelta si deve, da un lato, all'obiettivo della ricerca (individuare un fattore latente comune), nonché alla volontà di semplificare l'interpretazione dei dati e, potenzialmente, aumentare il potere esplicativo del modello: caricare le saturazioni fattoriali su un fattore principale, infatti, ha l'effetto di ridurle su tutti altri, dando modo al ricercatore di distinguere con maggior semplicità i diversi fattori tra loro; questi, pertanto, risultano più circoscritti, più pregnanti e maggiormente aderenti ai dati (*Analisi Fattoriale*, 2023). Infatti, sebbene anche di poco, la varianza che tale modello è in grado di spiegare (il 69,3%) è superiore a quella ottenuta utilizzando altre rotazioni e, nello specifico, “*varimax*” e “*oblimin*”, le quali spiegano, in confronto, il 69% e il 51,4% della variabilità osservata nei dati.

Inoltre, postulare l'esistenza di una “generale abilità nel trattare i dati” ha un potenziale valore teorico, in quanto potrebbe condurre a meglio definire la competenza di *data literacy per gli educatori* in modo organico ed orientato allo sviluppo delle capacità di approccio critico ai dati, superando i limiti degli attuali *framework* di competenze, come il *DigCompEdu* (Raffaghelli, 2017). Si parla anche di utilità diagnostica, poiché può condurre all'elaborazione di un indice sintetico che quantifichi il grado di *alfabetizzazione ai dati* posseduto da un docente. In questo senso, i fattori di second'ordine individuati all'interno del modello offrono l'opportunità di esaminare analiticamente le singole competenze, nonché il grado di appropriazione raggiunto nelle specifiche abilità d'uso dei dati.

Oltre al fattore generale, infatti, i sette costrutti restanti sono relativi a capacità, particolari e tra loro distinte, che caratterizzano l'azione del docente rispetto ai dati; riepilogando, esse riguardano: le pratiche di *assessment* basate su dati analogici (Fattore 2); le pratiche di progettazione didattica e/o *management* istituzionale basate sui dati (Fattore 3); le pratiche di *assessment* basate su dati digitali e/o automatici (Fattore 6); uso dei dati come contenuto didattico (Fattore 5); utilizzo partecipato dei dati (Fattore 4); uso dei dati per il processo decisionale docente (Fattore 7); e, infine, all'uso dei dati per la promozione della *data literacy* degli studenti (Fattore 8).

Pertanto, tra gli otto fattori osserviamo – come già accennato in precedenza – una sostanziale sovrapposizione con la scala teorica. Tuttavia, le pratiche di *assessment* si dividono tra processi di valutazione basati su dati analogici, e processi basati su dati digitali; tale aspetto indica diversificazioni strutturali rispetto alle pratiche basate sui dati.

I fattori latenti – con particolare riferimento a quelli appena citati – si sono dimostrati sostanzialmente indipendenti l'uno dall'altro: le componenti che, all'interno del modello, correlano maggiormente (il Fattore 3, progettazione didattica e management istituzionale, con il Fattore 5, uso dei dati come contenuto didattico, e il Fattore 2 con il Fattore 6, inerenti rispettivamente all'*assessment* con dati analogici, e a quello con dati digitali), mostrano valori di covarianza moderati (0,48 e 0,41) (Chiorri, 2014, p. 246): ciò significa che la varianza spiegata dagli elementi ad essi comuni equivale rispettivamente al 23% e al 17% percento; tali risultati confermano un buon grado di indipendenza, ed autonomia, tra i fattori latenti individuati.

L'analisi fattoriale ha messo in luce un uso consistente delle pratiche di dati per l'*assessment* e come risorsa per l'apprendimento; tuttavia, si sono constatate prassi di accentramento delle decisioni nelle mani del docente rispetto alla gestione della classe.

In relazione a questo, il costrutto 7 – che riflette le pratiche decisionali dei docenti tramite i dati – correla negativamente con gli item SDL3 (-0,12), SDL4 (-0,24) ed SDL5 (-0,19), relativi alla riflessione, nonché alla discussione, assieme agli studenti, rispetto all'uso dei dati e all'impatto di questi nella società; anche nell'uso dei dati come contenuto didattico (Fattore 5), si osservano covarianze negative rispetto a questi elementi (-0,11, -0,16, e -0,19) e con DTL4 (Riflessione del docente rispetto all'uso di dati automatici), con un valore di -0,14. Ciò denota uno stile didattico caratterizzato da un approccio tecnicistico rispetto ai dati, che non favorisce l'appropriazione critica ed autonoma di tali pratiche da parte degli studenti.

Queste considerazioni si potrebbero ripetere, in maniera del tutto analoga, per il costrutto 8 (Promozione della data literacy degli studenti), il quale mostra covarianze negative con SDL4 (-0,15) e SDL5 (-0,14).

L'esigenza di sviluppare l'alfabetizzazione digitale negli studenti si scontra con una realtà pedagogica che, all'atto pratico, non favorisce l'agentività degli studenti, né attraverso la condivisione di informazioni e di potere decisionale da parte del docente, né attraverso prassi educative rivolte ad una *data literacy* che sia riflessiva, critica e trasformativa. Al contrario, lo scenario emergente è quello di competenze digitali considerate come *skill* meramente tecniche, utili per costruire solo un "profilo di competenze" spendibili nel mondo del lavoro.

Elaborare un quadro teorico delle competenze digitali degli educatori che sia coerente, significativo, orientato ad un approccio critico verso la tecnologia, empiricamente fondato e operativamente utile, è, nell'attuale contesto di crescente datificazione e digitalizzazione, a tutti i livelli della società, una necessità stringente e non differibile. Come sottolineato da Yang e Li (2020, p. 8-9; TdA), «*l'alfabetizzazione ai dati fa parte delle competenze del ventunesimo secolo che contribuiscono all'occupabilità*».

Gli educatori, infatti, contribuiscono al successo accademico dei propri studenti attraverso l'uso dei dati relativi ad insegnamento e apprendimento; in quest'ottica, i dati che gli studenti possiedono – su loro stessi, e sull'efficacia dell'insegnamento – possono alimentare una sinergia capace di migliorare la qualità dell'insegnamento stesso (ibidem), attraverso la condivisione di queste informazioni con educatori in grado di gestirli in maniera adeguata.

L'integrazione dei propri dati di ricerca nella didattica risulta, inoltre, uno degli elementi fondamentali di una pratica educativa di alfabetizzazione ai dati che sia efficace, esortando quindi i docenti a migliorare le proprie capacità di impiegare i propri dati di ricerca nella pratica d'insegnamento; queste pratiche professionali contribuirebbero all'*engagement* degli studenti e alla loro esperienza d'apprendimento (ivi). Come sottolinea Raffaghelli (2021, p. 9), citando Atenas et al. (2015), l'*«utilizzo dei dati prodotti in ambito di ricerca, che come dati aperti secondo i canoni della Open Science, possono costituire importanti risorse educative per gli studenti»*.

Ciò avviene con maggiore probabilità se i docenti svolgono l'attività di ricercatori; come affermato in precedenza, l'esperienza di ricerca è il secondo fattore più rilevante per le *data practices*, almeno per quanto concerne il campione analizzato. Tuttavia, aggiunge l'autrice, *«queste capacità non possono essere solo tecniche, ma richiedono una contestualizzazione critica dei dati prodotti, trattati e utilizzati con finalità sociali, istituzionali e politiche»*.

L'attuale contesto socio-culturale ha assistito, infatti, all'emergere dei movimenti di “*Open Access*” (contro la privatizzazione della conoscenza), di “*Open Science*” (che vede la libera condivisione di dati tra i ricercatori) e di “*Open Government*” (per la trasparenza nell'uso delle risorse pubbliche) (Raffaghelli, 2017, p. 300), i quali hanno prodotto, e producono, una enorme varietà e volume di dati, liberamente accessibili.

Questa abbondanza di informazioni nella nostra quotidianità, tuttavia, pone delle sfide non banali; essa richiede una cittadinanza, attiva e coinvolta, che sia in grado di approcciarsi in maniera autonoma, e di gestire appropriatamente, tale volume di dati: come sottolinea Raffaghelli (2017, p. 307), citando Zuiderwijk & Janssen (2012, 2014), *«se si desidera che il semplice accesso agli Open Data si trasformi in effettivo monitoraggio civico e responsabilizzazione della cittadinanza, sarebbe necessario sviluppare le capacità minime per comprendere quali siano i problemi sociali monitorati attraverso i dati e per leggere le eventuali rappresentazioni grafiche e statistiche già disponibili, al fine di formulare nuovi quesiti»*.

In questo senso, è necessario considerare come le capacità di elaborazione e interpretazione delle informazioni siano, in primo luogo, di natura cognitiva; è, pertanto, necessario abituare le menti a confrontarsi con generi diversi di dati, a valutarne la credibilità e l'affidabilità, ad integrarli in

schemi che diano loro un senso, e a essere capaci di comunicarli all'esterno; nonché a saper valutare gli aspetti etici e sociali ad essi correlati.

Tale forma di controllo, consapevole e da parte dei cittadini non può essere slegata da uno slancio verso una società più equa e giusta: come riporta Raffaghelli (2020, p. 16; TdA), in una revisione della letteratura sugli effetti della *data literacy* come catalizzatore di giustizia sociale, «*non c'è possibilità di promuovere la giustizia sociale in uno spazio in cui il fenomeno della datificazione costringe gli individui ad un uso passivo (e in qualche modo inconsapevole) della tecnostuttura*».

Più nello specifico, riassume l'autrice (ibidem), «*l'alfabetizzazione ai dati, come forma specifica di intervento educativo, espande lo spazio di opportunità, connettendo lo sviluppo di consapevolezza, abilità, e conoscenze, con forme d'interazione agentiva: uso creativo e produttivo di dati pubblici (Dimensione 1); progettazione partecipata della tecnostuttura [...], o reazione verso forme di manipolazione (Dimensioni 2 e 3); e coinvolgimento nell'uso dei dati per informare la pratica educativa (Dimensione 4)*».

Sviluppare le competenze digitali nei docenti universitari risulta, perciò, prioritario ai fini della trasmissione e promozione tali abilità negli studenti; Raffaghelli (2021, p. 11), citando Felisatti e Serbati (2019) e Raffaghelli (2017), afferma: «*La DS (Digital Scholarship, NdA) definita come un insieme di competenze del docente per lavorare in contesti di ricerca e insegnamento digitali, in rete e aperti, avrà un impatto sulla competenza digitale, il pensiero critico e l'impegno civico degli studenti*». Lo sviluppo di competenze critiche, riflessive, di *judgement* e *decision-making*, deve essere posto al centro dei futuri quadri di competenze, in special modo se riferiti all'ambito del digitale. Sarà proprio all'interno di tale contesto che le persone si giocheranno l'occasione di essere cittadini e lavoratori consapevoli, critici, responsabili, attivi e proattivi verso le innovazioni tecnologiche.

Va considerato, a questo punto, che il contesto globale evolve con rapidità, e vedrà una sempre più massiva integrazione di sistemi di Intelligenza Artificiale all'interno delle infrastrutture digitali, le quali saranno sempre più presenti in ogni contesto di studio, di vita, e di lavoro.

Per garantire il rispetto dei propri diritti fondamentali, sarà indispensabile mettere gli individui nelle condizioni di cogliere il funzionamento di tali sistemi e tali infrastrutture, nonché delle numerose implicazioni – in primo luogo, etiche e sociali – correlate all'introduzione delle IA: senza una regolamentazione etica, il rischio è che tali applicazioni riproducano – in maniera del tutto automatica ed impersonale – *bias* e discriminazioni, il cui effetto sarebbe quello di esacerbare disuguaglianze ed ingiustizie sociali (*Ethics of Artificial Intelligence* | UNESCO, s.d.).

Allo scopo di evitare il verificarsi di un simile scenario, la Commissione Europea, grazie all'attività di un *pool* di esperti, ha elaborato gli “Orientamenti etici per gli educatori sull'uso dell'intelligenza artificiali (IA) e dei dati nell'insegnamento e nell'apprendimento” (Directorate-General for Education, 2022), attraverso il quale indirizzare i professionisti nel campo dell'educazione ad un uso etico, equo, sicuro, rispettoso della privacy e al servizio del bene comune delle IA e dei dati nell'istruzione (ibidem, p. 11).

Il lavoro che è stato compiuto dalla Commissione supporta, attraverso la diffusione della *data literacy*, la comprensione di questi sistemi, in particolare in un'ottica di *IA literacy*.

Per concludere, la cultura dei dati e del digitale è, e sarà ancora di più negli scenari della IA, centrale nel garantire ad ognuno delle concrete opportunità di equità e giustizia sociale.

6. Conclusioni

Il presente lavoro ha cercato di offrire un excursus, sufficientemente ampio ed approfondito, dei fenomeni di digitalizzazione e datificazione, esaminando i mutamenti in atto nella società e nell'educazione. Lo scopo era quello di discutere opportunità e problematicità di questi nuovi fenomeni, analizzandoli nel contesto dell'istruzione superiore e con particolare riferimento alle pratiche d'uso dei dati in due contesti accademici differenti. I risultati hanno permesso di elaborare un modello teorico delle attuali prassi, nonché di identificare alcuni profili di criticità nell'agire educativo, che andrebbero necessariamente affrontati con opportuni programmi di *faculty development*.

L'orientamento attuale, infatti, considera le competenze digitali da una prospettiva tecnicistica, connessa con esigenze di competitività professionale dei laureati, e di istanze che provengono dal mondo delle imprese: questa trattazione, infatti, ha sottolineato come i sistemi educativi siano stati oggetto di azioni volte, da un lato, a fornire agli studenti le competenze richieste dal mondo del lavoro; dall'altro, a rendere le istituzioni educative – sulla scia di una gestione di tipo imprenditoriale – più “efficaci” ed “efficienti”, introducendo parametri di valutazione delle “performance” educative che mal si adattano a progetti pedagogici orientati alla promozione autentica dell'agentività individuale.

Quindi, oltre alle considerazioni sull' “occupabilità” – tema di sicuro ricorrente nei dibattiti riguardanti il futuro dell'educazione – e sull'efficacia dell'insegnamento, risulta essenziale al giorno d'oggi offrire agli studenti, in special modo se saranno gli educatori di domani, l'opportunità di sviluppare le capacità necessarie ad operare in modo critico, con autonomia di pensiero e giudizio, nel mondo di oggi e, per quanto possibile, nella realtà di domani. Tale realtà, ad ora puramente ipotetica, vedrà la diffusione di innovazioni – quali l'intelligenza artificiale, la realtà aumentata, una sempre maggior datificazione e digitalizzazione – in ogni ambito dell'esperienza umana. È, e sarà sempre più necessario, fornire a educatori e studenti, gli strumenti adatti a navigare in queste impetuose, nonché dirompenti, trasformazioni.

7. Bibliografia:

Raffaghelli, J. E., Manca, S., Stewart, B., Prinsloo, P., & Sangrà, A. (2020). Supporting the development of critical data literacies in higher education: Building blocks for fair data cultures in society. *Int J Educ Technol High Educ*, 22. <https://doi.org/10.1186/s41239-020-00235-w>

Williamson, B., & Hogan, A. (2021). *Pandemic Privatisation in Higher Education: Edtech & University Reform*. Education International.

Raffaghelli, J. E. (2017). Alfabetizzare ai dati nella società dei big e open data: Una sfida formativa Data Literacy in the context of Big and Open Data: An educational challenge. *Formazione & Insegnamento*. https://doi.org/doi:107346/-fei-XV-03-17_21

Lane, M., & Saint-Martin, A. (2021). *The impact of Artificial Intelligence on the labour market: What do we know so far?* (OECD Social, Employment and Migration Working Papers Fasc. 256; OECD Social, Employment and Migration Working Papers, Vol. 256). <https://doi.org/10.1787/7c895724-en>

OpenAI. (2023). In *Wikipedia*. <https://it.wikipedia.org/w/index.php?title=OpenAI&oldid=132183563#ChatGPT>

ChatGPT. (2023). In *Wikipedia*. <https://it.wikipedia.org/w/index.php?title=ChatGPT&oldid=132257299>

GPT-3. (2023). In *Wikipedia*. <https://it.wikipedia.org/w/index.php?title=GPT-3&oldid=132245137>

Stewart, B. (2023). The problem of the web: Can we prioritize both participatory practices and privacy? *Contemporary Educational Technology*, 15(1), ep402. <https://doi.org/10.30935/cedtech/12668>

Crawford, K. (2022). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*.

Terre rare. (2023, gennaio 24). Wikipedia. Recuperato 21 febbraio 2023, da https://it.wikipedia.org/w/index.php?title=Terre_rare&oldid=131720837

European Union Agency for Cybersecurity. (2022). *ENISA threat landscape 2022: July 2021 to July 2022*. Publications Office. <https://data.europa.eu/doi/10.2824/764318>

Dijck, J.van, Poell, T. and Waal, M.de (2018) *The Platform Society: public values in a Connective World*. New York, N.Y: Oxford University Press.

Raffaghelli, J. E. (2019). Alfabetizzare ai dati attraverso l'uso degli Open Data. Uno studio di caso nella formazione iniziale degli educatori. *Idee in form@zione. Educare alla cittadinanza attiva*, 7, 101–128. <https://doi.org/10.4399/97888255202625>

Atenas, J., Havemann, L., & Timmermann, C. (2020). Critical literacies for a datafied society: Academic development and curriculum design in higher education. *Research in Learning Technology*, 28(0). <https://doi.org/10.25304/rlt.v28.2468>

Selwyn, N., & Jandrić, P. (2020). Postdigital Living in the Age of Covid-19: Unsettling What We See as Possible. *Postdigital Science and Education*, 2(3), 989–1005. <https://doi.org/10.1007/s42438-020-00166-9>

Raffaghelli, J. E., Manca, S., Stewart, B., Prinsloo, P., & Sangrà, A. (2020). Supporting the development of critical data literacies in higher education: Building blocks for fair data cultures in society. *Int J Educ Technol High Educ*, 22. <https://doi.org/10.1186/s41239-020-00235-w>

Raffaghelli, J. E., Grion, V., & Rossi, M. (2021). Data practices in quality evaluation and assessment: Two universities at a glance. *Higher Education Quarterly*, hequ.12361. <https://doi.org/10.1111/hequ.12361>

Wikimedia Foundation. (2023, 28 gennaio). *Lavoro Agile*. Wikipedia. URL consultato il 9 febbraio 2023, from https://it.wikipedia.org/wiki/Lavoro_agile

Windows 11 e la privacy: Il confronto con Windows XP è impietoso. (URL consultato in data 13/02/2023). Hardware Upgrade. Recuperato 13 febbraio 2023, da https://www.hwupgrade.it/news/sistemi-operativi/windows-11-e-la-privacy-il-confronto-con-windows-xp-e-impietoso_113978.html

Knox, J. (2019). Postdigital as (Re)Turn to the Political. *Postdigital Science and Education*, 1(2), 280–282. <https://doi.org/10.1007/s42438-019-00058-7>

Floridi, Luciano, AI as Agency Without Intelligence: On ChatGPT, Large Language Models, and Other Generative Models (February 14, 2023). *Philosophy and Technology*, 2023, Disponibile al SSRN: <https://ssrn.com/abstract=4358789> or <http://dx.doi.org/10.2139/ssrn.4358789>

Terre rare nell'Enciclopedia Treccani. (s.d.). Consultato il 21 febbraio 2023, da <https://www.treccani.it/enciclopedia/terre-rare>

Competenze chiave per l'apprendimento permanente. (2006).

<https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:394:0010:0018:it:PDF>

Joint Research Centre (European Commission), Brande, L. V. den, Carretero, S., Vuorikari, R., & Punie, Y. (2016). *DigComp 2.0: The digital competence framework for citizens*. Publications Office of the European Union. <https://data.europa.eu/doi/10.2791/11517>

Vuorikari, R., Kluzer, S., & Punie, Y. (2022). *DigComp 2.2, The Digital Competence framework for citizens: With new examples of knowledge, skills and attitudes*. Publications Office of the European Union.

Calvani, A., Fini, A., & Ranieri, M. (2009). LA COMPETENZA DIGITALE NELLA SCUOLA. MODELLI TEORICI E STRUMENTI DI VALUTAZIONE. *International Journal of Developmental and Educational Psychology*, 4(1), 11.

Raffaghelli, J. E., & Stewart, B. (2020). Centering complexity in ‘educators’ data literacy’ to support future practices in faculty development: A systematic review of the literature. *Teaching in Higher Education*, 25(4), 435–455. <https://doi.org/10.1080/13562517.2019.1696301>

Raffaghelli, J., & Stewart, B. (2021). The Educators’ Datafied Present and Future: Complexity as an Approach to Developing Educators’ Data Literacies. *The Open/Technology in Education, Society, and Scholarship Association Conference*, 1(1), 1–8. <https://doi.org/10.18357/otessac.2021.1.1.49>

Raffaghelli, J. E. (2017). Exploring the (missed) connections between digital scholarship and faculty development: A conceptual analysis. *International Journal of Educational Technology in Higher Education*, 14(1), 20. <https://doi.org/10.1186/s41239-017-0058-x>

Raffaghelli, J. E. (2021). «Datificazione» e istruzione superiore: Verso la costruzione di un quadro competenziale per una rinnovata Digital Scholarship. *EXCELLENCE AND INNOVATION IN LEARNING AND TEACHING*, 0, 127–147. <https://doi.org/10.3280/exioa0-2021oa11132>

PNRR Italia.pdf, 273 (2022).

Fedeli, M., & Tino, C. (2019). Teaching4Learning@Unipd: Instruments for faculty development. *Form@re - Open Journal per la formazione in rete*, 105-121 Pages. <https://doi.org/10.13128/FORMARE-25191>

De Rossi, M., & Fedeli, M. (2022). *De Rossi e Fedeli (2022) Costruire percorsi di faculty development.pdf*. Pensa MultiMedia Editore.

Perla, L., Felisatti, E., Grion, V., Agrati, L. S., Gallelli, R., Vinci, V., Amati, I., & Bonelli, R. (2020). Oltre l'era Covid-19: Dall'emergenza alle prospettive di sviluppo professionale. *EXCELLENCE AND INNOVATION IN LEARNING AND TEACHING*, 2, 2. <https://doi.org/10.3280/exioa2-2020oa10802>

Faculty development e didattica universitaria. (2023). SIPED. <https://www.siped.it/gruppi-di-lavoro/faculty-development-e-didattica-universitaria/>

Raffaghelli, J. E., Grion, V., & De Rossi, M. (2021). Pratiche basate sui dati nella valutazione e l'analisi della qualità didattica: Il caso dell'Università di Padova. *Qwerty. Open and Interdisciplinary Journal of Technology, Culture and Education*, 16(1), 58–79. <https://doi.org/10.30557/QW000036>

Ranieri, M. (2022). Le competenze digitali degli insegnanti. In R. Biagioli & S. Oliviero (A c. Di), *Strumenti per la didattica e la ricerca* (1^a ed., Vol. 222, pp. 49–60). Firenze University Press. <https://doi.org/10.36253/978-88-5518-587-5.6>

Raffaghelli, J. E. (2023). *Construir culturas de datos justas en la universidad. Desafíos para el profesorado* (Prima ed.). Ediciones Octaedro. <https://octaedro.com/wp-content/uploads/2023/04/9788419506450.pdf>

Farrow, R., Iniesto, F., Weller, M. & Pitt., R. (2020). The GO-GN Research Methods Handbook. Open Education Research Hub. The Open University, UK. CC-BY 4.0. http://go-gn.net/gogn_outputs/research-methods-handbook/

Critical realism. (2022, 30 Aprile). *New World Encyclopedia*, . Consultato il 24 maggio 2023, da https://www.newworldencyclopedia.org/p/index.php?title=Critical_realism&oldid=1068635.

Complexity. (2023). In *Wikipedia*. Consultato il 24 maggio 2023, da <https://en.wikipedia.org/w/index.php?title=Complexity&oldid=1154862148>

Uncertainty. (2023). In *Wikipedia*. Consultato il 24 maggio 2023, da <https://en.wikipedia.org/w/index.php?title=Uncertainty&oldid=1153441033>

Raffaghelli, J. E. (2019a). *Delphi Study to validate the survey exploring academic awareness and engagement with data-driven practices*. Barcelona. <https://doi.org/10.5281/ZENODO.3581290>

Juliana Elisa Raffaghelli, Valentina Grion, & Marina De Rossi. (2020). Studio esplorativo sulle pratiche basate sui dati nella didattica universitaria: il caso dell'Università di Padova [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.4110000>

Raffaghelli, J. E. (2019). *ALLEGATO I-StruttQuestionario.docx*.

R: *What is R?* (2023). Recuperato 24 maggio 2023, da <https://www.r-project.org/about.html>

Posit. (2023). Posit. <https://www.posit.co/>

Home—RDocumentation. (2023). <https://www.rdocumentation.org/>

LMS Platform—Moodle LMS - Learning Management System. (2023). Moodle. <https://moodle.com/solutions/lms/>

Overview | Moodle Developer Resources. (2023, gennaio 20). <https://moodledev.io/general/community>

Chiorri, C. (2014). *Fondamenti di Psicometria* (Seconda edizione). McGraw-Hill Education.

Mair, P. (2018). *Modern Psychometrics with R*. Springer. <https://doi.org/10.1007/978-3-319-93175-3>

Pozzolo, P. (2021, dicembre 18). *ANOVA: L'analisi della varianza spiegata semplice*. Paola Pozzolo. <https://paolapozzolo.it/anova-introduzione/>

Analisi Fattoriale. (2023, giugno 1). [www.rescoop.com. http://www.rescoop.com/multivariata/AnalisiFattoriale.htm#3](http://www.rescoop.com/multivariata/AnalisiFattoriale.htm#3)

Giannini, M., & Pannocchia, L. (2006). *L'Analisi Fattoriale Esplorativa in Psicologia — Una guida pratica per la ricerca*. Giunti O.S.

van Geel, M., Keuning, T., Visscher, A., & Fox, J.-P. (2017). Changes in educators' data literacy during a data-based decision making intervention. *Teaching and Teacher Education*, 64, 187–198. <https://doi.org/10.1016/j.tate.2017.02.015>

Yang, N., & Li, T. (2020). How stakeholders' data literacy contributes to student success in higher education: A goal-oriented analysis. *International Journal of Educational Technology in Higher Education*, 17(1). Scopus. <https://doi.org/10.1186/s41239-020-00220-3>

Elisa Raffaghelli, J. (2020). Is Data Literacy a Catalyst of Social Justice? A Response from Nine Data Literacy Initiatives in Higher Education. *Education Sciences*, 10(9), 233. <https://doi.org/10.3390/educsci10090233>

Ethics of Artificial Intelligence | UNESCO. (s.d.). Recuperato 12 giugno 2023, da <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>

European Commission, Directorate-General for Education, Youth, Sport and Culture, (2022). *Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators* – , Publications Office of the European Union. <https://data.europa.eu/doi/10.2766/153756>

8. Allegati

8.1. Script di R per l'analisi dei dati

8.1.1. Script per l'analisi univariata e bivariata

```
# Script per l'analisi univariata e bivariata

# 1. Preparo il workspace

# install.packages("flextable")
# install.packages("rempsyc")
# install.packages("summarytools")

library(dplyr)
library(flextable)
library(psych)
library(rempsyc)
library(rstatix)
library(summarytools)

setwd("C:/Users/pc/Documents/UNIV/TESI DI LAUREA/Dataset")
dati_two_universities <- read.table("DataPractices_TwoUniversities.csv",
                                   header = TRUE, sep = ",")

dati_two_universities[dati_two_universities == "0,00"] <- "0"
dati_two_universities[dati_two_universities == "1,00"] <- "1"
dati_two_universities[dati_two_universities == "2,00"] <- "2"
dati_two_universities[dati_two_universities == "3,00"] <- "3"
dati_two_universities[dati_two_universities == "4,00"] <- "4"
dati_two_universities[dati_two_universities == "5,00"] <- "5"
View(dati_two_universities)

# 2. Convento le variabili
# 2.1. Genere

Gender <- dati_two_universities$Gender
Gender <- factor(Gender,
                levels = c("Male", "Female"))

# 2.2. Et 

Age <- dati_two_universities$Age
Age <- factor(Age,
             levels = c("Less than 25", "25-34", "35-44",
                       "45-54", "More than 55"))

# 2.3. Nazione
```

```
Country <- dati_two_universities$Country
Country <- factor(Country,
  levels = c("Italy","Spain"))
```

2.4. Area disciplinare

```
Subject.field.1 <- factor(dati_two_universities$Subject.field.1,
  levels = c("Agricultural Sciences",
    "Anthropology",
    "Anthropology,History",
    "Anthropology,Pedagogy",
    "Astronomy and astrophysics",
    "Chemistry",
    "Chemistry,Life Sciences,Agricultural Sciences",
    "Demographics",
    "Demographics,Sociology",
    "Earth and Space Sciences",
    "Earth and Space Sciences,Agricultural Sciences",
    "Earth and Space Sciences,Technological Sciences",
    "Earth and Space Sciences,Agricultural Sciences,Geography",
    "Economic Sciences",
    "Economic Sciences,Ethics",
    "Economic Sciences,History",
    "Economic Sciences,Juridical Sciences and Law",
    "Economic Sciences,Pedagogy",
    "Economic Sciences,Pedagogy,Psychology",
    "Ethics",
    "Geography",
    "History",
    "Juridical Sciences and Law",
    "Juridical Sciences and Law,Psychology",
    "Juridical Sciences and Law,Political Science,Sociology",
    "Juridical Sciences and Law,Science of Arts and Letters",
    "Life Sciences,Medical Sciences",
    "Life Sciences,Medical Sciences,Linguistics,Pedagogy,Psychology",
    "Life Sciences,Technological Sciences",
    "Life Sciences,Technological Sciences,Pedagogy",
    "Linguistics,Pedagogy",
    "Logic",
    "Mathematics",
    "Mathematics,Physics,Technological Sciences",
    "Mathematics,Economic Sciences,Pedagogy",
    "Mathematics,Economic Sciences,Sociology,Philosophy",
    "Mathematics,Life Sciences",
    "Mathematics,Technological Sciences",
    "Medical Sciences",
    "Medical Sciences,Anthropology",
    "Medical Sciences,Pedagogy",
    "Medical Sciences,Technological Sciences,Pedagogy",
    "Technological Sciences",
```

"Technological Sciences,Economic Sciences",
 "Technological Sciences,Economic Sciences,Pedagogy",
 "Technological Sciences,Geography",
 "Technological Sciences,Juridical Sciences and Law",
 "Technological Sciences,Linguistics,Pedagogy,Science of Arts and Letters,Philosophy",
 "Technological Sciences,Pedagogy,Philosophy",
 "Technological Sciences,Pedagogy",
 "Technological Sciences,Science of Arts and Letters",
 "Pedagogy",
 "Pedagogy,Psychology",
 "Pedagogy,Sociology",
 "Philosophy",
 "Physics",
 "Physics,Chemistry,Life Sciences,Earth and Space Sciences,Medical Sciences,Technological Sciences,Pedagogy,Ethics",
 "Physics,Technological Sciences",
 "Physics,Technological Sciences,Geography",
 "Psychology",
 "Psychology,Ethics",
 "Political Science",
 "Political Science,Sociology",
 "Science of Arts and Letters",
 "Sociology",
 "Life Sciences",
 "Linguistics & Humanities",
 "NA",
 "Natural & Formal Sciences",
 "Technology",
 "Social Sciences"))

Subject.field.1[Subject.field.1 == "Science of Arts and Letters"] <- "Linguistics & Humanities"
 Subject.field.1[Subject.field.1 == "Philosophy"] <- "Linguistics & Humanities"
 Subject.field.1[Subject.field.1 == "History"] <- "Linguistics & Humanities"
 Subject.field.1[Subject.field.1 == "Ethics"] <- "Linguistics & Humanities"
 Subject.field.1[Subject.field.1 == "Linguistics"] <- "Linguistics & Humanities"

Subject.field.1[Subject.field.1 == "Medical Sciences"] <- "Life Sciences"
 Subject.field.1[Subject.field.1 == "Psychology"] <- "Life Sciences"
 Subject.field.1[Subject.field.1 == "Life Sciences"] <- "Life Sciences"
 Subject.field.1[Subject.field.1 == "Life Sciences,Technological Sciences,Pedagogy"] <- "Life Sciences"
 Subject.field.1[Subject.field.1 == "Life Sciences,Medical Sciences,Linguistics,Pedagogy,Psychology"] <- "Life Sciences"
 Subject.field.1[Subject.field.1 == "Physics,Chemistry,Life Sciences,Earth and Space Sciences,Medical Sciences,Technological Sciences,Pedagogy,Ethics"] <- "Life Sciences"
 Subject.field.1[Subject.field.1 == "Medical Sciences,Technological Sciences,Pedagogy"] <- "Life Sciences"
 Subject.field.1[Subject.field.1 == "Life Sciences,Medical Sciences"] <- "Life Sciences"
 Subject.field.1[Subject.field.1 == "Psychology,Ethics"] <- "Life Sciences"
 Subject.field.1[Subject.field.1 == "Mathematics,Life Sciences"] <- "Life Sciences"

Subject.field.1[Subject.field.1 == "Medical Sciences,Pedagogy"] <- "Life Sciences"

Subject.field.1[Subject.field.1 == "Earth and Space Sciences,Agricultural Sciences"] <- "Natural & Formal Sciences"

Subject.field.1[Subject.field.1 == "Physics"] <- "Natural & Formal Sciences"

Subject.field.1[Subject.field.1 == "Geography"] <- "Natural & Formal Sciences"

Subject.field.1[Subject.field.1 == "Earth and Space Sciences"] <- "Natural & Formal Sciences"

Subject.field.1[Subject.field.1 == "Chemistry"] <- "Natural & Formal Sciences"

Subject.field.1[Subject.field.1 == "Astronomy and astrophysics"] <- "Natural & Formal Sciences"

Subject.field.1[Subject.field.1 == "Earth and Space Sciences,Technological Sciences"] <- "Natural & Formal Sciences"

Subject.field.1[Subject.field.1 == "Chemistry,Life Sciences,Agricultural Sciences"] <- "Natural & Formal Sciences"

Subject.field.1[Subject.field.1 == "Earth and Space Sciences,Agricultural Sciences,Geography"] <- "Natural & Formal Sciences"

Subject.field.1[Subject.field.1 == "Mathematics"] <- "Natural & Formal Sciences"

Subject.field.1[Subject.field.1 == "Logic"] <- "Natural & Formal Sciences"

Subject.field.1[Subject.field.1 == "Pedagogy"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Political Science"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Sociology"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Anthropology"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Pedagogy"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Economic Sciences"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Juridical Sciences and Law"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Demographics"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Juridical Sciences and Law,Psychology"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Pedagogy,Psychology"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Mathematics,Economic Sciences,Sociology,Philosophy"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Technological Sciences,Pedagogy"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Pedagogy,Sociology"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Anthropology,History"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Anthropology,Pedagogy"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Political Science,Sociology"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Economic Sciences,Pedagogy"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Demographics,Sociology"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Juridical Sciences and Law,Science of Arts and Letters"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Mathematics,Economic Sciences,Pedagogy"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Juridical Sciences and Law,Political Science,Sociology"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Economic Sciences,History"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Technological Sciences,Juridical Sciences and Law"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Economic Sciences,Ethics"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Economic Sciences,Pedagogy,Psychology"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Technological Sciences,Economic Sciences,Pedagogy"] <- "Social Sciences"

```
Subject.field.1[Subject.field.1 == "Technological Sciences,Linguistics,Pedagogy,Science of Arts
and Letters,Philosophy"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Linguistics,Pedagogy"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Economic Sciences,Juridical Sciences and Law"] <- "Social
Sciences"
Subject.field.1[Subject.field.1 == "Medical Sciences,Anthropology"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Technological Sciences"] <- "Technology"
Subject.field.1[Subject.field.1 == "Technological Sciences,Science of Arts and Letters"] <-
"Technology"
Subject.field.1[Subject.field.1 == "Agricultural Sciences"] <- "Technology"
Subject.field.1[Subject.field.1 == "Technological Sciences,Economic Sciences"] <- "Technology"
Subject.field.1[Subject.field.1 == "Technological Sciences,Pedagogy,Philosophy"] <-
"Technology"
Subject.field.1[Subject.field.1 == "Technological Sciences,Economic Sciences"] <- "Technology"
Subject.field.1[Subject.field.1 == "Mathematics,Physics,Technological Sciences"] <- "Technology"
Subject.field.1[Subject.field.1 == "Mathematics,Technological Sciences"] <- "Technology"
Subject.field.1[Subject.field.1 == "Physics,Technological Sciences,Geography"] <- "Technology"
Subject.field.1[Subject.field.1 == "Technological Sciences,Science of Arts and Letters"] <-
"Technology"
Subject.field.1[Subject.field.1 == "Physics,Technological Sciences"] <- "Technology"
Subject.field.1[Subject.field.1 == "Life Sciences,Technological Sciences"] <- "Technology"
Subject.field.1[Subject.field.1 == "Technological Sciences,Geography"] <- "Technology"

Subjectfield_agg <- factor(Subject.field.1,
  levels = c("Linguistics & Humanities",
    "Life Sciences",
    "Natural & Formal Sciences",
    "Technology",
    "Social Sciences"))

# 2.5. Esperienza d'insegnamento

Teach.exp <- dati_two_universities$Teach_exp
Teach.exp <- factor(Teach.exp,
  levels = c("Less than 3","3-4","5-10","10-15",
    "More than 15"))

# 2.6. Esperienza di ricerca

Research.exp <- dati_two_universities$Research_exp
Research.exp <- factor(Research.exp,
  levels = c("No experience","Less than 3",
    "3-5","6-10","11-15",
    "More than 15"))

# 2.7 Variabili quantitative

EDMQ1 <- as.numeric(dati_two_universities$EDMQ1)
EDMQ2 <- as.numeric(dati_two_universities$EDMQ2)
EDMQ3 <- as.numeric(dati_two_universities$EDMQ3)
```

```
EDMQ4 <- as.numeric(dati_two_universities$EDMQ4)
EDMQ5 <- as.numeric(dati_two_universities$EDMQ5)
EDMQ6 <- as.numeric(dati_two_universities$EDMQ6)
EDMQ7 <- as.numeric(dati_two_universities$EDMQ7)
EDMQ8 <- as.numeric(dati_two_universities$EDMQ8)
DER1 <- as.numeric(dati_two_universities$DER1)
DER2 <- as.numeric(dati_two_universities$DER2)
DER3 <- as.numeric(dati_two_universities$DER3)
DER4 <- as.numeric(dati_two_universities$DER4)
SED1 <- as.numeric(dati_two_universities$SED1)
SED2 <- as.numeric(dati_two_universities$SED2)
SED3 <- as.numeric(dati_two_universities$SED3)
DTL1 <- as.numeric(dati_two_universities$DTL1)
DTL2 <- as.numeric(dati_two_universities$DTL2)
DTL3 <- as.numeric(dati_two_universities$DTL3)
DTL4 <- as.numeric(dati_two_universities$DTL4)
DTL5 <- as.numeric(dati_two_universities$DTL5)
DA1 <- as.numeric(dati_two_universities$DA1)
DA10 <- as.numeric(dati_two_universities$DA10)
DA2 <- as.numeric(dati_two_universities$DA2)
DA3 <- as.numeric(dati_two_universities$DA3)
DA4 <- as.numeric(dati_two_universities$DA4)
DA5 <- as.numeric(dati_two_universities$DA5)
DA6 <- as.numeric(dati_two_universities$DA6)
DA7 <- as.numeric(dati_two_universities$DA7)
DA8 <- as.numeric(dati_two_universities$DA8)
DA9 <- as.numeric(dati_two_universities$DA9)
SDL1 <- as.numeric(dati_two_universities$SDL1)
SDL2 <- as.numeric(dati_two_universities$SDL2)
SDL3 <- as.numeric(dati_two_universities$SDL3)
SDL4 <- as.numeric(dati_two_universities$SDL4)
SDL5 <- as.numeric(dati_two_universities$SDL5)

# 3. Costruisco i data frame
# 3.1. Costruisco il data frame per tutte le variabili

dati_two_universities <- data.frame(Gender, Age, Country, Subjectfield_agg, Teach.exp,
Research.exp,
    EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5, EDMQ6, EDMQ7, EDMQ8,
    DER1, DER2, DER3, DER4,
    SED1, SED2, SED3,
    DTL1, DTL2, DTL3, DTL4, DTL5,
    DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,
    SDL1, SDL2, SDL3, SDL4, SDL5)
dati_two_universities <- na.omit(dati_two_universities)

# 3.2. Costruisco il dataframe per le variabili quantitative

aldepvars <- data.frame(EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5, EDMQ6, EDMQ7,
    EDMQ8,
    DER1, DER2, DER3, DER4,
```

```
SED1, SED2, SED3,
DTL1, DTL2, DTL3, DTL4, DTL5,
DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,
SDL1, SDL2, SDL3, SDL4, SDL5)
dati_analisi <- na.omit(alldepvars)

# 4. Identifico e rimuovo gli outlier

md <- mahalanobis((dati_analisi), center = colMeans(dati_analisi), cov = cov(dati_analisi))
alpha <- (.001)
cutoff <- (qchisq(p = 1 - alpha, df = ncol(dati_analisi)))
names_outliers_MH <- which(md > cutoff)

dati_two_universities_so <- dati_two_universities[-names_outliers_MH, ]
alldepvars_so <- dati_two_universities_so[,-c(1:6)]

# 5. Eseguo le analisi
# 5.1. Eseguo l'analisi univariata

summary(dati_two_universities_so$Gender)
summary(dati_two_universities_so$Age)
summary(dati_two_universities_so$Country)
summary(dati_two_universities_so$Subjectfield_agg)
summary(dati_two_universities_so$Teach.exp)
summary(dati_two_universities_so$Research.exp)

tabella_stat_quant <- describe(alldepvars_so)
tabella_stat_quant <- nice_table(tabella_stat_quant)
save_as_docx(tabella_stat_quant, path = "C:/Users/pc/Documents/UNIV/TESI DI
LAUREA/Dataset/tabella_stat_quant.docx")

# 5.2. Analizzo le variabili per gruppo

EDMQ <- c(EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5, EDMQ6, EDMQ7, EDMQ8)
DER <- c(DER1, DER2, DER3, DER4)
SED <- c(SED1, SED2, SED3)
DTL <- c(DTL1, DTL2, DTL3, DTL4, DTL5)
DA <- c(DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10)
SDL <- c(SDL1, SDL2, SDL3, SDL4, SDL5)

tab_EDMQ <- describe(EDMQ)
tab_DER <- describe(DER)
tab_SED <- describe(SED)
tab_DTL <- describe(DTL)
tab_DA <- describe(DA)
tab_SDL <- describe(SDL)

tab_EDMQ <- nice_table(tab_EDMQ)
save_as_docx(tab_EDMQ, path = "C:/Users/pc/Documents/UNIV/TESI DI
LAUREA/Dataset/tab_EDMQ.docx")
```

```
tab_DER <- nice_table(tab_DER)
save_as_docx(tab_DER, path = "C:/Users/pc/Documents/UNIV/TESI DI
LAUREA/Dataset/tab_DER.docx")
```

```
tab_SED <- nice_table(tab_SED)
save_as_docx(tab_SED, path = "C:/Users/pc/Documents/UNIV/TESI DI
LAUREA/Dataset/tab_SED.docx")
```

```
tab_DTL <- nice_table(tab_DTL)
save_as_docx(tab_DTL, path = "C:/Users/pc/Documents/UNIV/TESI DI
LAUREA/Dataset/tab_DTL.docx")
```

```
tab_DA <- nice_table(tab_DA)
save_as_docx(tab_DA, path = "C:/Users/pc/Documents/UNIV/TESI DI
LAUREA/Dataset/tab_DA.docx")
```

```
tab_SDL <- nice_table(tab_SDL)
save_as_docx(tab_SDL, path = "C:/Users/pc/Documents/UNIV/TESI DI
LAUREA/Dataset/tab_SDL.docx")
```

6. Eseguo l'analisi bivariata

```
dati_bivariata_gender <- dati_two_universities_so %>%
  group_by(Gender) %>%
  summarise(EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5,
            EDMQ6, EDMQ7, EDMQ8,
            DER1, DER2, DER3, DER4,
            SED1, SED2, SED3,
            DTL1, DTL2, DTL3, DTL4, DTL5,
            DA1, DA2, DA3, DA4, DA5,
            DA6, DA7, DA8, DA9, DA10,
            SDL1, SDL2, SDL3, SDL4, SDL5)
```

```
gender_tab <- dfSummary(dati_bivariata_gender)
gender_tab
```

```
dati_bivariata_age <- dati_two_universities_so %>%
  group_by(Age) %>%
  summarise(EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5,
            EDMQ6, EDMQ7, EDMQ8,
            DER1, DER2, DER3, DER4,
            SED1, SED2, SED3,
            DTL1, DTL2, DTL3, DTL4, DTL5,
            DA1, DA2, DA3, DA4, DA5,
            DA6, DA7, DA8, DA9, DA10,
            SDL1, SDL2, SDL3, SDL4, SDL5)
```

```
age_tab <- dfSummary(dati_bivariata_age)
age_tab
```

```
dati_bivariata_country <- dati_two_universities_so %>%
```



```
group_by(Country) %>%  
summarise(EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5,  
          EDMQ6, EDMQ7, EDMQ8,  
          DER1, DER2, DER3, DER4,  
          SED1, SED2, SED3,  
          DTL1, DTL2, DTL3, DTL4, DTL5,  
          DA1, DA2, DA3, DA4, DA5,  
          DA6, DA7, DA8, DA9, DA10,  
          SDL1, SDL2, SDL3, SDL4, SDL5)
```

```
country_tab <- dfSummary(dati_bivariata_country)  
country_tab
```

```
dati_bivariata_sf <- dati_two_universities_so %>%  
group_by(Subjectfield_agg) %>%  
summarise(EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5,  
          EDMQ6, EDMQ7, EDMQ8,  
          DER1, DER2, DER3, DER4,  
          SED1, SED2, SED3,  
          DTL1, DTL2, DTL3, DTL4, DTL5,  
          DA1, DA2, DA3, DA4, DA5,  
          DA6, DA7, DA8, DA9, DA10,  
          SDL1, SDL2, SDL3, SDL4, SDL5)
```

```
sf_tab <- dfSummary(dati_bivariata_sf)  
View(sf_tab)
```

```
dati_bivariata_te <- dati_two_universities_so %>%  
group_by(Teach.exp) %>%  
summarise(EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5,  
          EDMQ6, EDMQ7, EDMQ8,  
          DER1, DER2, DER3, DER4,  
          SED1, SED2, SED3,  
          DTL1, DTL2, DTL3, DTL4, DTL5,  
          DA1, DA2, DA3, DA4, DA5,  
          DA6, DA7, DA8, DA9, DA10,  
          SDL1, SDL2, SDL3, SDL4, SDL5)
```

```
sf_tab <- dfSummary(dati_bivariata_te)  
View(sf_tab)
```

```
dati_bivariata_re <- dati_two_universities_so %>%  
group_by(Research.exp) %>%  
summarise(EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5,  
          EDMQ6, EDMQ7, EDMQ8,  
          DER1, DER2, DER3, DER4,  
          SED1, SED2, SED3,  
          DTL1, DTL2, DTL3, DTL4, DTL5,  
          DA1, DA2, DA3, DA4, DA5,  
          DA6, DA7, DA8, DA9, DA10,  
          SDL1, SDL2, SDL3, SDL4, SDL5)
```

```
re_tab <- dfSummary(dati_bivariata_re)
re_tab

# 7. Eseguo una One-way ANOVA per cercare relazioni tra le variabili
# 7.1. Raccolgo i dati per variabile

dati_ragg <- dati_two_universities_so %>%
  gather(key = "variable", value = "value", EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5,
    EDMQ6, EDMQ7, EDMQ8,
    DER1, DER2, DER3, DER4,
    SED1, SED2, SED3,
    DTL1, DTL2, DTL3, DTL4, DTL5,
    DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,
    SDL1, SDL2, SDL3, SDL4, SDL5) %>%
  group_by(variable)

# 7.2. Eseguo i test

aov_dati_gender <- dati_ragg %>% welch_anova_test(value ~ Gender)
print(aov_dati_gender, n = 35)
tab_aov_gender <- data.frame(aov_dati_gender)
View(tab_aov_gender)

aov_dati_age <- dati_ragg %>% anova_test(value ~ Age)
print(aov_dati_age, n = 35)
tab_aov_age <- data.frame(aov_dati_age)
View(tab_aov_age)

aov_dati_country <- dati_ragg %>% welch_anova_test(value ~ Country)
print(aov_dati_country, n = 35)
tab_aov_country <- data.frame(aov_dati_country)
View(tab_aov_country)

aov_dati_sf <- dati_ragg %>% welch_anova_test(value ~ Subjectfield_agg)
print(aov_dati_sf, n = 35)
tab_aov_sf <- data.frame(aov_dati_sf)
View(tab_aov_sf)

aov_dati_te <- dati_ragg %>% welch_anova_test(value ~ Teach.exp)
print(aov_dati_te, n = 35)
tab_aov_te <- data.frame(aov_dati_te)
View(tab_aov_te)

aov_dati_re <- dati_ragg %>% welch_anova_test(value ~ Research.exp)
print(aov_dati_re, n = 35)
tab_aov_re <- data.frame(aov_dati_re)
View(tab_aov_re)

# 8. Realizzo la matrice di correlazione
```

```
CorrMat <- cor(alldepvars_so)
cor_tab <- as.data.frame(CorrMat)
View(cor_tab)
```

```
cor_tab <- nice_table(cor_tab)
save_as_docx(cor_tab, path = "C:/Users/pc/Documents/UNIV/TESI DI
LAUREA/Dataset/cor_tab.docx")
```

8.1. Disegno il grafico della matrice di correlazione

```
windows()
Grafico_CorrMat <- corrplot(CorrMat,
  method = "circle",
  type = "lower",
  is.corr = FALSE,
  order = "original",
  diag = TRUE,
  outline = FALSE,
  plotCI = NULL)
```

9. Boxplot

```
boxplot_gender <- ggboxplot(dati_two_universities_so, x = "Gender", y = c("DA1", "DA2",
"DA3", "DA4", "EDMQ5", "SDL1"),
  merge = TRUE, palette = "rainbow", fill = "grey", repel = TRUE, xlab = FALSE,
ylab = FALSE, notch = TRUE)
facet(boxplot_gender, facet.by = "Gender", scales = "free")
```

```
windows()
boxplot_age <- ggboxplot(dati_two_universities_so, x = "Age", y = c("DA6", "DA8", "DER1",
"DER2", "DER3", "DER4", "EDMQ1", "EDMQ2", "EDMQ3", "EDMQ4", "SDL2"),
  merge = TRUE, palette = "rainbow", fill = "grey", repel = TRUE, xlab = FALSE,
ylab = FALSE)
facet(boxplot_age, facet.by = "Age", scales = "free")
```

```
windows()
boxplot_country <- ggboxplot(dati_two_universities_so, x = "Country", y = c("DA1", "DA2",
"DA6", "DA7", "DA9", "DER2", "DER3", "DTL2", "DTL5", "EDMQ2", "EDMQ3", "EDMQ4",
"EDMQ5", "EDMQ6", "SDL1", "SDL5", "SED2"),
  merge = TRUE, palette = "rainbow", fill = "grey", repel = TRUE, xlab = FALSE,
ylab = FALSE)
facet(boxplot_country, facet.by = "Country", scales = "free", ncol = 1, nrow = 2)
```

```
windows()
boxplot_sf <- ggboxplot(dati_two_universities_so, x = "Subjectfield_agg", y = c("DA1", "DA2",
"DA6", "DA7", "DA9", "DER2", "DER3", "DTL2", "DTL5", "EDMQ2", "EDMQ3", "EDMQ4",
"EDMQ5", "EDMQ6", "SDL1", "SDL5", "SED2"),
  merge = TRUE, palette = "rainbow", fill = "grey", repel = TRUE, xlab = FALSE,
ylab = FALSE)
facet(boxplot_sf, facet.by = "Subjectfield_agg", scales = "free", ncol = 2, nrow = 3)
```

```
windows()
boxplot_te <- ggboxplot(dati_two_universities_so, x = "Teach.exp", y = c("DA7", "EDMQ1",
"EDMQ2", "EDMQ3", "EDMQ4"),
merge = TRUE, palette = "rainbow", fill = "grey", repel = TRUE, xlab = FALSE,
ylab = FALSE)
facet(boxplot_te, facet.by = "Teach.exp", scales = "free", ncol = 2, nrow = 3)
```

```
windows()
boxplot_re <- ggboxplot(dati_two_universities_so, x = "Research.exp", y = c("DA6", "DA7",
"DA9", "DER2", "DER3", "EDMQ2", "EDMQ3", "EDMQ4", "EDMQ5"),
merge = TRUE, palette = "rainbow", fill = "grey", repel = TRUE, xlab = FALSE,
ylab = FALSE)
facet(boxplot_re, facet.by = "Research.exp", scales = "free", ncol = 2, nrow = 3)
```

8.1.2. Script per MANOVA

```
# Script per MANOVA

# 1. Preparo il workspace

# install.packages("corrplot")
# install.packages("heplots")
# install.packages("ggpubr")
# install.packages("psych")
# install.packages("rstatix")
# install.packages("tidyverse")

library(corrplot)
library(ggpubr)
library(heplots)
library(performance)
library(psych)
library(rstatix)
library(tidyverse)

setwd("C:/Users/pc/Documents/UNIV/TESI DI LAUREA/Dataset")
dati_two_universities <- read.table("DataPractices_TwoUniversities.csv",
                                   header = TRUE, sep = ",")

dati_two_universities[dati_two_universities == "0,00"] <- "0"
dati_two_universities[dati_two_universities == "1,00"] <- "1"
dati_two_universities[dati_two_universities == "2,00"] <- "2"
dati_two_universities[dati_two_universities == "3,00"] <- "3"
dati_two_universities[dati_two_universities == "4,00"] <- "4"
dati_two_universities[dati_two_universities == "5,00"] <- "5"
View(dati_two_universities)

# 2. Converto le variabili
# 2.1. Genere

Gender <- dati_two_universities$Gender
Gender <- factor(Gender,
                 levels = c("Male","Female"))

# 2.2. Età

Age <- dati_two_universities$Age
Age <- factor(Age,
              levels = c("25-34","35-44",
                         "45-54","More than 55"))

# 2.3. Nazione
```

```
Country <- dati_two_universities$Country  
Country <- factor(Country,  
  levels = c("Italy","Spain"))
```

2.4. Area disciplinare

```
Subject.field.1 <- factor(dati_two_universities$Subject.field.1,  
  levels = c("Agricultural Sciences",  
    "Anthropology",  
    "Anthropology,History",  
    "Anthropology,Pedagogy",  
    "Astronomy and astrophysics",  
    "Chemistry",  
    "Chemistry,Life Sciences,Agricultural Sciences",  
    "Demographics",  
    "Demographics,Sociology",  
    "Earth and Space Sciences",  
    "Earth and Space Sciences,Agricultural Sciences",  
    "Earth and Space Sciences,Technological Sciences",  
    "Earth and Space Sciences,Agricultural Sciences,Geography",  
    "Economic Sciences",  
    "Economic Sciences,Ethics",  
    "Economic Sciences,History",  
    "Economic Sciences,Juridical Sciences and Law",  
    "Economic Sciences,Pedagogy",  
    "Economic Sciences,Pedagogy,Psychology",  
    "Ethics",  
    "Geography",  
    "History",  
    "Juridical Sciences and Law",  
    "Juridical Sciences and Law,Psychology",  
    "Juridical Sciences and Law,Political Science,Sociology",  
    "Juridical Sciences and Law,Science of Arts and Letters",  
    "Life Sciences,Medical Sciences",  
    "Life Sciences,Medical Sciences,Linguistics,Pedagogy,Psychology",  
    "Life Sciences,Technological Sciences",  
    "Life Sciences,Technological Sciences,Pedagogy",  
    "Linguistics,Pedagogy",  
    "Logic",  
    "Mathematics",  
    "Mathematics,Physics,Technological Sciences",  
    "Mathematics,Economic Sciences,Pedagogy",  
    "Mathematics,Economic Sciences,Sociology,Philosophy",  
    "Mathematics,Life Sciences",  
    "Mathematics,Technological Sciences",  
    "Medical Sciences",  
    "Medical Sciences,Anthropology",  
    "Medical Sciences,Pedagogy",  
    "Medical Sciences,Technological Sciences,Pedagogy",  
    "Technological Sciences",  
    "Technological Sciences,Economic Sciences",
```

"Technological Sciences,Economic Sciences,Pedagogy",
"Technological Sciences,Geography",
"Technological Sciences,Juridical Sciences and Law",
"Technological Sciences,Linguistics,Pedagogy,Science of Arts and
Letters,Philosophy",
"Technological Sciences,Pedagogy,Philosophy",
"Technological Sciences,Pedagogy",
"Technological Sciences,Science of Arts and Letters",
"Pedagogy",
"Pedagogy,Psychology",
"Pedagogy,Sociology",
"Philosophy",
"Physics",
"Physics,Chemistry,Life Sciences,Earth and Space Sciences,Medical
Sciences,Technological Sciences,Pedagogy,Ethics",
"Physics,Technological Sciences",
"Physics,Technological Sciences,Geography",
"Psychology",
"Psychology,Ethics",
"Political Science",
"Political Science,Sociology",
"Science of Arts and Letters",
"Sociology",
"Life Sciences",
"Linguistics & Humanities",
"NA",
"Natural & Formal Sciences",
"Technology",
"Social Sciences"))

Subject.field.1[Subject.field.1 == "Science of Arts and Letters"] <- "Linguistics & Humanities"
Subject.field.1[Subject.field.1 == "Philosophy"] <- "Linguistics & Humanities"
Subject.field.1[Subject.field.1 == "History"] <- "Linguistics & Humanities"
Subject.field.1[Subject.field.1 == "Ethics"] <- "Linguistics & Humanities"
Subject.field.1[Subject.field.1 == "Linguistics"] <- "Linguistics & Humanities"

Subject.field.1[Subject.field.1 == "Medical Sciences"] <- "Life Sciences"
Subject.field.1[Subject.field.1 == "Psychology"] <- "Life Sciences"
Subject.field.1[Subject.field.1 == "Life Sciences"] <- "Life Sciences"
Subject.field.1[Subject.field.1 == "Life Sciences,Technological Sciences,Pedagogy"] <- "Life
Sciences"
Subject.field.1[Subject.field.1 == "Life Sciences,Medical
Sciences,Linguistics,Pedagogy,Psychology"] <- "Life Sciences"
Subject.field.1[Subject.field.1 == "Physics,Chemistry,Life Sciences,Earth and Space
Sciences,Medical Sciences,Technological Sciences,Pedagogy,Ethics"] <- "Life Sciences"
Subject.field.1[Subject.field.1 == "Medical Sciences,Technological Sciences,Pedagogy"] <- "Life
Sciences"
Subject.field.1[Subject.field.1 == "Life Sciences,Medical Sciences"] <- "Life Sciences"
Subject.field.1[Subject.field.1 == "Psychology,Ethics"] <- "Life Sciences"
Subject.field.1[Subject.field.1 == "Mathematics,Life Sciences"] <- "Life Sciences"
Subject.field.1[Subject.field.1 == "Medical Sciences,Pedagogy"] <- "Life Sciences"

Subject.field.1[Subject.field.1 == "Earth and Space Sciences,Agricultural Sciences"] <- "Natural & Formal Sciences"
Subject.field.1[Subject.field.1 == "Physics"] <- "Natural & Formal Sciences"
Subject.field.1[Subject.field.1 == "Geography"] <- "Natural & Formal Sciences"
Subject.field.1[Subject.field.1 == "Earth and Space Sciences"] <- "Natural & Formal Sciences"
Subject.field.1[Subject.field.1 == "Chemistry"] <- "Natural & Formal Sciences"
Subject.field.1[Subject.field.1 == "Astronomy and astrophysics"] <- "Natural & Formal Sciences"
Subject.field.1[Subject.field.1 == "Earth and Space Sciences,Technological Sciences"] <- "Natural & Formal Sciences"
Subject.field.1[Subject.field.1 == "Chemistry,Life Sciences,Agricultural Sciences"] <- "Natural & Formal Sciences"
Subject.field.1[Subject.field.1 == "Earth and Space Sciences,Agricultural Sciences,Geography"] <- "Natural & Formal Sciences"
Subject.field.1[Subject.field.1 == "Mathematics"] <- "Natural & Formal Sciences"
Subject.field.1[Subject.field.1 == "Logic"] <- "Natural & Formal Sciences"

Subject.field.1[Subject.field.1 == "Pedagogy"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Political Science"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Sociology"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Anthropology"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Pedagogy"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Economic Sciences"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Juridical Sciences and Law"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Demographics"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Juridical Sciences and Law,Psychology"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Pedagogy,Psychology"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Mathematics,Economic Sciences,Sociology,Philosophy"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Technological Sciences,Pedagogy"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Pedagogy,Sociology"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Anthropology,History"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Anthropology,Pedagogy"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Political Science,Sociology"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Economic Sciences,Pedagogy"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Demographics,Sociology"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Juridical Sciences and Law,Science of Arts and Letters"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Mathematics,Economic Sciences,Pedagogy"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Juridical Sciences and Law,Political Science,Sociology"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Economic Sciences,History"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Technological Sciences,Juridical Sciences and Law"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Economic Sciences,Ethics"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Economic Sciences,Pedagogy,Psychology"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Technological Sciences,Economic Sciences,Pedagogy"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Technological Sciences,Linguistics,Pedagogy,Science of Arts and Letters,Philosophy"] <- "Social Sciences"


```
Subject.field.1[Subject.field.1 == "Linguistics,Pedagogy"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Economic Sciences,Juridical Sciences and Law"] <- "Social
Sciences"
Subject.field.1[Subject.field.1 == "Medical Sciences,Anthropology"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Technological Sciences"] <- "Technology"
Subject.field.1[Subject.field.1 == "Technological Sciences,Science of Arts and Letters"] <-
"Technology"
Subject.field.1[Subject.field.1 == "Agricultural Sciences"] <- "Technology"
Subject.field.1[Subject.field.1 == "Technological Sciences,Economic Sciences"] <- "Technology"
Subject.field.1[Subject.field.1 == "Technological Sciences,Pedagogy,Philosophy"] <-
"Technology"
Subject.field.1[Subject.field.1 == "Technological Sciences,Economic Sciences"] <- "Technology"
Subject.field.1[Subject.field.1 == "Mathematics,Physics,Technological Sciences"] <- "Technology"
Subject.field.1[Subject.field.1 == "Mathematics,Technological Sciences"] <- "Technology"
Subject.field.1[Subject.field.1 == "Physics,Technological Sciences,Geography"] <- "Technology"
Subject.field.1[Subject.field.1 == "Technological Sciences,Science of Arts and Letters"] <-
"Technology"
Subject.field.1[Subject.field.1 == "Physics,Technological Sciences"] <- "Technology"
Subject.field.1[Subject.field.1 == "Life Sciences,Technological Sciences"] <- "Technology"
Subject.field.1[Subject.field.1 == "Technological Sciences,Geography"] <- "Technology"

Subjectfield_agg <- factor(Subject.field.1,
  levels = c("Linguistics & Humanities",
            "Life Sciences",
            "Natural & Formal Sciences",
            "Technology",
            "Social Sciences"))

# 2.5. Esperienza d'insegnamento

Teach.exp <- dati_two_universities$Teach_exp
Teach.exp <- factor(Teach.exp,
  levels = c("Less than 3","3-4","5-10","10-15",
            "More than 15"))

# 2.6. Esperienza di ricerca

Research.exp <- dati_two_universities$Research_exp
Research.exp <- factor(Research.exp,
  levels = c("No experience","Less than 3",
            "3-5","6-10","11-15",
            "More than 15"))

# 2.7. Converto le variabili

EDMQ1 <- as.numeric(dati_two_universities$EDMQ1)
EDMQ2 <- as.numeric(dati_two_universities$EDMQ2)
EDMQ3 <- as.numeric(dati_two_universities$EDMQ3)
EDMQ4 <- as.numeric(dati_two_universities$EDMQ4)
EDMQ5 <- as.numeric(dati_two_universities$EDMQ5)
```

```
EDMQ6 <- as.numeric(dati_two_universities$EDMQ6)
EDMQ7 <- as.numeric(dati_two_universities$EDMQ7)
EDMQ8 <- as.numeric(dati_two_universities$EDMQ8)
DER1 <- as.numeric(dati_two_universities$DER1)
DER2 <- as.numeric(dati_two_universities$DER2)
DER3 <- as.numeric(dati_two_universities$DER3)
DER4 <- as.numeric(dati_two_universities$DER4)
SED1 <- as.numeric(dati_two_universities$SED1)
SED2 <- as.numeric(dati_two_universities$SED2)
SED3 <- as.numeric(dati_two_universities$SED3)
DTL1 <- as.numeric(dati_two_universities$DTL1)
DTL2 <- as.numeric(dati_two_universities$DTL2)
DTL3 <- as.numeric(dati_two_universities$DTL3)
DTL4 <- as.numeric(dati_two_universities$DTL4)
DTL5 <- as.numeric(dati_two_universities$DTL5)
DA1 <- as.numeric(dati_two_universities$DA1)
DA10 <- as.numeric(dati_two_universities$DA10)
DA2 <- as.numeric(dati_two_universities$DA2)
DA3 <- as.numeric(dati_two_universities$DA3)
DA4 <- as.numeric(dati_two_universities$DA4)
DA5 <- as.numeric(dati_two_universities$DA5)
DA6 <- as.numeric(dati_two_universities$DA6)
DA7 <- as.numeric(dati_two_universities$DA7)
DA8 <- as.numeric(dati_two_universities$DA8)
DA9 <- as.numeric(dati_two_universities$DA9)
SDL1 <- as.numeric(dati_two_universities$SDL1)
SDL2 <- as.numeric(dati_two_universities$SDL2)
SDL3 <- as.numeric(dati_two_universities$SDL3)
SDL4 <- as.numeric(dati_two_universities$SDL4)
SDL5 <- as.numeric(dati_two_universities$SDL5)
```

3. Costruisco i data frame

3.1. Costruisco il data frame per tutte le variabili

```
dati_MANOVA <- data.frame(Gender, Age, Country, Subjectfield_agg, Teach.exp, Research.exp,
  EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5, EDMQ6, EDMQ7, EDMQ8,
  DER1, DER2, DER3, DER4,
  SED1, SED2, SED3,
  DTL1, DTL2, DTL3, DTL4, DTL5,
  DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,
  SDL1, SDL2, SDL3, SDL4, SDL5)
dati_MANOVA <- na.omit(dati_MANOVA)
```

3.2. Costruisco il data frame per le variabili dipendenti

```
alldpvars <- data.frame(EDMQ1,EDMQ2,EDMQ3,EDMQ4,EDMQ5,EDMQ6,EDMQ7,EDMQ8,
  DER1, DER2, DER3, DER4,
  SED1, SED2, SED3,
  DTL1, DTL2, DTL3, DTL4, DTL5,
  DA1, DA2, DA3, DA4,DA5,DA6,DA7,DA8,DA9,DA10,
  SDL1, SDL2, SDL3, SDL4, SDL5)
```

```
alldepvars <- na.omit(alldepvars)
```

```
# 4. Controllo le assunzioni della MANOVA
```

```
# 4.1. Controllo la presenza di outlier univariati
```

```
out_EDMQ1 <- table(is_outlier(EDMQ1))
out_EDMQ2 <- table(is_outlier(EDMQ2))
out_EDMQ3 <- table(is_outlier(EDMQ3))
out_EDMQ4 <- table(is_outlier(EDMQ4))
out_EDMQ5 <- table(is_outlier(EDMQ5))
out_EDMQ6 <- table(is_outlier(EDMQ6))
out_EDMQ7 <- table(is_outlier(EDMQ7))
out_EDMQ8 <- table(is_outlier(EDMQ8))
out_DER1 <- table(is_outlier(DER1))
out_DER2 <- table(is_outlier(DER2))
out_DER3 <- table(is_outlier(DER3))
out_DER4 <- table(is_outlier(DER4))
out_SED1 <- table(is_outlier(SED1))
out_SED2 <- table(is_outlier(SED2))
out_SED3 <- table(is_outlier(SED3))
out_DTL1 <- table(is_outlier(DTL1))
out_DTL2 <- table(is_outlier(DTL2))
out_DTL3 <- table(is_outlier(DTL3))
out_DTL4 <- table(is_outlier(DTL4))
out_DTL5 <- table(is_outlier(DTL5))
out_DA1 <- table(is_outlier(DA1))
out_DA10 <- table(is_outlier(DA10))
out_DA2 <- table(is_outlier(DA2))
out_DA3 <- table(is_outlier(DA3))
out_DA4 <- table(is_outlier(DA4))
out_DA5 <- table(is_outlier(DA5))
out_DA6 <- table(is_outlier(DA6))
out_DA7 <- table(is_outlier(DA7))
out_DA8 <- table(is_outlier(DA8))
out_DA9 <- table(is_outlier(DA9))
out_SDL1 <- table(is_outlier(SDL1))
out_SDL2 <- table(is_outlier(SDL2))
out_SDL3 <- table(is_outlier(SDL3))
out_SDL4 <- table(is_outlier(SDL4))
out_SDL5 <- table(is_outlier(SDL5))
```

```
out_EDMQ1
out_EDMQ2
out_EDMQ3
out_EDMQ4
out_EDMQ5
out_EDMQ6
out_EDMQ7
out_EDMQ8
out_DER1
out_DER2
```

```
out_DER3
out_DER4
out_SED1
out_SED2
out_SED3
out_DTL1
out_DTL2
out_DTL3
out_DTL4
out_DTL5
out_DA1
out_DA10
out_DA2
out_DA3
out_DA4
out_DA5
out_DA6
out_DA7
out_DA8
out_DA9
out_SDL1
out_SDL2
out_SDL3
out_SDL4
out_SDL5
```

4.2. Outlier multivariati

```
multivariate_outliers <- dati_MANOVA %>%
  mahalanobis_distance() %>%
  filter(is.outlier == TRUE) %>%
  as.data.frame()
```

```
View(multivariate_outliers)
```

5. Identifico e rimuovo gli outlier

```
md <- mahalanobis((alldepvars), center = colMeans(alldepvars), cov = cov(alldepvars))
alpha <- (.001)
cutoff <- (qchisq(p = 1 - alpha, df = ncol(alldepvars)))
names_outliers_MH <- which(md > cutoff)
```

```
dati_MANOVA_so <- dati_MANOVA[-names_outliers_MH, ]
alldepvars_so <- dati_MANOVA_so[,-c(1:6)]
```

6. Testo l'assunto di normalità

6.1. Testo l'assunzione di normalità univariata

```
Shapiro_test <- alldepvars_so %>% shapiro_test(EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5,
EDMQ6, EDMQ7, EDMQ8,
  DER1, DER2, DER3, DER4,
```

```
SED1, SED2, SED3,  
DTL1, DTL2, DTL3, DTL4, DTL5,  
DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,  
SDL1, SDL2, SDL3, SDL4, SDL5)
```

```
View(Shapiro_test)
```

```
tab_norm_uni <- nice_table(Shapiro_test)  
save_as_docx(tab_norm_uni, path = "C:/Users/pc/Documents/UNIV/TESI DI  
LAUREA/Dataset/tab_norm_uni.docx")
```

```
# 6.2. Test l'assunzione di normalità multivariata
```

```
mardia(alldepvars_so)
```

```
alldepvars_so %>%  
  select(EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5, EDMQ6, EDMQ7, EDMQ8,  
         DER1, DER2, DER3, DER4,  
         SED1, SED2, SED3,  
         DTL1, DTL2, DTL3, DTL4, DTL5,  
         DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,  
         SDL1, SDL2, SDL3, SDL4, SDL5) %>% mshapiro_test()
```

```
# 7. Verifico l'omogeneità delle varianze
```

```
levene_Gender <- dati_MANOVA_so %>%  
  gather(key = "variable", value = "value", EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5,  
         EDMQ6, EDMQ7, EDMQ8,  
         DER1, DER2, DER3, DER4,  
         SED1, SED2, SED3,  
         DTL1, DTL2, DTL3, DTL4, DTL5,  
         DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,  
         SDL1, SDL2, SDL3, SDL4, SDL5) %>%  
  group_by(variable) %>%  
  levene_test(value ~ Gender, center = median)  
print(levene_Gender, n = 35)
```

```
View(levene_Gender)
```

```
levene_Age <- dati_MANOVA_so %>%  
  gather(key = "variable", value = "value", EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5,  
         EDMQ6, EDMQ7, EDMQ8,  
         DER1, DER2, DER3, DER4,  
         SED1, SED2, SED3,  
         DTL1, DTL2, DTL3, DTL4, DTL5,  
         DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,  
         SDL1, SDL2, SDL3, SDL4, SDL5) %>%  
  group_by(variable) %>%  
  levene_test(value ~ Age, center = median)  
print(levene_Age, n = 35)
```

View(levene_Age)

```
levene_Country <- dati_MANOVA_so %>%  
  gather(key = "variable", value = "value", EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5,  
EDMQ6, EDMQ7, EDMQ8,  
  DER1, DER2, DER3, DER4,  
  SED1, SED2, SED3,  
  DTL1, DTL2, DTL3, DTL4, DTL5,  
  DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,  
  SDL1, SDL2, SDL3, SDL4, SDL5) %>%  
  group_by(variable) %>%  
  levene_test(value ~ Country, center = median)  
print(levene_Country, n = 35)
```

View(levene_Country)

```
levene_sf <- dati_MANOVA_so %>%  
  gather(key = "variable", value = "value", EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5,  
EDMQ6, EDMQ7, EDMQ8,  
  DER1, DER2, DER3, DER4,  
  SED1, SED2, SED3,  
  DTL1, DTL2, DTL3, DTL4, DTL5,  
  DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,  
  SDL1, SDL2, SDL3, SDL4, SDL5) %>%  
  group_by(variable) %>%  
  levene_test(value ~ Subjectfield_agg, center = median)  
print(levene_sf, n = 35)
```

View(levene_sf)

```
levene_te <- dati_MANOVA_so %>%  
  gather(key = "variable", value = "value", EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5,  
EDMQ6, EDMQ7, EDMQ8,  
  DER1, DER2, DER3, DER4,  
  SED1, SED2, SED3,  
  DTL1, DTL2, DTL3, DTL4, DTL5,  
  DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,  
  SDL1, SDL2, SDL3, SDL4, SDL5) %>%  
  group_by(variable) %>%  
  levene_test(value ~ Teach.exp, center = median)  
print(levene_te, n = 35)
```

```
levene_re <- dati_MANOVA_so %>%  
  gather(key = "variable", value = "value", EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5,  
EDMQ6, EDMQ7, EDMQ8,  
  DER1, DER2, DER3, DER4,  
  SED1, SED2, SED3,  
  DTL1, DTL2, DTL3, DTL4, DTL5,  
  DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,  
  SDL1, SDL2, SDL3, SDL4, SDL5) %>%  
  group_by(variable) %>%
```

```
levene_test(value ~ Research.exp, center = median)
print(levene_re, n = 35)
```

```
# 8. Testo l'assunto di multicollinearità
```

```
CorrMat <- cor(alldepvars_so)
```

```
# 9. Calcolo il Cronbach's alfa
```

```
alpha(alldepvars_so)
```

```
# 10. Eseguo il MANOVA
```

```
MANOVA_Gender_allvars_so <- manova(cbind(EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5,
EDMQ6, EDMQ7, EDMQ8,
      DER1, DER2, DER3, DER4,
      SED1, SED2, SED3,
      DTL1, DTL2, DTL3, DTL4, DTL5,
      DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,
      SDL1, SDL2, SDL3, SDL4, SDL5) ~ Gender, dati_MANOVA_so)
summary(MANOVA_Gender_allvars_so)
summary.aov(MANOVA_Gender_allvars_so)
```

```
MANOVA_Age_allvars_so <- manova(cbind(EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5,
EDMQ6, EDMQ7, EDMQ8,
      DER1, DER2, DER3, DER4,
      SED1, SED2, SED3,
      DTL1, DTL2, DTL3, DTL4, DTL5,
      DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,
      SDL1, SDL2, SDL3, SDL4, SDL5) ~ Age, dati_MANOVA_so)
summary(MANOVA_Age_allvars_so)
summary.aov(MANOVA_Age_allvars_so)
```

```
MANOVA_Country_allvars_so <- manova(cbind(EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5,
EDMQ6, EDMQ7, EDMQ8,
      DER1, DER2, DER3, DER4,
      SED1, SED2, SED3,
      DTL1, DTL2, DTL3, DTL4, DTL5,
      DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,
      SDL1, SDL2, SDL3, SDL4, SDL5) ~ Country, dati_MANOVA_so)
summary(MANOVA_Country_allvars_so)
summary.aov(MANOVA_Country_allvars_so)
```

```
MANOVA_Subjectfield_agg_allvars_so <- manova(cbind(EDMQ1, EDMQ2, EDMQ3, EDMQ4,
EDMQ5, EDMQ6, EDMQ7, EDMQ8,
      DER1, DER2, DER3, DER4,
      SED1, SED2, SED3,
      DTL1, DTL2, DTL3, DTL4, DTL5,
      DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,
      SDL1, SDL2, SDL3, SDL4, SDL5) ~ Subjectfield_agg,
dati_MANOVA_so)
```

```
summary(MANOVA_Subjectfield_agg_allvars_so)
summary.aov(MANOVA_Subjectfield_agg_allvars_so)
```

```
MANOVA_Teach.exp_allvars_so <- manova(cbind(EDMQ1, EDMQ2, EDMQ3, EDMQ4,
EDMQ5, EDMQ6, EDMQ7, EDMQ8,
      DER1, DER2, DER3, DER4,
      SED1, SED2, SED3,
      DTL1, DTL2, DTL3, DTL4, DTL5,
      DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,
      SDL1, SDL2, SDL3, SDL4, SDL5) ~ Teach.exp, dati_MANOVA_so)
summary(MANOVA_Teach.exp_allvars_so)
summary.aov(MANOVA_Teach.exp_allvars_so)
```

```
MANOVA_Research.exp_allvars_so <- manova(cbind(EDMQ1, EDMQ2, EDMQ3, EDMQ4,
EDMQ5, EDMQ6, EDMQ7, EDMQ8,
      DER1, DER2, DER3, DER4,
      SED1, SED2, SED3,
      DTL1, DTL2, DTL3, DTL4, DTL5,
      DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,
      SDL1, SDL2, SDL3, SDL4, SDL5) ~ Research.exp,
```

```
dati_MANOVA_so)
summary(MANOVA_Research.exp_allvars_so)
summary.aov(MANOVA_Research.exp_allvars_so)
```

10. Effect size

```
effectsize::eta_squared(MANOVA_Gender_allvars_so)
effectsize::eta_squared(MANOVA_Age_allvars_so)
effectsize::eta_squared(MANOVA_Country_allvars_so)
effectsize::eta_squared(MANOVA_Subjectfield_agg_allvars_so)
effectsize::eta_squared(MANOVA_Teach.exp_allvars_so)
effectsize::eta_squared(MANOVA_Research.exp_allvars_so)
```

11. Eseguo una comparazione multipla a coppie

```
cmc_gender <- dati_MANOVA_so %>%
  gather(key = "variables", value = "value", EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5,
EDMQ6, EDMQ7, EDMQ8,
      DER1, DER2, DER3, DER4,
      SED1, SED2, SED3,
      DTL1, DTL2, DTL3, DTL4, DTL5,
      DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,
      SDL1, SDL2, SDL3, SDL4, SDL5) %>%
  group_by(variables) %>%
  games_howell_test(value ~ Gender)
```

```
tab_cmc_gender <- data.frame(cmc_gender)
View(tab_cmc_gender)
```

```
cmc_age <- dati_MANOVA_so %>%
```



```
gather(key = "variables", value = "value", EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5,  
EDMQ6, EDMQ7, EDMQ8,  
  DER1, DER2, DER3, DER4,  
  SED1, SED2, SED3,  
  DTL1, DTL2, DTL3, DTL4, DTL5,  
  DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,  
  SDL1, SDL2, SDL3, SDL4, SDL5) %>%  
group_by(variables) %>%  
games_howell_test(value ~ Age)
```

```
tab_cmc_age <- data.frame(cmc_age)  
View(tab_cmc_age)
```

```
write.table(tab_cmc_age, file = "tabella cmc age.txt", sep = ",", quote = FALSE)
```

```
cmc_country <- dati_MANOVA_so %>%  
gather(key = "variables", value = "value", EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5,  
EDMQ6, EDMQ7, EDMQ8,  
  DER1, DER2, DER3, DER4,  
  SED1, SED2, SED3,  
  DTL1, DTL2, DTL3, DTL4, DTL5,  
  DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,  
  SDL1, SDL2, SDL3, SDL4, SDL5) %>%  
group_by(variables) %>%  
games_howell_test(value ~ Country)
```

```
tab_cmc_country <- data.frame(cmc_country)  
View(tab_cmc_country)
```

```
write.table(tab_cmc_country, file = "tabella cmc country.txt", sep = ",", quote = FALSE)
```

```
cmc_sf <- dati_MANOVA_so %>%  
gather(key = "variables", value = "value", EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5,  
EDMQ6, EDMQ7, EDMQ8,  
  DER1, DER2, DER3, DER4,  
  SED1, SED2, SED3,  
  DTL1, DTL2, DTL3, DTL4, DTL5,  
  DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,  
  SDL1, SDL2, SDL3, SDL4, SDL5) %>%  
group_by(variables) %>%  
games_howell_test(value ~ Subjectfield_agg)
```

```
tab_cmc_sf <- data.frame(cmc_sf)  
View(tab_cmc_sf)
```

```
write.table(tab_cmc_sf, file = "tabella cmc sf.txt", sep = ",", quote = FALSE)
```

```
cmc_te <- dati_MANOVA_so %>%  
gather(key = "variables", value = "value", EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5,  
EDMQ6, EDMQ7, EDMQ8,  
  DER1, DER2, DER3, DER4,
```

```
    SED1, SED2, SED3,  
    DTL1, DTL2, DTL3, DTL4, DTL5,  
    DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,  
    SDL1, SDL2, SDL3, SDL4, SDL5) %>%  
group_by(variables) %>%  
games_howell_test(value ~ Teach.exp)  
  
tab_cmc_te <- data.frame(cmc_te)  
View(tab_cmc_te)  
  
write.table(tab_cmc_te, file = "tabella cmc te.txt", sep = ",", quote = FALSE)  
  
cmc_re <- dati_MANOVA_so %>%  
  gather(key = "variables", value = "value", EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5,  
EDMQ6, EDMQ7, EDMQ8,  
    DER1, DER2, DER3, DER4,  
    SED1, SED2, SED3,  
    DTL1, DTL2, DTL3, DTL4, DTL5,  
    DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,  
    SDL1, SDL2, SDL3, SDL4, SDL5) %>%  
group_by(variables) %>%  
games_howell_test(value ~ Research.exp)  
  
tab_cmc_re <- data.frame(cmc_re)  
View(tab_cmc_re)  
  
write.table(tab_cmc_re, file = "tabella cmc re.txt", sep = ",", quote = FALSE)
```

8.1.3. Script per l'Analisi Fattoriale

```
# Script per l'Analisi Fattoriale Esplorativa

# 1. Preparo il workspace

# install.packages("corrplot")
# install.packages("GPArotation")
# install.packages("lavaan")
# install.packages("psych")
# install.packages("semPlot")

library(corrplot)
library(GPArotation)
library(lavaan)
library(psych)
library(semPlot)

setwd("C:/Users/pc/Documents/UNIV/TESI DI LAUREA/Dataset")
dati_two_universities <- read.table("DataPractices_TwoUniversities.csv",
                                   header = TRUE, sep = ",")
dati_two_universities[dati_two_universities == "0,00"] <- "0"
dati_two_universities[dati_two_universities == "1,00"] <- "1"
dati_two_universities[dati_two_universities == "2,00"] <- "2"
dati_two_universities[dati_two_universities == "3,00"] <- "3"
dati_two_universities[dati_two_universities == "4,00"] <- "4"
dati_two_universities[dati_two_universities == "5,00"] <- "5"
View(dati_two_universities)

# 2. Convento le variabili
# 2.1. Genere

Gender <- dati_two_universities$Gender
Gender <- factor(Gender,
                 levels = c("Male", "Female"))

# 2.2. Età

Age <- dati_two_universities$Age
Age <- factor(Age,
              levels = c("25-34", "35-44",
                         "45-54", "More than 55"))

# 2.3. Nazione

Country <- dati_two_universities$Country
Country <- factor(Country,
                  levels = c("Italy", "Spain"))

# 2.4. Area disciplinare
```

```
Subject.field.1 <- factor(dati_two_universities$Subject.field.1,  
  levels = c("Agricultural Sciences",  
    "Anthropology",  
    "Anthropology,History",  
    "Anthropology,Pedagogy",  
    "Astronomy and astrophysics",  
    "Chemistry",  
    "Chemistry,Life Sciences,Agricultural Sciences",  
    "Demographics",  
    "Demographics,Sociology",  
    "Earth and Space Sciences",  
    "Earth and Space Sciences,Agricultural Sciences",  
    "Earth and Space Sciences,Technological Sciences",  
    "Earth and Space Sciences,Agricultural Sciences,Geography",  
    "Economic Sciences",  
    "Economic Sciences,Ethics",  
    "Economic Sciences,History",  
    "Economic Sciences,Juridical Sciences and Law",  
    "Economic Sciences,Pedagogy",  
    "Economic Sciences,Pedagogy,Psychology",  
    "Ethics",  
    "Geography",  
    "History",  
    "Humanities",  
    "Juridical Sciences and Law",  
    "Juridical Sciences and Law,Psychology",  
    "Juridical Sciences and Law,Political Science,Sociology",  
    "Juridical Sciences and Law,Science of Arts and Letters",  
    "Life Sciences,Medical Sciences",  
    "Life Sciences,Medical Sciences,Linguistics,Pedagogy,Psychology",  
    "Life Sciences,Technological Sciences",  
    "Life Sciences,Technological Sciences,Pedagogy",  
    "Linguistics,Pedagogy",  
    "Logic",  
    "Mathematics",  
    "Mathematics,Physics,Technological Sciences",  
    "Mathematics,Economic Sciences,Pedagogy",  
    "Mathematics,Economic Sciences,Sociology,Philosophy",  
    "Mathematics,Life Sciences",  
    "Mathematics,Technological Sciences",  
    "Medical Sciences",  
    "Medical Sciences,Anthropology",  
    "Medical Sciences,Pedagogy",  
    "Medical Sciences,Technological Sciences,Pedagogy",  
    "Technological Sciences",  
    "Technological Sciences,Economic Sciences",  
    "Technological Sciences,Economic Sciences,Pedagogy",  
    "Technological Sciences,Geography",  
    "Technological Sciences,Juridical Sciences and Law",
```

Letters,Philosophy",
"Technological Sciences,Linguistics,Pedagogy,Science of Arts and
Letters,Philosophy",
"Technological Sciences,Pedagogy,Philosophy",
"Technological Sciences,Pedagogy",
"Technological Sciences,Science of Arts and Letters",
"Pedagogy",
"Pedagogy,Psychology",
"Pedagogy,Sociology",
"Philosophy",
"Physics",
"Physics,Chemistry,Life Sciences,Earth and Space Sciences,Medical
Sciences,Technological Sciences,Pedagogy,Ethics",
"Physics,Technological Sciences",
"Physics,Technological Sciences,Geography",
"Psychology",
"Psychology,Ethics",
"Political Science",
"Political Science,Sociology",
"Science of Arts and Letters",
"Sociology",
"Formal Sciences",
"Life Sciences",
"Linguistics",
"NA",
"Natural Sciences",
"Technology",
"Social Sciences"))

Subject.field.1[Subject.field.1 == "Mathematics"] <- "Formal Sciences"
Subject.field.1[Subject.field.1 == "Logic"] <- "Formal Sciences"

Subject.field.1[Subject.field.1 == "Science of Arts and Letters"] <- "Humanities"
Subject.field.1[Subject.field.1 == "Philosophy"] <- "Humanities"
Subject.field.1[Subject.field.1 == "History"] <- "Humanities"
Subject.field.1[Subject.field.1 == "Ethics"] <- "Humanities"

Subject.field.1[Subject.field.1 == "Linguistics"] <- "Linguistics"

Subject.field.1[Subject.field.1 == "Medical Sciences"] <- "Life Sciences"
Subject.field.1[Subject.field.1 == "Psychology"] <- "Life Sciences"
Subject.field.1[Subject.field.1 == "Life Sciences"] <- "Life Sciences"
Subject.field.1[Subject.field.1 == "Life Sciences,Technological Sciences,Pedagogy"] <- "Life
Sciences"
Subject.field.1[Subject.field.1 == "Life Sciences,Medical
Sciences,Linguistics,Pedagogy,Psychology"] <- "Life Sciences"
Subject.field.1[Subject.field.1 == "Physics,Chemistry,Life Sciences,Earth and Space
Sciences,Medical Sciences,Technological Sciences,Pedagogy,Ethics"] <- "Life Sciences"
Subject.field.1[Subject.field.1 == "Medical Sciences,Technological Sciences,Pedagogy"] <- "Life
Sciences"
Subject.field.1[Subject.field.1 == "Life Sciences,Medical Sciences"] <- "Life Sciences"
Subject.field.1[Subject.field.1 == "Psychology,Ethics"] <- "Life Sciences"

Subject.field.1[Subject.field.1 == "Mathematics,Life Sciences"] <- "Life Sciences"
Subject.field.1[Subject.field.1 == "Medical Sciences,Pedagogy"] <- "Life Sciences"

Subject.field.1[Subject.field.1 == "Earth and Space Sciences,Agricultural Sciences"] <- "Natural Sciences"
Subject.field.1[Subject.field.1 == "Physics"] <- "Natural Sciences"
Subject.field.1[Subject.field.1 == "Geography"] <- "Natural Sciences"
Subject.field.1[Subject.field.1 == "Earth and Space Sciences"] <- "Natural Sciences"
Subject.field.1[Subject.field.1 == "Chemistry"] <- "Natural Sciences"
Subject.field.1[Subject.field.1 == "Astronomy and astrophysics"] <- "Natural Sciences"
Subject.field.1[Subject.field.1 == "Earth and Space Sciences,Technological Sciences"] <- "Natural Sciences"
Subject.field.1[Subject.field.1 == "Chemistry,Life Sciences,Agricultural Sciences"] <- "Natural Sciences"
Subject.field.1[Subject.field.1 == "Earth and Space Sciences,Agricultural Sciences,Geography"] <- "Natural Sciences"

Subject.field.1[Subject.field.1 == "Pedagogy"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Political Science"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Sociology"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Anthropology"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Pedagogy"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Economic Sciences"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Juridical Sciences and Law"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Demographics"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Juridical Sciences and Law,Psychology"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Pedagogy,Psychology"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Mathematics,Economic Sciences,Sociology,Philosophy"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Technological Sciences,Pedagogy"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Pedagogy,Sociology"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Anthropology,History"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Anthropology,Pedagogy"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Political Science,Sociology"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Economic Sciences,Pedagogy"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Demographics,Sociology"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Juridical Sciences and Law,Science of Arts and Letters"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Mathematics,Economic Sciences,Pedagogy"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Juridical Sciences and Law,Political Science,Sociology"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Economic Sciences,History"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Technological Sciences,Juridical Sciences and Law"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Economic Sciences,Ethics"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Economic Sciences,Pedagogy,Psychology"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Technological Sciences,Economic Sciences,Pedagogy"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Technological Sciences,Linguistics,Pedagogy,Science of Arts and Letters,Philosophy"] <- "Social Sciences"

```
Subject.field.1[Subject.field.1 == "Linguistics,Pedagogy"] <- "Social Sciences"
Subject.field.1[Subject.field.1 == "Economic Sciences,Juridical Sciences and Law"] <- "Social
Sciences"
Subject.field.1[Subject.field.1 == "Medical Sciences,Anthropology"] <- "Social Sciences"

Subject.field.1[Subject.field.1 == "Technological Sciences"] <- "Technology"
Subject.field.1[Subject.field.1 == "Technological Sciences,Science of Arts and Letters"] <-
"Technology"
Subject.field.1[Subject.field.1 == "Agricultural Sciences"] <- "Technology"
Subject.field.1[Subject.field.1 == "Technological Sciences,Economic Sciences"] <- "Technology"
Subject.field.1[Subject.field.1 == "Technological Sciences,Pedagogy,Philosophy"] <-
"Technology"
Subject.field.1[Subject.field.1 == "Technological Sciences,Economic Sciences"] <- "Technology"
Subject.field.1[Subject.field.1 == "Mathematics,Physics,Technological Sciences"] <- "Technology"
Subject.field.1[Subject.field.1 == "Mathematics,Technological Sciences"] <- "Technology"
Subject.field.1[Subject.field.1 == "Physics,Technological Sciences,Geography"] <- "Technology"
Subject.field.1[Subject.field.1 == "Technological Sciences,Science of Arts and Letters"] <-
"Technology"
Subject.field.1[Subject.field.1 == "Physics,Technological Sciences"] <- "Technology"
Subject.field.1[Subject.field.1 == "Life Sciences,Technological Sciences"] <- "Technology"
Subject.field.1[Subject.field.1 == "Technological Sciences,Geography"] <- "Technology"

Subjectfield_agg <- factor(Subject.field.1,
  levels = c("Formal Sciences",
            "Humanities",
            "Life Sciences",
            "Linguistics",
            "Natural Sciences",
            "Technology",
            "Social Sciences"))

# 2.5. Esperienza d'insegnamento

Teach.exp <- dati_two_universities$Teach_exp
Teach.exp <- factor(Teach.exp,
  levels = c("Less than 3","3-4","5-10","10-15",
            "More than 15"))

# 2.6. Esperienza di ricerca

Research.exp <- dati_two_universities$Research_exp
Research.exp <- factor(Research.exp,
  levels = c("No experience","Less than 3",
            "3-5","6-10","11-15",
            "More than 15"))

# 2.7. Converto le variabili quantitative

EDMQ1 <- as.numeric(dati_two_universities$EDMQ1)
EDMQ2 <- as.numeric(dati_two_universities$EDMQ2)
EDMQ3 <- as.numeric(dati_two_universities$EDMQ3)
```

```
EDMQ4 <- as.numeric(dati_two_universities$EDMQ4)
EDMQ5 <- as.numeric(dati_two_universities$EDMQ5)
EDMQ6 <- as.numeric(dati_two_universities$EDMQ6)
EDMQ7 <- as.numeric(dati_two_universities$EDMQ7)
EDMQ8 <- as.numeric(dati_two_universities$EDMQ8)
DER1 <- as.numeric(dati_two_universities$DER1)
DER2 <- as.numeric(dati_two_universities$DER2)
DER3 <- as.numeric(dati_two_universities$DER3)
DER4 <- as.numeric(dati_two_universities$DER4)
SED1 <- as.numeric(dati_two_universities$SED1)
SED2 <- as.numeric(dati_two_universities$SED2)
SED3 <- as.numeric(dati_two_universities$SED3)
DTL1 <- as.numeric(dati_two_universities$DTL1)
DTL2 <- as.numeric(dati_two_universities$DTL2)
DTL3 <- as.numeric(dati_two_universities$DTL3)
DTL4 <- as.numeric(dati_two_universities$DTL4)
DTL5 <- as.numeric(dati_two_universities$DTL5)
DA1 <- as.numeric(dati_two_universities$DA1)
DA10 <- as.numeric(dati_two_universities$DA10)
DA2 <- as.numeric(dati_two_universities$DA2)
DA3 <- as.numeric(dati_two_universities$DA3)
DA4 <- as.numeric(dati_two_universities$DA4)
DA5 <- as.numeric(dati_two_universities$DA5)
DA6 <- as.numeric(dati_two_universities$DA6)
DA7 <- as.numeric(dati_two_universities$DA7)
DA8 <- as.numeric(dati_two_universities$DA8)
DA9 <- as.numeric(dati_two_universities$DA9)
SDL1 <- as.numeric(dati_two_universities$SDL1)
SDL2 <- as.numeric(dati_two_universities$SDL2)
SDL3 <- as.numeric(dati_two_universities$SDL3)
SDL4 <- as.numeric(dati_two_universities$SDL4)
SDL5 <- as.numeric(dati_two_universities$SDL5)
```

3. Costruisco il data frame per le variabili quantitative

3.1. Costruisco il data frame per tutte le variabili

```
dati_MANOVA <- data.frame(Gender, Age, Country, Subjectfield_agg, Teach.exp, Research.exp,
  EDMQ1, EDMQ2, EDMQ3, EDMQ4, EDMQ5, EDMQ6, EDMQ7, EDMQ8,
  DER1, DER2, DER3, DER4,
  SED1, SED2, SED3,
  DTL1, DTL2, DTL3, DTL4, DTL5,
  DA1, DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9, DA10,
  SDL1, SDL2, SDL3, SDL4, SDL5)
```

```
dati_MANOVA <- na.omit(dati_MANOVA)
```

3.2. Costruisco il data frame per le variabili dipendenti

```
alldpvars <- dati_MANOVA[,-c(1:6)]
```

4. Creo un dataset senza outlier


```
md <- mahalanobis((alldepvars), center = colMeans(alldepvars), cov = cov(alldepvars))  
alpha <- (.001)
```

```
cutoff <- (qchisq(p = 1 - alpha, df = ncol(alldepvars)))  
names_outliers_MH <- which(md > cutoff)
```

```
dati_alldepvars_so <- alldepvars[-names_outliers_MH, ]
```

```
# 5. Test di fattibilità AFE
```

```
# 5.1. Realizzo la matrice di correlazione
```

```
MatriceCorrelazione <- cor(dati_alldepvars_so)  
View(data.frame(MatriceCorrelazione))
```

```
windows()
```

```
Grafico_MatriceCorrelazione <- corrplot(MatriceCorrelazione,  
                                         method = "circle",  
                                         type = "lower",  
                                         title = "Matrice di correlazione",  
                                         is.corr = FALSE,  
                                         order = "original",  
                                         diag = TRUE,  
                                         outline = FALSE,  
                                         plotCI = NULL)
```

```
# 5.2. Test di controllo Kaiser-Meyer-Olkin
```

```
KMO(r=cor(dati_alldepvars_so))
```

```
# 5.3. Test di sfericità di Bartlett
```

```
cortest.bartlett(dati_alldepvars_so, n = 494)  
det(cor(dati_alldepvars_so))
```

```
# 5.4. Analisi parallela con scree plot
```

```
windows()
```

```
parallel <- fa.parallel(dati_alldepvars_so)
```

```
# 6. Analisi fattoriale con funzione fa
```

```
EFA_quartimax <- fa(r = dati_alldepvars_so,  
                   nfactors = 8,  
                   rotate = "quartimax",  
                   fm = "ml")
```

```
tab_saturazioni <- print(EFA_quartimax$loadings, cutoff = 0.2)  
write.table(tab_saturazioni, file = "tab_saturazioni.txt", sep = ",", quote = FALSE)
```

```
tab_comunalita_var <- round(EFA_quartimax$communality, 8)
```

```
write.table(tab_comunalita, file = "tab_comunalita.txt", sep = ",", quote = FALSE)
```

```
# 7. Eseguo l'Analisi Fattoriale Confermativa
```

```
var_model <-'
```

```
ML2 =~ DA1 + DA2 + DA3 + DA4 + DA5 + DA6 + DA9
```

```
ML3 =~ EDMQ1 + EDMQ2 + EDMQ3 + EDMQ4 + EDMQ5 + DER1 + SED2 + DA1 + DA2 +  
DA6
```

```
ML6 =~ DA4 + DA5 + DA6 + DA7 + DA8 + DA9 + DA10
```

```
ML5 =~ EDMQ1 + EDMQ2 + DER1 + DER2 + DER3 + SED1
```

```
ML4 =~ EDMQ1 + EDMQ7 + EDMQ8 + DER3 + SED3 + DA5 + DA8
```

```
ML7 =~ DA5 + DTL2 + DTL3
```

```
ML8 =~ SED1 + SED2 + SED3 + DA5 + DA9
```

```
ML1 =~ ML2 + ML3 + ML4 + ML5 + ML6 + ML7 + ML8 + DA5 + DA9
```

```
EDMQ2 =~ EDMQ1
```

```
EDMQ4 =~ EDMQ5
```

```
EDMQ5 =~ DA7
```

```
DER1 =~ DER2
```

```
DER2 =~ DER3
```

```
DA1 =~ DA2
```

```
DA8 =~ DA10
```

```
DA6 =~ DA7
```

```
DA2 =~ EDMQ2
```

```
ML2 =~ ML6
```

```
ML2 =~ ML3
```

```
ML3 =~ ML4
```

```
ML6 =~ ML4
```

```
ML6 =~ ML7
```

```
ML3 =~ ML5'
```

```
fit_var <-cfa(var_model, data = dati_alldepvars_so,  
ordered = names(dati_alldepvars_so))
```

```
summary(fit_var, standardized = TRUE, fit.measures = T)
```

```
mod_ind <- modindices(fit_var)
```

```
head(mod_ind[order(mod_ind$mi, decreasing=TRUE),], 10)
```

```
# User Model versus Baseline Model:
```

```
# Comparative Fit Index (CFI)          1.000    0.992
```

```
# Tucker-Lewis Index (TLI)            1.000    0.989
```

```
# Root Mean Square Error of Approximation:
```

```
# RMSEA                                0.000    0.044
```

```
# 90 Percent confidence interval - lower 0.000    0.031
```

```
# 90 Percent confidence interval - upper 0.009    0.050
```

```
# Standardized Root Mean Square Residual:
```

```
# SRMR                                  0.027    0.027
```

```
latent_vars <- list("ML1","ML2","ML3","ML4","ML5","ML6","ML7","ML8")
```

```
latent_var_labels <- c("ML1", "ML8", "ML7", "ML6", "ML5", "ML4", "ML3", "ML2")
```

```
manifest_var_labels <- c("SED1", "SED2", "SED3",  
  "DER1", "DER2", "DER3",  
  "DTL2", "DTL3",  
  "DA7", "DA8", "DA9", "DA10",  
  "EDMQ1", "EDMQ2", "EDMQ3", "EDMQ4", "EDMQ5", "EDMQ7", "EDMQ8",  
  "DA1", "DA2", "DA3", "DA4", "DA5", "DA6")
```

```
Path_diagram_CFA_circle <- semPaths(fit_var, what = "std", layout = "circle2",  
  whatLabels = "std", edge.label.cex = 0.3, edge.color = 1, label.color = 1,  
  esize = 1, asize = 1,  
  sizeMan = 4,  
  intercepts = F, rotation = 3,  
  mar = c(1,5,1.5,5), fade = T,  
  nCharNodes = 2, nCharEdges = 2,  
  shapeMan = "rectangle", shapeLat = "ellipse",  
  latents = latent_var_labels,  
  manifests = manifest_var_labels,  
  groups = latent_vars,  
  thresholds = F,  
  reorder = F,  
  weighted = T,  
  arrows = T,  
  filename = "PathDiagram_circle", filetype = "jpg",  
  width = 30, height = 25)
```

```
Path_diagram_CFA_tree <- semPaths(fit_var, what = "est",  
  whatLabels = "std",  
  layout = "tree",  
  style = "lisrel",  
  sizeMan = 3, sizeMan2 = 2.5,  
  sizeLat = 4, sizeLat2 = 4,  
  shapeLat = "square",  
  shapeMan = "rectangle",  
  edge.label.cex = 0.2, edge.color = 1,  
  label.color = 1,  
  label.scale = T,  
  label.norm = T,  
  edge.label.position = 0.7,  
  esize = 1, asize = 1,  
  curveAdjacent = 'cov',  
  optimizeLatRes = T,  
  intAtSide = T,  
  intercepts = F,  
  residuals = F,  
  thresholds = F,  
  normalize = T,  
  intStyle = "multi",  
  reorder = F,  
  rotation = 3,
```

```
latents = latent_var_labels,  
manifests = manifest_var_labels,  
groups = latent_vars,  
filename = "PathDiagram_tree", filetype = "jpg",  
width = 30, height = 25)
```

```
Path_diagram_CFA_structure <- semPaths(fit_var, what = "est",  
  whatLabels = "std",  
  layout = "tree",  
  style = "mx",  
  structural = T,  
  sizeLat = 7, sizeLat2 = 7,  
  shapeLat = "circle",  
  edge.label.cex = 0.5, edge.color = 1,  
  label.color = 1,  
  label.scale = T,  
  label.norm = T,  
  edge.label.position = 0.5,  
  esize = 1, asize = 1,  
  curveAdjacent = 'cov',  
  intAtSide = T,  
  intercepts = F,  
  residuals = T,  
  thresholds = F,  
  normalize = T,  
  intStyle = "multi",  
  reorder = T,  
  rotation = 3,  
  fade = F,  
  latents = latent_var_labels,  
  groups = latent_vars,  
  filename = "PathDiagram_tree_structure", filetype = "jpg",  
  width = 30, height = 25)
```