

**Matricola n. 0000997866**

**ALMA MATER STUDIORUM  
UNIVERSITA' DI BOLOGNA**

**SCUOLA DI SCIENZE**

**Corso di laurea in Informatica per il Management**

**Analisi del Minutaggio Under 21 in Serie  
A: Dalle Affermazioni Comuni all'Evidenza  
Statistica**

**Tesi di laurea in Architettura di Internet**

**Relatore**

**Prof. Marco Rocetti**

**Presentata da**

**Chiara Anna Cartarasa**

**Sessione III  
Anno Accademico 2022/2023**

*Alla mia famiglia, la parte migliore di me.*

## *Indice*

<b>Introduzione.....</b>	<b>5</b>
--------------------------	----------

### **I Moneyball e l'analisi statistica nello sport**

1.1. Storia e concetto di “Moneyball” .....	6
1.2. Applicazione nello sport. ....	8
1.3. Importanza dei modelli statistici nel calcio e nel mondo sportivo.....	11

### **II Metodologia**

2.1 Raccolta dei dati.....	12
2.2 Utilizzo di R Studio e il test di Kolmogorov-Smirnov .....	12
2.2.1 Strumenti di analisi dati .....	13
2.2.2 Test di Kolmogorov-Smirnov: metodologia e scopo .....	14
2.3 Campione e periodo di riferimento .....	16
2.3.1 Selezione del campione .....	16
2.3.2 Definizione del periodo di riferimento.....	17

### **III Analisi dei dati**

3.1 Caricamento delle librerie e inizializzazione del dataset.....	19
3.2 Normalizzazione dei dati .....	23
3.3 Analisi dei grafici delle singole stagioni .....	25
3.4 Preparazione ai test .....	29

3.5 Test di Kolmogorov-Smirnov .....	35
3.5.1 Gruppo1 vs gruppo2:Test a una coda.....	35
3.5.2 Gruppo3 vs gruppo2:Test a una coda.....	36
3.5.3 Gruppo3 vs gruppo1:Test a due code.....	38
3.6 Esito e interpretazione dei risultati .....	39

#### **IV Discussione**

4.1 Limiti dell'approccio .....	40
4.1.1 Divisione in soli tre gruppi.....	40
4.1.2 Scelta periodo di studio .....	41
4.1.3 Scelta di non distinguere la nazionalità dei giocatori.....	41
4.1.4 Non aver cercato correlazioni con la sentenza Bosman.....	42
4.2 Fattori che influenzano il minutaggio degli under 21 .....	43
4.2.1 Livello di fiducia del club .....	44
4.2.2 Maturità fisica e psicofisica .....	44
4.2.3 Posizione in campo .....	44
4.2.4 Esperienza di gioco .....	44
4.2.5 Prestazioni durante l'allenamento .....	45
4.2.6 Influenza dei dati statistici .....	45
4.3 Confronto risultati ottenuti e affermazioni comuni .....	45
4.4 Implicazioni per lo sviluppo dei giovani talenti .....	46

<b>Elenco figure .....</b>	<b>49</b>
----------------------------	-----------

<b>Ringraziamenti .....</b>	<b>50</b>
-----------------------------	-----------

<b>Bibliografia e Sitografia.....</b>	<b>52</b>
---------------------------------------	-----------

## Introduzione

L'espressione 'Italia non è un Paese per giovani' è spesso evocata nei dibattiti sul minutaggio degli *under 21* nella Serie A, suggerendo un declino delle opportunità per i giovani talenti nel calcio italiano. Questa tesi si pone l'obiettivo di analizzare approfonditamente queste affermazioni attraverso un'indagine statistica dettagliata. La percezione comune di una riduzione delle opportunità per gli *under 21* costituisce un punto cruciale di partenza, che si intende mettere in discussione, evidenziando la complessità della situazione.

Inoltre, l'attualità di questo tema è sottolineata dal caso di Francesco Camarda, il talento del Milan che ha recentemente fatto il suo debutto in Serie A a soli 15 anni, segnando un nuovo record per il calcio italiano. Questo evento evidenzia la dinamicità del panorama calcistico italiano e aggiunge un elemento di attualità alla discussione sulle opportunità per i giovani giocatori.

Attraverso un'analisi statistica, si esploreranno le dinamiche del minutaggio degli *under 21* nel corso degli anni, confrontandole con il passato e rivelando eventuali differenze; inoltre, ci si concentrerà sull'importanza dei dati, considerando come le informazioni raccolte tramite analisi statistica possano influenzare la comprensione del minutaggio. Questo approccio si basa su principi informati dalla centralità dell'*Architettura di Internet*, che fornisce le basi concettuali e metodologiche per trattare i dati come risorsa critica, enfatizzando la necessità di un approccio informato basato su evidenze.

In linea con il concetto più ampio di analisi statistica nello sport, questa ricerca si ispira al '*Moneyball*' e la sua rilevanza nella guida delle decisioni sportive.

# I *Moneyball* e l'analisi statistica nello sport

## 1.1 Storia e concetto di *Moneyball*.

Il concetto di "*Moneyball*" è strettamente legato all'analisi statistica applicata allo sport professionistico per la selezione dei giocatori. Questo termine deriva dal film di successo "*Moneyball*", basato sul libro di Michael Lewis. Il film, uscito nel 2011, racconta la storia di Billy Beane, il *general manager* della squadra di *baseball* degli Oakland Athletics, e il suo tentativo di assemblare una squadra competitiva con un *budget* limitato, utilizzando un'analisi generata al computer per acquisire nuovi giocatori<sup>i</sup>.

La storia è ambientata nel contesto della stagione di Major League Baseball del 2002. Gli Oakland Athletics avevano appena perso tre dei loro giocatori stella - Jason Giambi, Jason Isringhausen e Johnny Damon - e si trovavano di fronte a gravi limitazioni di *budget*. Billy Beane, interpretato da Brad Pitt, insieme al suo assistente *general manager* Peter Brand, interpretato da Jonah Hill, ha deciso di utilizzare una strategia chiamata "*sabermetrics*"<sup>1</sup> per identificare e reclutare giocatori sottovalutati<sup>ii</sup>.

La *sabermetrics* è un approccio analitico al *baseball* che utilizza l'analisi statistica per misurare le prestazioni dei giocatori e prendere decisioni strategiche. Questo metodo è stato promosso per la prima volta da Bill James negli anni '80 ed è stato adottato da Beane per superare la disparità finanziaria tra gli Athletics e squadre di mercato più ampio.

---

<sup>1</sup> Il termine deriva dall'acronimo SABR, che sta per "Society for American Baseball Research" (Società americana per la ricerca sul baseball) e fu coniato da Bill James, che è stato tra i suoi primi proponenti e per lungo tempo il suo principale sostenitore.

Dal Manifesto della *sabermetrica* di David Grabiner([Sabermetrica - Wikipedia](#))

Nonostante la resistenza iniziale e le critiche dei tradizionalisti, l'approccio innovativo di Beane ebbe successo. Sotto la sua guida e l'applicazione della *sabermetrics*, gli Oakland Athletics sono riusciti a raggiungere i *playoff* per quattro anni consecutivi tra il 2000 e il 2003, ottenendo addirittura una serie *record* di 20 vittorie consecutive, nonostante avessero un *budget* quasi un terzo delle squadre come i New York Yankees<sup>iii</sup>.

Il film presenta una rappresentazione avvincente della serie di 20 vittorie degli Athletics, la più lunga mai realizzata da una squadra della American League all'epoca. La 20<sup>a</sup> partita, in particolare, è stata un momento cruciale che ha dimostrato il potenziale dell'analisi nel prendere decisioni nel *baseball*. In questa partita, gli A's erano in vantaggio 11-0 contro i Kansas City Royals dopo il terzo *inning*. Tuttavia, a causa di errori non forzati, i Royals sono riusciti a pareggiare la partita al nono *inning*. In una conclusione climatica, Scott Hatteberg, un giocatore il cui valore è stato identificato attraverso la strategia Moneyball, ha colpito un fuoricampo per assicurare la vittoria degli Athletics.

Dal punto di vista statistico, l'approccio "*Moneyball*" si fonda su una serie di metriche chiave per valutare le prestazioni dei giocatori, tra cui l'*On-Base Percentage* (OBP), che misura la frequenza con cui un giocatore raggiunge la base. Questa statistica è stata utilizzata per individuare giocatori sottovalutati, spesso trascurati nonostante le loro abilità<sup>iv</sup>.

Un altro aspetto fondamentale è l'utilizzo dei dati storici per prevedere le future prestazioni dei giocatori, basandosi su modelli statistici che considerano variabili come le performance passate, l'età e le condizioni di salute. Questi modelli guidano le decisioni di reclutamento e formazione della squadra. Inoltre, l'approccio "*Moneyball*" si focalizza sull'efficienza e sull'ottimizzazione delle risorse, cercando di massimizzare il rendimento da ogni investimento. Invece di concentrarsi su pochi giocatori stellari, suggerisce di distribuire le risorse su un gruppo più ampio, dimostratosi efficiente secondo le metriche chiave.

In sintesi, il concetto di "*Moneyball*" rappresenta l'applicazione di metodi statistici avanzati e analisi dei dati nello sport, con l'obiettivo di fare previsioni informate, ottimizzare l'uso delle risorse e migliorare le prestazioni della squadra.

## 1.2 Applicazione nello sport.

L'approccio "*Moneyball*" è stato applicato in vari sport, oltre al *baseball*, per guidare le decisioni strategiche basate su dati e analisi.

- **Football Americano**

Nell'NFL, diversi team hanno abbracciato con successo l'approccio "*Moneyball*" nel reclutamento dei giocatori e nella costruzione delle squadre. Ad esempio, i Cleveland Browns hanno implementato questo metodo con l'aiuto di Paul DePodesta, che precedentemente aveva collaborato con Billy Beane per raffinare il processo di acquisto di giocatori per gli Oakland Athletics<sup>v</sup>. L'uso di dati statistici avanzati ha consentito ai Browns di identificare talenti sottovalutati e di prendere decisioni più informate nel reclutamento dei giocatori. In modo simile, i New England Patriots sono noti per la loro adozione precoce dell'analisi dei dati, che ha contribuito a plasmare le strategie di squadra e a scoprire giocatori che erano stati trascurati da altre squadre. I Philadelphia Eagles sono un altro esempio di successo nell'applicazione del "*Moneyball*" nella NFL, utilizzando metriche avanzate per guidare le loro decisioni in materia di roster e strategia di gioco. Questi esempi dimostrano come l'approccio basato sui dati abbia rivoluzionato la costruzione delle squadre nella NFL, portando a successi e risultati di rilievo in termini di vittorie e prestazioni nei *playoff*<sup>vi</sup>.



- **Basket**

*Moneyball* ha portato all'adozione massiccia di metriche avanzate nella valutazione delle prestazioni dei giocatori, nella pianificazione delle strategie di gioco e nelle decisioni di gestione del personale. L'efficienza è diventata una parola d'ordine, con una maggiore enfasi su tiri da tre punti e schiacciate, considerati più efficienti rispetto a tiri da due punti più lunghi. *Moneyball* ha anche evidenziato il valore di giocatori sottovalutati, incoraggiando le squadre NBA a identificare talenti non convenzionali attraverso analisi avanzate e *scouting* dettagliato. La gestione del **salary cap** è stata influenzata dall'approccio *Moneyball*, con una ricerca attenta di giocatori sottostimati che offrano produzione a costi inferiori, ottimizzando così l'*allocation* delle risorse. Inoltre, c'è stata un'accentuata enfasi nello sviluppo giovanile, con le squadre NBA che investono maggiormente nella crescita dei giovani talenti attraverso programmi di sviluppo e leghe affiliate.

Sebbene l'impatto diretto di *Moneyball* nel *basket* potrebbe non essere paragonabile a quello nel *baseball*, i principi fondamentali, come l'analisi dati, l'efficienza e la valorizzazione dei talenti trascurati, hanno sicuramente plasmato l'approccio della lega alla valutazione dei giocatori, alla gestione delle squadre e alla strategia di gioco<sup>vii</sup>.

- **Calcio**

Il modello "*Moneyball*" è stato applicato anche nel calcio per migliorare la strategia di reclutamento e le prestazioni della squadra. Questo approccio si basa sull'uso di analisi dei dati per identificare i giocatori che offrono il miglior rapporto qualità-prezzo, piuttosto che concentrarsi esclusivamente sulle stelle più luminose e più costose.

Un esempio di applicazione del modello "*Moneyball*" nel calcio è Il Brentford FC, squadra di *Premier League* inglese. Il proprietario, Matthew Benham, un fisico con un *background* nell'industria delle scommesse

sportive, ha applicato le sue competenze in analisi dati e statistica per plasmare la strategia del *club*. Collaborando con una squadra di analisti, Benham ha adottato modelli statistici per individuare giocatori sottovalutati e sviluppare strategie di gioco efficaci. Questo approccio ha fruttato notevoli successi, portando il Brentford FC a raggiungere la finale dei *playoff* per la promozione in Premier League nel 2020, un traguardo che il *club* non aveva toccato per anni. L'implementazione attenta e accurata dell'approccio "*Moneyball*" ha richiesto un'analisi dettagliata delle *performance* dei giocatori, l'identificazione di talenti trascurati basati su variabili chiave e l'utilizzo di modelli statistici per anticipare le future prestazioni dei giocatori. In conclusione, il Brentford FC si erge come un esempio eloquente di come il modello "*Moneyball*" possa essere applicato con successo nel calcio, consentendo al *club* di competere in modo efficace nonostante un *budget* più limitato rispetto ad altri club.

Un altro esempio più recente è il Milan che, sotto la gestione del fondo RedBird, ha adottato un approccio "*Moneyball*" nel calciomercato, utilizzando metodi statistici per massimizzare la performance attesa per ogni euro investito. Questo non implica necessariamente la ricerca di giocatori sconosciuti solo perché selezionati da un algoritmo, ma piuttosto la realizzazione di scelte di investimento più sagge basate su dati solidi<sup>viii</sup>. Un esempio tangibile di questa strategia è l'attenzione particolare ai giocatori con un anno rimanente nei loro contratti, migliorando così la posizione negoziale del club. Negli ultimi giorni di mercato, la dirigenza milanista era disposta a investire 18 milioni in Taremi, ma quando l'operazione non si è concretizzata, hanno optato per Jovic a costo zero. Questi movimenti suggeriscono che il Milan potrebbe avere un *budget* a gennaio o che la quota non spesa potrebbe essere aggiunta allo stesso nella prossima finestra di trasferimenti<sup>ix</sup>.

### **1.3 Importanza dei modelli statistici nel calcio e nel mondo sportivo.**

Nel panorama sportivo contemporaneo, l'epoca romantica dello sport, basata su passione, intuizione e talento, sta cedendo il passo a una nuova era guidata dalla tecnologia e dalla statistica. I *Big Data* sono diventati i protagonisti di questa rivoluzione, unendo numeri, dati e algoritmi per trasformare radicalmente la gestione degli atleti e delle squadre. Questa trasformazione è evidente in tutti gli sport, soprattutto quelli di grande impatto economico, dove le decisioni degli allenatori e delle società sono sempre più basate sui dati. I risultati tangibili di questa tendenza si riflettono negli investimenti crescenti in servizi di dati e soluzioni analitiche, con il mercato dei dati nello sport che ha raggiunto i 3 miliardi di euro nel 2022 e si prevede che crescerà fino a 8 miliardi di euro nel 2026.

Dal baseball alla Formula 1, la *data analytics* si è diffusa ampiamente coinvolgendo diversi sport, inclusi il calcio, dove l'utilizzo di astrofisici da parte del Manchester City e di neuroscienziati da parte del Liverpool evidenziano il ruolo cruciale dei dati nella massimizzazione delle performance dei giocatori. Questa sinergia tra scienza e sport ha portato a risultati sorprendenti, come dimostrato dalla capacità del Liverpool di segnare più gol su azione da palla inattiva. L'importanza dei modelli statistici e della *data analytics* nel mondo sportivo sta delineando una nuova era, dove i dati non solo guidano decisioni e strategie, ma contribuiscono a migliorare le prestazioni e l'esperienza complessiva per squadre, tifosi e l'industria sportiva nel suo complesso. Questa rivoluzione, sebbene guidata dai dati, conserva la componente umana che rende lo sport unico e affascinante, preservando la centralità del talento anche in un contesto sempre più tecnologico<sup>x</sup>.

## II Metodologia

### 2.1 Raccolta dei dati

La raccolta dei dati è stata condotta utilizzando il sito *web* FBref<sup>2</sup> che fornisce dettagliate informazioni sul minutaggio dei giocatori di tutte le squadre partecipanti alle competizioni calcistiche e ha dimostrato di essere una fonte affidabile per l'analisi nel contesto del calcio italiano.

I dati sono stati raccolti da questo sito per le stagioni dal 1988-1989 fino alla stagione calcistica più recente 2022-2023 per analizzare gli ultimi 30 anni.

Nel processo di raccolta dei dati, sono state estratte le informazioni relative a squadre e giocatori per ogni stagione e organizzate in un foglio Excel.

Successivamente, sono stati selezionati i dati relativi ai giocatori che avevano 21 anni all'inizio della stagione corrente.

Questi dati sono stati quindi trasferiti in un secondo foglio denominato "*under21*," consentendo un'analisi specifica del minutaggio e delle prestazioni dei giocatori di questa fascia d'età nel corso delle diverse stagioni.

### 2.2 Utilizzo di RStudio e il test di Kolmogorov-Smirnov

La sezione 2.2 si concentra sull'impiego di RStudio<sup>3</sup> come principale ambiente di sviluppo integrato (IDE) e sull'applicazione del test di Kolmogorov-Smirnov nell'analisi statistica.

RStudio è stato scelto per la sua flessibilità e robustezza nel trattamento dei dati, mentre il test di Kolmogorov-Smirnov si presenta come uno strumento chiave per

---

<sup>2</sup>FBRef. (s.d.). Statistiche di Serie A. Recuperato da <https://fbref.com/it/comp/11/Statistiche-di-Serie-A>

<sup>3</sup> <https://www.rstudio.com/>

valutare l'uguaglianza tra distribuzioni di probabilità, contribuendo in modo significativo alle analisi successive.

### 2.2.1 Strumenti di analisi dati

Nel contesto dell'analisi dei dati condotta, è stato fatto ampio utilizzo di strumenti rilevanti per garantire l'efficienza e l'accuratezza dell'indagine. In particolare, è stato selezionato RStudio per il trattamento dei dati e le analisi statistiche.

RStudio, noto per essere una piattaforma open source ampiamente utilizzata per il linguaggio di programmazione R, è stato scelto in quanto offre una serie di funzionalità avanzate per agevolare il processo analitico.

Attraverso RStudio, è stato possibile condurre analisi interattive in tempo reale utilizzando codice R, semplificando la manipolazione<sup>xi</sup>, la visualizzazione dei grafici<sup>xii</sup> e il calcolo dei dati<sup>xiii</sup>.

L'integrazione di RStudio con numerosi pacchetti specializzati di R ha ulteriormente esteso le capacità analitiche, consentendo l'impiego di metodi statistici avanzati e approcci personalizzati.

In aggiunta, sono state impiegate diverse librerie R, tra cui "ggplot2"<sup>4</sup>, "dplyr"<sup>5</sup>, e "stats"<sup>6</sup>, per fornire funzionalità supplementari nell'ambito dell'analisi dei dati. Queste librerie hanno giocato un ruolo cruciale nel migliorare l'efficienza e la completezza dell'analisi.

---

<sup>4</sup> ggplot2: Libreria R per la creazione di grafici e visualizzazioni

<sup>5</sup> dplyr: Libreria R per la manipolazione efficiente dei dati2.

<sup>6</sup> stats: Libreria statistica di base di R

Per l'importazione dei dati da una fonte esterna, sono state utilizzate le librerie "*readxl*"<sup>7</sup> e "*openxlsx*"<sup>8</sup>, semplificando il processo di lettura dei dati da un file <sup>xiv</sup>*Excel*<sup>xv</sup>.

Nella fase di organizzazione dei dati, è stato creato un nuovo *dataframe* denominato "*db\_analisi*," che ha fornito una struttura coerente per le operazioni analitiche successive.

Infine, sono stati eseguiti test di Kolmogorov-Smirnov per valutare la distribuzione delle variabili e condurre confronti tra gruppi di dati.

L'impiego di tali strumenti ha svolto un ruolo fondamentale nell'analisi dei dati condotta, agevolando l'effettuazione di analisi complesse e l'interpretazione dei risultati, nel rispetto dell'obiettivo di rispondere alle domande di ricerca in modo dettagliato ed esaustivo.

### **2.2.2 Test di Kolmogorov-Smirnov: metodologia e scopo**

Il test di Kolmogorov-Smirnov è un test non parametrico sviluppato per confrontare la forma delle distribuzioni campionarie. Può essere utilizzato per confrontare un campione con una distribuzione di riferimento oppure per confrontare due campioni <sup>9</sup>.

Il test è basato sulla massima differenza tra le funzioni di distribuzione cumulativa (CDF) dei due campioni che vengono confrontati. La statistica del test misura la massima distanza verticale tra le due CDF.

---

<sup>7</sup> readxl: Libreria R per l'importazione efficiente dei dati da file Excel1.

<sup>8</sup> openxlsx: Libreria R per la manipolazione e l'importazione dei dati da file Excel

<sup>9</sup> [https://it.wikipedia.org/wiki/Test\\_di\\_Kolmogorov-Smirnov](https://it.wikipedia.org/wiki/Test_di_Kolmogorov-Smirnov).

Più grande è la statistica del test, maggiore è la differenza tra le due distribuzioni che vengono confrontate.<sup>10</sup>

Esistono diverse varianti del test di Kolmogorov-Smirnov, ognuna con un diverso scopo.

Nel test a un campione, l'ipotesi nulla ( $H_0$ ) è che il campione provenga da una specifica distribuzione di riferimento. Se la differenza massima tra la funzione di distribuzione cumulativa osservata e quella teorica è inferiore al valore critico, l'ipotesi nulla non viene rifiutata. In caso contrario, se la differenza è maggiore del valore critico, l'ipotesi nulla viene rifiutata. Ciò significa che se l'ipotesi nulla viene rifiutata, si può concludere che il campione non proviene dalla distribuzione di riferimento [isolution.pro](http://isolution.pro).

Nel test a due campioni, l'ipotesi nulla ( $H_0$ ) è che i due campioni provengano dalla stessa distribuzione. Se la differenza massima tra le funzioni di distribuzione cumulativa dei due campioni è inferiore al valore critico, l'ipotesi nulla non viene rifiutata. In caso contrario, se la differenza è maggiore del valore critico, l'ipotesi nulla viene rifiutata. Ciò significa che se l'ipotesi nulla viene rifiutata, si può concludere che i due campioni non provengono dalla stessa distribuzione.

Il test di adattamento (*goodness-of-fit*) viene utilizzato per verificare se un campione proviene da una distribuzione specificata o teorica.

Lo scopo del test di Kolmogorov-Smirnov è determinare se due distribuzioni differiscono, o se una distribuzione di probabilità sottostante differisce da una distribuzione ipotizzata. Viene utilizzato

---

<sup>10</sup> <https://dataaspirant.com/kolmogorov-smirnov-test/>

quando abbiamo due campioni provenienti da due popolazioni che possono essere diverse<sup>11</sup>.

Nell'uso comune, il test di Kolmogorov-Smirnov è spesso utilizzato per verificare se una data  $F(x)$  è la distribuzione di probabilità sottostante di  $F_n(x)$ . In altre parole, viene utilizzato per decidere se la distribuzione osservata differisce significativamente dalla distribuzione di popolazione specificata<sup>12</sup>.

## 2.3 Campione e periodo di riferimento

Nella presente sezione, saranno esaminati due aspetti fondamentali del processo di ricerca: la selezione del campione e la definizione del periodo di riferimento. Questi elementi rivestono un ruolo determinante nell'esecuzione di un'analisi accurata dei dati relativi al minutaggio degli *under-21*.

### 2.3.1 Selezione del campione

La selezione del campione rappresenta il primo passo in qualsiasi studio statistico. Nel contesto di questa ricerca, il campione è composto dai dati concernenti il minutaggio dei giocatori di serie *A under-21*. Tuttavia, raccogliere semplicemente questi dati non è sufficiente; è necessario eseguire operazioni di pulizia e preparazione per garantire che il campione sia rappresentativo e privo di valori mancanti o dati non validi. Questa fase di selezione e pulizia dei dati è cruciale per ottenere risultati affidabili e significativi.

---

<sup>11</sup> [https://link.springer.com/referenceworkentry/10.1007/978-0-387-32833-1\\_214](https://link.springer.com/referenceworkentry/10.1007/978-0-387-32833-1_214)

<sup>12</sup> <https://ncbi.nlm.nih.gov/pmc/articles/PMC6964527/>



In particolare, il processo di selezione del campione ha coinvolto:

- Caricamento dei dati: I dati sono stati estratti da un *file Excel* utilizzando le librerie "*readxl*" e "*openxlsx*". Questa operazione è stata fondamentale per acquisire il dataset di partenza.
- Gestione dei valori mancanti: Nel corso del processo, è stata data particolare attenzione alla gestione dei valori mancanti (NA), che potrebbero influire negativamente sull'analisi. Al fine di affrontare questa situazione, sono stati attribuiti zeri al minutaggio nei casi appropriati.
- Creazione di un nuovo *dataframe*: È stato creato un nuovo *dataframe* denominato "*db\_analisi*," il quale ha mantenuto i medesimi dati del *dataset* iniziale. Questo nuovo *dataframe* ha costituito una struttura coesa per le analisi successive.

### **2.3.2 Definizione del periodo di riferimento**

La definizione del periodo di riferimento rappresenta un aspetto cruciale della ricerca. Questo parametro specifica l'intervallo temporale su cui si concentreranno le analisi. Nel caso specifico, il periodo di riferimento è stato delineato attraverso le diverse stagioni calcistiche. Ciascuna stagione è caratterizzata da numero di squadre e da variazioni nei minutaggi dei giocatori.

Una corretta definizione del periodo di riferimento è di importanza fondamentale per comprendere come le variabili si evolvano nel tempo e per rispondere con precisione alle domande di ricerca. Gli strumenti statistici impiegati nelle analisi successive sono strettamente legati alla definizione del

periodo di riferimento, poiché le dinamiche del minutaggio dei giocatori possono variare in modo significativo da una stagione all'altra.

Il campione analizzato copre il periodo dall'1988/1989 al 2022/23 ed è stato suddiviso in tre gruppi distinti:

- gruppo 1: 12 stagioni, dal 1988/1989 al 1999/2000
- gruppo 2: 12 stagioni, dal 2000/2001 al 2012/2013
- gruppo 3: 11 stagioni, dal 2013/2014 al 2022/2023

Questo tipo di suddivisione offre una panoramica chiara della distribuzione dei minuti giocati in ciascun gruppo di stagioni, consentendo un'analisi più dettagliata dei dati.

### III Analisi dei dati

#### 3.1 Caricamento delle librerie e inizializzazione del *dataset*

In questa sezione vengono introdotti i passaggi chiave per preparare il terreno per l'analisi dei dati. Questa fase iniziale risulta cruciale al fine di assicurare che i dati siano pronti per l'elaborazione e l'interpretazione.

```
1 library(readxl)
2 library(openxlsx)
3 # Carica le librerie necessarie, se non sono già caricate
4 if (!require("ggplot2")) {
5   install.packages("ggplot2")
6   library(ggplot2)
7 }
8 if (!require("dplyr")) {
9   install.packages("dplyr")
10  library(dplyr)
11 }
12 if (!require("stats")) {
13   install.packages("stats")
14   library(stats)
15 }
16 # Legge i dati da un file Excel
17 dati <-
18   read.xlsx("C:/Users/chiar/OneDrive/Documenti/copiatesiDatiCorretti_statisticheGiocatori.xlsx", sheet = "under21")
19 # Gestisce i valori NA
20 dati$Min <- ifelse(is.na(dati$Min), 0, dati$Min)
21 dati[is.na(dati)] <- 0
22 #creo un nuovo dataframe chiamato db_analisi, in cui setto gli stessi dati del db
23 #dati che ho caricato dal file excel per effettuare le operazioni
24 db_analisi=dati
25 db_analisi=db_analisi[complete.cases(db_analisi),] #controllo che tutte le righe abbiano dei valori validi
26
```

Figura 1: Caricamento librerie e inizializzazione dataset

Il processo ha inizio con il caricamento delle librerie necessarie per l'analisi dati. In tal senso, come mostrato nella figura 1, vengono caricate le librerie "readxl" e "openxlsx" al fine di consentire la lettura dei dati da un file Excel. Successivamente, si procede a un controllo per verificare che altre librerie fondamentali quali "ggplot2," "dplyr," e "stats" siano presenti. Nel caso in cui ciò non fosse verificato, vengono installate e caricate per garantire la disponibilità di tutte le funzionalità richieste.

Dopo aver predisposto l'ambiente di sviluppo, si procede alla lettura dei dati dal *file Excel* specificato. I dati vengono estratti dal foglio "*under-21*" del documento "*copiatesiDatiCorretti\_statisticheGiocatori.xlsx*". Questo passaggio risulta cruciale, poiché consente di ottenere i dati di partenza necessari per l'analisi.

Per garantire la coerenza dei dati, vengono gestiti i valori mancanti presenti nella colonna "*Min*". I valori NA vengono sostituiti con il valore 0, garantendo così che i dati siano completi e pronti per l'analisi successiva.

Infine, per mantenere intatto il *dataframe* originale, si procede con la creazione di un nuovo *dataframe* denominato "*db\_analisi*," che contiene gli stessi dati. Tale passaggio risulta utile, poiché consente di eseguire operazioni successive nell'ambito dell'analisi dei dati senza influire direttamente sui dati originali.

Inoltre, si effettua un controllo per verificare che tutte le righe del *dataframe* "*db\_analisi*" siano prive di valori mancanti, garantendo la qualità dei dati iniziali.

	Stagione	Squadra	Giocatore	Nazione	Ruolo	Età	PG	Tit	Min	90.min
1	1988/1989	Inter	Nicola Berti	it ITA	Cen	21	32	32	2775	30.8
2	1988/1989	Inter	Dario Morello	it ITA	Att,Cen	20	10	0	195	2.2
3	1988/1989	Inter	Alberto Rivolta	it ITA	Dif	20	1	0	12	0.1
4	1988/1989	Inter	Pasquale Rocco	it ITA	Cen	17	1	0	1	0.0
5	1988/1989	Inter	Gabriele Baraldi	it ITA	Dif	19	0	0	0	0.0
6	1988/1989	Inter	Luca Mondini	it ITA	Por	18	0	0	0	0.0
7	1988/1989	Inter	Raffaele Paolino	it ITA	Cen	19	0	0	0	0.0
8	1988/1989	Napoli	Ciro Ferrara	it ITA	Dif	21	27	27	2307	25.6
9	1988/1989	Napoli	Massimo Filardi	it ITA	Dif	21	12	4	515	5.7
10	1988/1989	Napoli	Simone Giacchetta	it ITA	Dif	19	3	1	149	1.7
11	1988/1989	Napoli	Antonio Bucciarelli	it ITA	Cen	17	2	1	111	1.2
12	1988/1989	Napoli	Marco Ferrante	it ITA	Att	17	1	1	90	1.0
13	1988/1989	Napoli	Cosimo Portaluri	it ITA	Dif	18	2	0	38	0.4
14	1988/1989	Napoli	Giovanni Di Rocco	it ITA	Dif	17	2	0	32	0.4
15	1988/1989	Napoli	Enrico Buonocore	it ITA	0	17	0	0	0	0.0
16	1988/1989	Napoli	Patrizio Chiaiese	0	0	21	0	0	0	0.0
17	1988/1989	Napoli	Mirco Marinelli	0	0	18	0	0	0	0.0
18	1988/1989	Milan	Paolo Maldini	it ITA	Dif	20	26	26	2259	25.1
19	1988/1989	Milan	Graziano Mannari	it ITA	Att	19	17	8	803	8.9
20	1988/1989	Milan	Christian Lantignotti	it ITA	Cen	18	8	2	354	3.9
21	1988/1989	Milan	Fabio Viviani	it ITA	Cen	21	6	1	158	1.8
22	1988/1989	Milan	Massimiliano Cappellini	it ITA	Att	17	3	1	100	1.1
23	1988/1989	Milan	Demetrio Albertini	it ITA	Att,Cen	16	1	0	12	0.1
24	1988/1989	Milan	Francesco Antonioli	it ITA	Por	18	0	0	0	0.0
25	1988/1989	Milan	Davide Corti	0	0	16	0	0	0	0.0
26	1988/1989	Juventus	Renato Buso	it ITA	Cen	18	24	14	1458	16.2
27	1988/1989	Juventus	Salvatore Avallone	it ITA	Cen	18	0	0	0	0.0
28	1988/1989	Juventus	Cristiano Bellucci	0	0	17	0	0	0	0.0
29	1988/1989	Juventus	Federico Giampaolo	it ITA	Att	18	0	0	0	0.0
30	1988/1989	Juventus	Hugo Rubini	ar ARG	Por	18	0	0	0	0.0
31	1988/1989	Sampdoria	Gianluca Pagliuca	it ITA	Por	21	33	33	2970	33.0
32	1988/1989	Sampdoria	Marco Lanna	it ITA	Dif	20	19	18	1572	17.5
33	1988/1989	Sampdoria	Stefano Pellegrini	it ITA	Dif	21	13	7	708	7.9

Showing 1 to 34 of 2,820 entries, 10 total columns

Figura 2: db\_analisi

Nella figura 2, è presente una sezione del database completo in cui non sono presenti valori nulli. L'indicazione 'Showing 1 to 34 of 2,820 entries' evidenzia che la visualizzazione attuale include solo una parte dei dati, composta da 34 righe e 10 colonne. Tuttavia, va notato che il dataset completo contiene un totale di 2,820 righe di dati.

Durante la fase di elaborazione dei dati, è stato inizializzato un vettore vuoto

denominato "num\_squadre". Successivamente, è stato creato un nuovo *dataframe* denominato "nuovo\_db"(figura3), contenente le informazioni sulle stagioni uniche estratte dal *dataframe* originale "db\_analisi". La colonna "num\_squadre" di "nuovo\_db" è stata popolata con il numero corrispondente di squadre per ogni stagione. Quest'ultimo è stato ottenuto attraverso la replicazione del valore 18 per le prime 16 stagioni e del valore 20 per le successive 19, riflettendo la variazione del numero di squadre da 18 a 20 a partire dal 2004. In tal modo, "nuovo\_db" offre un quadro sintetico delle stagioni e del numero di squadre associato a ciascuna di esse nell'analisi dei dati.

	stagioni	num_squadre
1	1988/1989	18
2	1989/1990	18
3	1990/1991	18
4	1991/1992	18
5	1992/1993	18
6	1993/1994	18
7	1994/1995	18
8	1995/1996	18
9	1996/1997	18
10	1997/1998	18
11	1998/1999	18
12	1999/2000	18
13	2000/2001	18
14	2001/2002	18
15	2002/2003	18
16	2003/2004	18
17	2004/2005	20
18	2005/2006	20
19	2006/2007	20
20	2007/2008	20
21	2008/2009	20
22	2009/2010	20
23	2010/2011	20
24	2011/2012	20
25	2012/2013	20
26	2013/2014	20
27	2014/2015	20
28	2015/2016	20
29	2016/2017	20
30	2017/2018	20
31	2018/2019	20
32	2019/2020	20
33	2020/2021	20
34	2021/2022	20
35	2022/2023	20

Figura 3:nuovo\_db

## 3.2 Normalizzazione dei dati

La decisione di normalizzare i dati è stata motivata dalla presenza di un numero variabile di squadre nelle diverse stagioni, variando da 18 a 20. Questa normalizzazione mira a garantire coerenza e uniformità nei minuti giocati dagli *under-21*, consentendo una comparazione accurata tra le stagioni.

L'obiettivo principale di questa procedura è rendere i dati comparabili e interpretabili in modo significativo, tenendo debitamente conto delle variazioni stagionali e delle fluttuazioni nel numero di squadre partecipanti, ognuna con il proprio minutaggio associato.

È stato generato un nuovo *dataframe* denominato "*min\_tot\_per\_stagione*", progettato per associare a ciascuna stagione il totale dei minuti giocati dagli under 21. L'operazione implica l'utilizzo della funzione *group\_by(Stagione)* per raggruppare i dati in base alle stagioni e della funzione *summarise(min\_tot = sum(Min))* per calcolare la somma dei minuti giocati in ciascuna stagione. Tale procedura consente di ottenere una chiara rappresentazione della quantità complessiva di tempo dedicato sul campo dagli *under-21* in ogni stagione.

Successivamente, i risultati della fase precedente sono sfruttati per arricchire il *dataframe* principale, noto come "*nuovo\_db*".

Vengono introdotte due nuove colonne:

- "*min\_totali\_per\_stagione*", associata al totale dei minuti giocati dagli under 21 in ciascuna stagione nel *dataframe* principale, offrendo una visione panoramica dei dati relativi ai minuti totali giocati in ogni stagione.
- La seconda colonna, "*min\_tot\_normalizzati*", si ottiene normalizzando i minuti totali per stagione rispetto al numero di squadre coinvolte in quella specifica

stagione. Tale procedimento tiene conto delle variazioni di scala dovute al numero di squadre, consentendo una comparazione più accurata dei dati.

	stagioni	num_squadre	min_totali_per_stagione	min_tot_normalizzati
1	1988/1989	18	70962	3942.333
2	1989/1990	18	52011	2889.500
3	1990/1991	18	68085	3782.500
4	1991/1992	18	58143	3230.167
5	1992/1993	18	55579	3087.722
6	1993/1994	18	60508	3361.556
7	1994/1995	18	54325	3018.056
8	1995/1996	18	67165	3731.389
9	1996/1997	18	50879	2826.611
10	1997/1998	18	55474	3081.889
11	1998/1999	18	72389	4021.611
12	1999/2000	18	52561	2920.056
13	2000/2001	18	59043	3280.167
14	2001/2002	18	49247	2735.944
15	2002/2003	18	50514	2806.333
16	2003/2004	18	59947	3330.389
17	2004/2005	20	69625	3481.250
18	2005/2006	20	58198	2909.900
19	2006/2007	20	49641	2482.050
20	2007/2008	20	60863	3043.150
21	2008/2009	20	71149	3557.450
22	2009/2010	20	49069	2453.450
23	2010/2011	20	47754	2387.700
24	2011/2012	20	50190	2509.500
25	2012/2013	20	82192	4109.600
26	2013/2014	20	77797	3889.850
27	2014/2015	20	67489	3374.450
28	2015/2016	20	68310	3415.500
29	2016/2017	20	91643	4582.150
30	2017/2018	20	68289	3414.450
31	2018/2019	20	95280	4764.000
32	2019/2020	20	75679	3783.950
33	2020/2021	20	75108	3755.400
34	2021/2022	20	76035	3801.750
35	2022/2023	20	64054	3202.700

*Figura 4:nuovo\_db aggiornato*



### 3.3 Analisi dei grafici delle singole stagioni

Attraverso l'utilizzo della libreria *ggplot2* in RStudio, è stata realizzata un'analisi dettagliata dei minuti giocati dagli *under-21* in diverse stagioni.

Sono stati generati tre grafici distinti che offrono una visione approfondita dell'andamento stagionale.

L'obiettivo di questa analisi visiva è consentire una comprensione più chiara e dettagliata della distribuzione del minutaggio, evidenziando eventuali tendenze o variazioni nel corso delle singole stagioni.

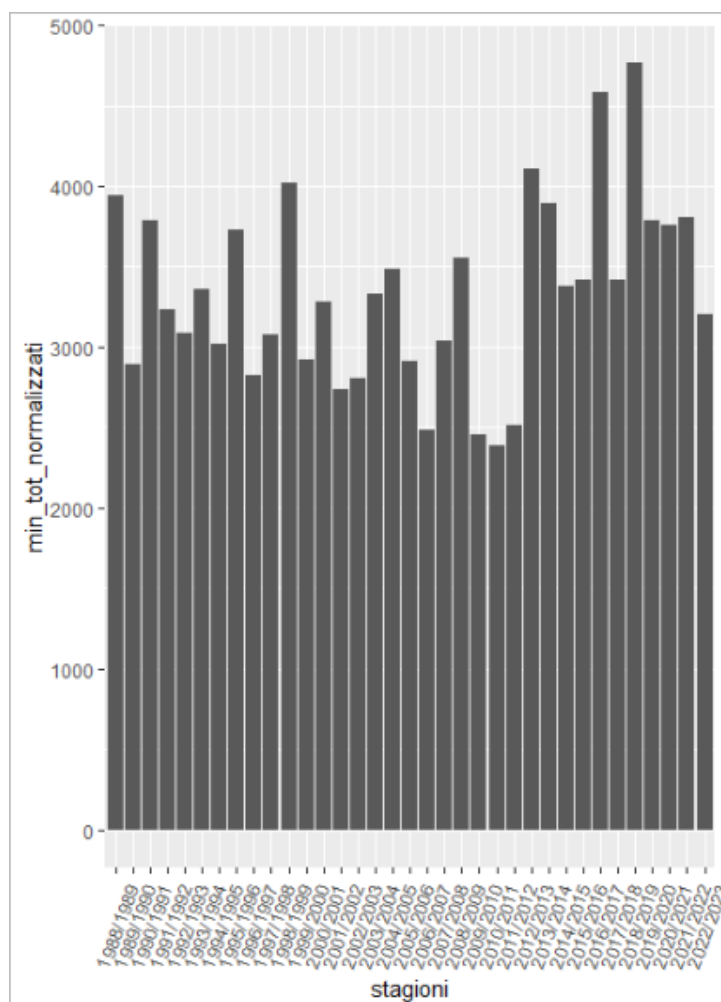


Figura 5: grafico 1

Il primo grafico (**figura 5**) rappresenta un istogramma che illustra la distribuzione dei minuti normalizzati giocati dagli *under-21* nelle diverse stagioni del campionato di Serie A. Sull'asse x sono riportate le singole stagioni, mentre sull'asse y sono indicati i minuti normalizzati. Ciascuna barra verticale corrisponde a una stagione specifica, evidenziando il quantitativo di minuti giocati dagli *under-21* in quel periodo, considerando la normalizzazione per adeguarsi alle variazioni nel numero di squadre partecipanti. Le etichette delle stagioni sull'asse x sono ruotate di 70 gradi per garantire una migliore leggibilità del testo.

Questo grafico fornisce una panoramica chiara delle fluttuazioni nei minuti giocati dagli under 21 nelle diverse stagioni, offrendo un'analisi visiva dettagliata

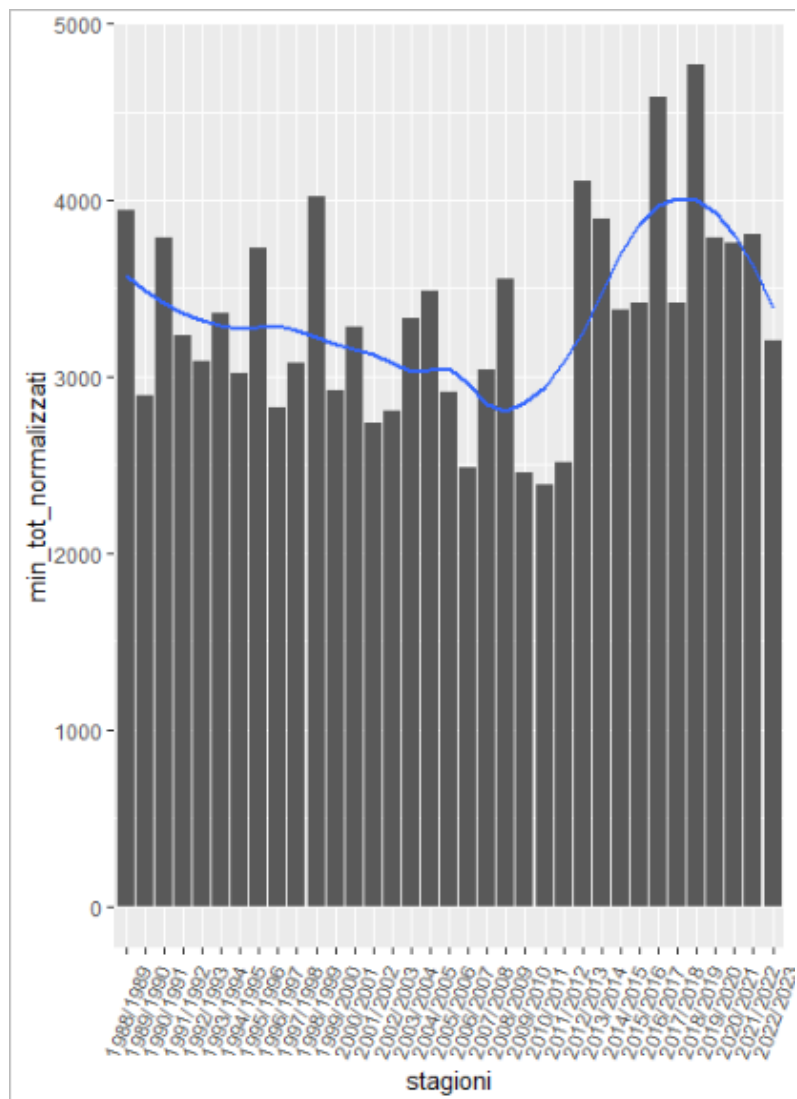


Figura 6: grafico 2

Il secondo grafico (figura 6), una versione ampliata del primo, include una curva di regressione polinomiale, modellando la relazione tra le variabili. Questa curva riflette come i minuti giocati dagli *under-21* variano stagione dopo stagione. L'inclinazione della curva indica la presenza di una tendenza positiva (aumento dei minuti giocati) o negativa (diminuzione).

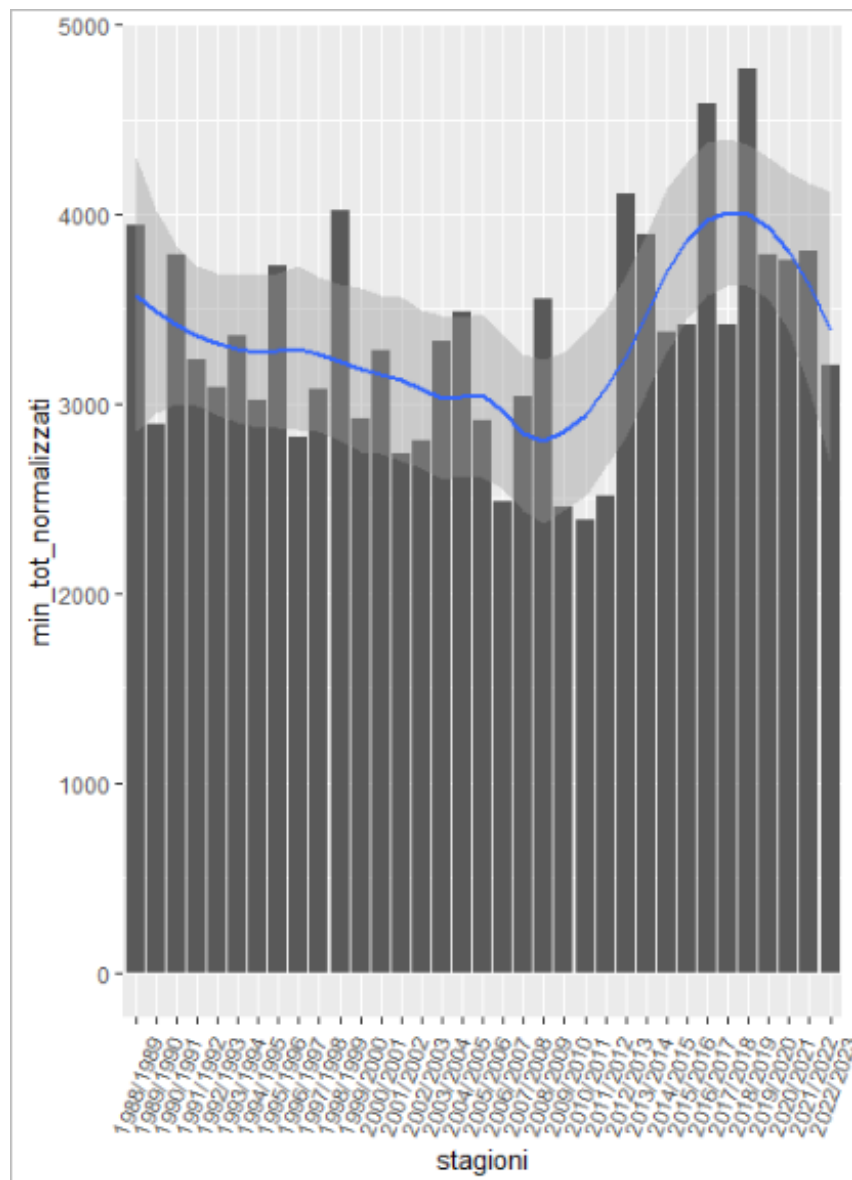


Figura 7: grafico 3

Nel terzo grafico (figura 7), la curva di regressione lineare è sovrapposta all'istogramma dei minuti normalizzati dagli *under-21*. L'opzione `se = TRUE` evidenzia l'intervallo di confidenza, fornendo una stima della precisione della regressione. Questo intervallo è cruciale per interpretare l'incertezza nei valori stimati, offrendo una prospettiva approfondita sulle tendenze dei minuti giocati dagli *under-21* nel corso delle stagioni.

### 3.4 Preparazione ai test

La preparazione dei dati per il test di Kolmogorov-Smirnov riveste un ruolo cruciale nell'analisi statistica condotta.

Nel processo di preparazione, i dati sono esportati in un nuovo *dataset* (figura 9) con tre colonne significative.

La colonna "Stagioni" indica le diverse annate in cui sono stati raccolti i dati, mentre la colonna "*min\_tot\_normalizzati*" rappresenta i minuti totali normalizzati associati a ciascuna stagione.

La suddivisione delle stagioni in "Gruppi" è particolarmente rilevante, poiché consente di creare tre distinti campioni di dati. Questa categorizzazione è fondamentale per condurre analisi specifiche su ogni gruppo di dati, soprattutto considerando l'obiettivo di ottenere tre classi di annate di minutaggio normalizzato.

	Stagioni	min_tot_normalizzati	gruppi
1	1988/1989	3942.333	1
2	1989/1990	2889.500	1
3	1990/1991	3782.500	1
4	1991/1992	3230.167	1
5	1992/1993	3087.722	1
6	1993/1994	3361.556	1
7	1994/1995	3018.056	1
8	1995/1996	3731.389	1
9	1996/1997	2826.611	1
10	1997/1998	3081.889	1
11	1998/1999	4021.611	1
12	1999/2000	2920.056	1
13	2000/2001	3280.167	2
14	2001/2002	2735.944	2
15	2002/2003	2806.333	2
16	2003/2004	3330.389	2
17	2004/2005	3481.250	2
18	2005/2006	2909.900	2
19	2006/2007	2482.050	2
20	2007/2008	3043.150	2
21	2008/2009	3557.450	2
22	2009/2010	2453.450	2
23	2010/2011	2387.700	2
24	2011/2012	2509.500	2
25	2012/2013	4109.600	3
26	2013/2014	3889.850	3
27	2014/2015	3374.450	3
28	2015/2016	3415.500	3
29	2016/2017	4582.150	3
30	2017/2018	3414.450	3
31	2018/2019	4764.000	3
32	2019/2020	3783.950	3
33	2020/2021	3755.400	3
34	2021/2022	3801.750	3
35	2022/2023	3202.700	3

*Figura 8: Stagioni con i relativi gruppi e minutaggio normalizzato*

Nel contesto dell'analisi condotta, è stato generato un istogramma (figura 9) che presenta la somma normalizzata dei minuti giocati per stagione in un formato aggregato.

Questo grafico è finalizzato a evidenziare la tendenza generale osservata nei dati di ciascun gruppo stagionale. La rappresentazione visiva fornisce un'istantanea chiara e

intuitiva della distribuzione complessiva dei minuti giocati dagli *under-21* nelle diverse stagioni considerate.

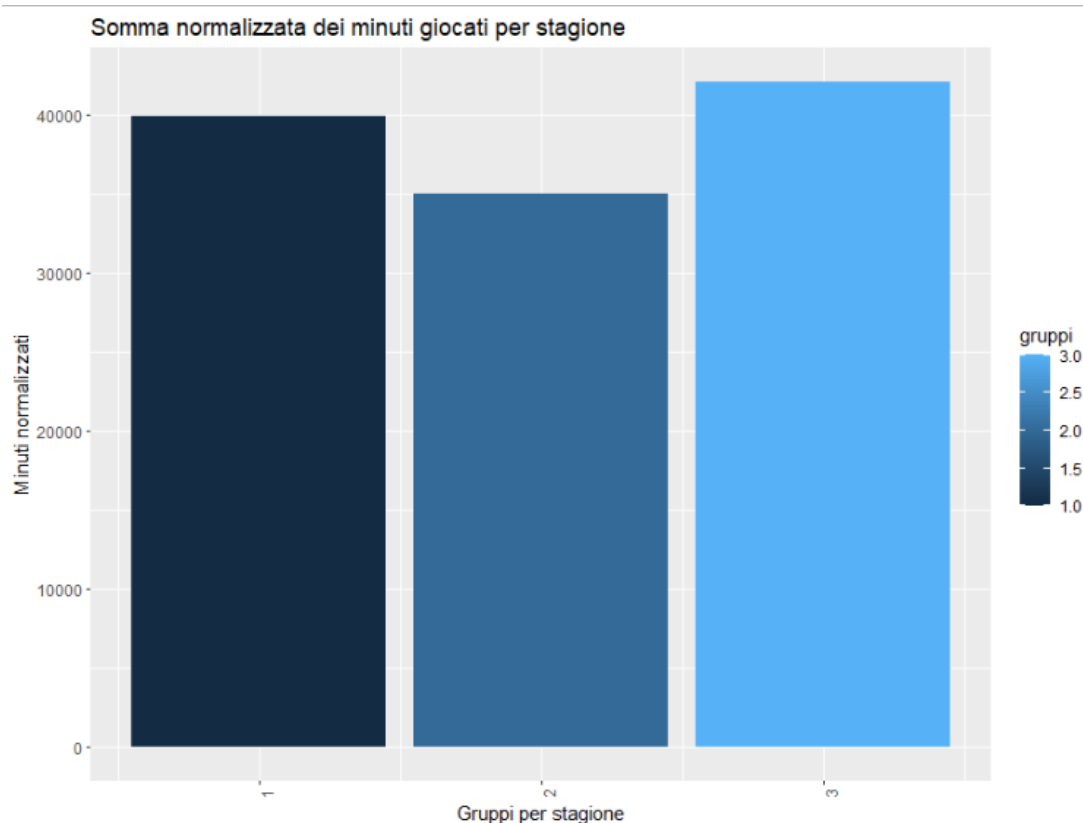


Figura 9: istogramma Somma aggregata minuti normalizzati (grafico 4)

L'istogramma mostra la distribuzione dei minuti giocati in ciascun gruppo stagionale. I tre gruppi considerati riguardano le annate dal 1988/1989 al 2022/2023:

- Gruppo 1: comprende 12 stagioni, dal 1988/1989 al 1999/2000, con una somma normalizzata di minuti giocati pari a 39.893,39.
- Gruppo 2: si estende su 12 stagioni, dal 2000/2001 al 2012/2013, con una somma normalizzata di minuti giocati pari a 34.977,28.
- Gruppo 3: include 11 stagioni, dal 2013/2014 al 2022/2023, con una somma normalizzata di minuti giocati pari a 42.093,8.

Nell'analisi esplorativa dei dati relativi ai minuti giocati dagli *under-21* nelle diverse stagioni di ciascun gruppo, è stato impiegato il *boxplot* (figura 10) come strumento visuale per la rappresentazione della distribuzione, oltreché per fornire informazioni cruciali, tra cui le mediane, i quartili e la presenza di eventuali valori anomali.

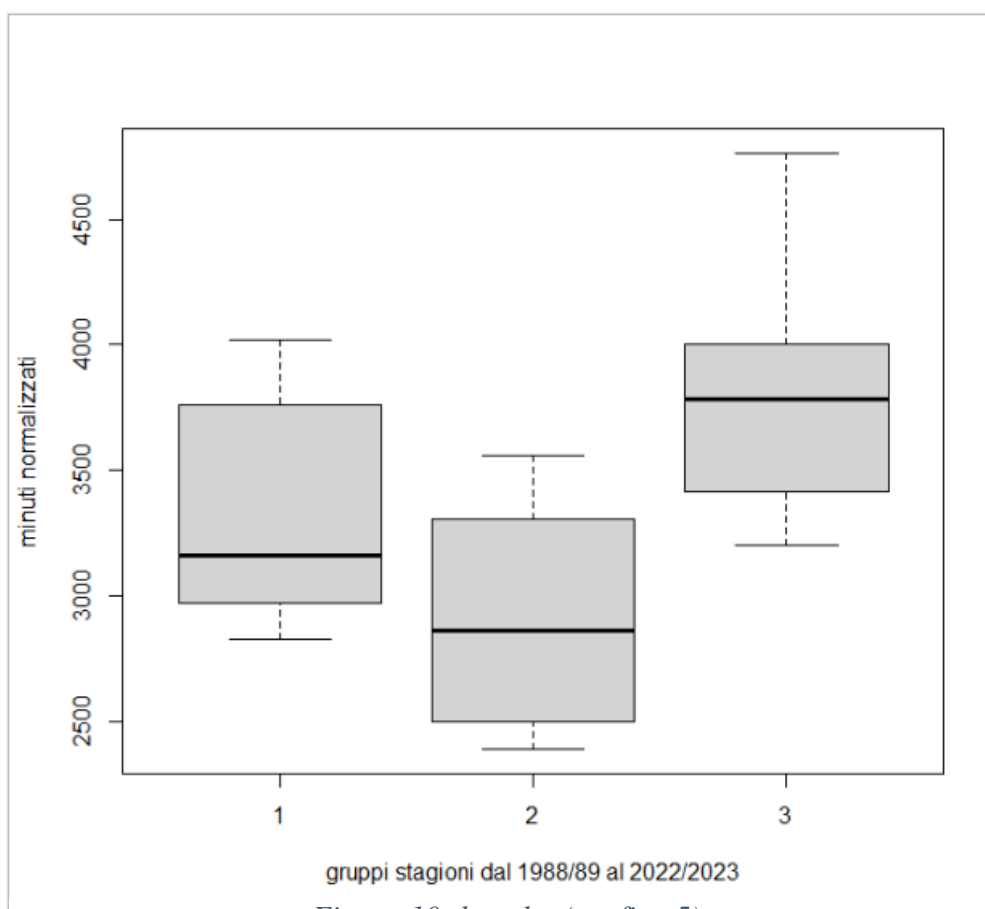


Figura 10: *boxplot* (grafico 5)



```

> summary(gruppo1)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 2827  2994   3159   3324   3744   4022
> summary(gruppo2)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 2388  2503   2858   2915   3293   3557
> summary(gruppo3)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 3203  3415   3784   3827   4000   4764

```

Figura 11: valori boxplot

L'analisi del *boxplot* nel Gruppo 1, che comprende le stagioni dal 1988/1989 al 1999/2000, rileva una distribuzione dei minuti giocati dagli *under-21* caratterizzata da una mediana di 3159 minuti (figura 11). Il *range*, che varia da 2827 a 4022 minuti, sottolinea una considerevole diversità nelle opportunità di gioco per i giovani calciatori durante questo periodo. La presenza di questo ampio intervallo indica una notevole varietà nella partecipazione degli *under-21*, con alcuni giocatori che hanno accumulato un notevole minutaggio rispetto ad altri.

Nel Gruppo 2, che spazia dalle stagioni 2000/2001 al 2012/2013, la mediana di 2858 minuti suggerisce una quantità tipica leggermente inferiore rispetto al Gruppo 1. Il *range*, compreso tra 2388 e 3557 minuti, evidenzia una distribuzione dei minuti giocati dagli *under-21* con una certa variabilità. In questo periodo, la partecipazione dei giovani calciatori ha mantenuto una diversità considerevole, ma con una tendenza a una tipica quantità leggermente più bassa rispetto al decennio precedente.

Per quanto riguarda il Gruppo 3 (stagioni 2013/2014 al 2022/2023), la mediana di 3784 minuti indica una notevole crescita nella quantità tipica di gioco rispetto ai due gruppi precedenti. Il *range*, che va da 3203 a 4764 minuti, mette in luce un'ampia variabilità nelle opportunità di gioco per gli *under-21*. Questi risultati suggeriscono un aumento significativo del minutaggio medio durante questo periodo, indicando una maggiore considerazione dei giovani talenti nel contesto calcistico.

In seguito, è stato eseguito un calcolo delle medie normalizzate per tre distinti gruppi di dati sugli *under-21*. Il procedimento prevede l'aggregazione del minutaggio normalizzato per ciascun gruppo, seguita da una normalizzazione di questa somma, ottenuta dividendo per il numero di elementi nel gruppo. Tale processo è stato applicato separatamente a ciascun gruppo al fine di calcolare le rispettive medie normalizzate dei minuti giocati dagli *under-21*.

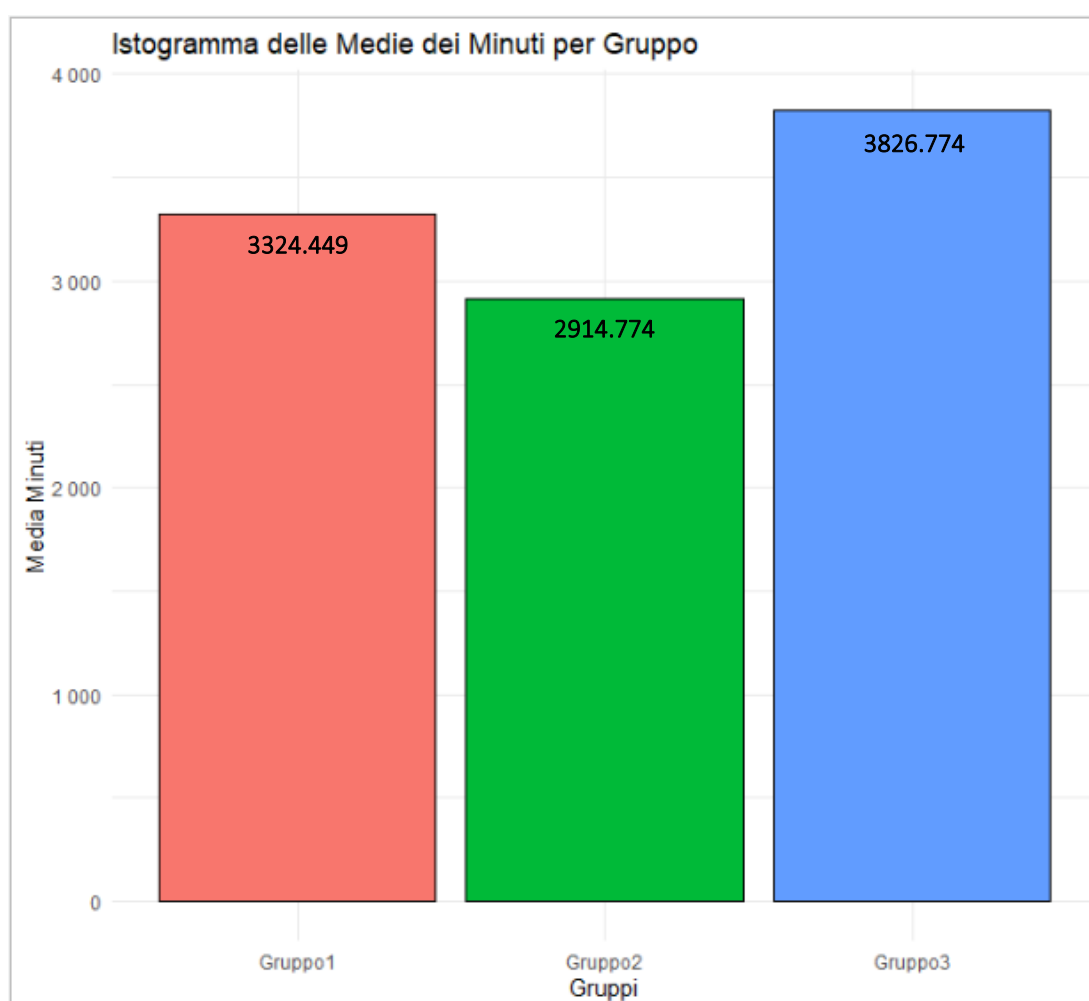


Figura 12: Media dei minuti per gruppo (grafico 6)

I risultati di questa analisi sono stati successivamente rappresentati attraverso un istogramma denominato nella figura 12.

Questo grafico visualizza chiaramente le medie normalizzate dei minuti giocati per ogni gruppo. Le colonne colorate distintive rappresentano i vari gruppi, facilitando una comparazione visiva delle medie tra Gruppo 1, Gruppo 2 e Gruppo 3.

L'asse x del grafico indica i nomi dei gruppi, mentre l'asse y rappresenta le medie normalizzate, offrendo una visione intuitiva della distribuzione del minutaggio normalizzato negli specifici periodi temporali considerati.

### 3.5 Test Kolmogorov-Smirnov

In questo studio sono stati confrontati due gruppi di campioni di minuti giocati da giocatori under 21 in diverse stagioni di Serie A. L'obiettivo era di valutare se ci fosse una differenza significativa tra le distribuzioni di minuti giocati nei due gruppi.

#### 3.5.1 gruppo1 vs gruppo 2: Test a una coda

```
Two-sample Kolmogorov-Smirnov test
data:  gruppo1 and gruppo2
D^- = 0.5, p-value = 0.04979
```

#### IPOTESI

- Ipotesi Nulla (H0): La distribuzione dei minuti giocati nel Gruppo 1 è inferiore o uguale a quella del Gruppo 2.
- Ipotesi Alternativa(H1): La distribuzione dei minuti giocati nel Gruppo 1 è maggiore di quella del Gruppo 2.

Ho condotto un test di Kolmogorov-Smirnov unilaterale a coda sinistra per verificare questa ipotesi.

## RISULTATO

Data: gruppo1 e gruppo2 rappresentano i minuti giocati nei due gruppi.

$D^- = 0.5$ : Questo è il valore statisticamente calcolato di  $D^-$ . Indica la massima differenza tra le due funzioni cumulative di distribuzione (CDF).

$p\text{-value} = 0.04979$ : Il  $p\text{-value}$  rappresenta la probabilità che l'osservazione di  $D^-$  sia casuale. In questo caso, il  $p\text{-value}$  è 0.04979, che è inferiore al livello di significatività comune del 5% (0.05).

## CONCLUSIONI

Sono presenti evidenze statistiche a favore del rifiuto dell'ipotesi nulla ( $H_0$ ).

Viene dimostrato che la distribuzione dei minuti giocati nel Gruppo 1 è maggiore di quella del Gruppo 2 a un livello di significatività del 5%. Ciò indica che c'è una differenza significativa tra i due gruppi in termini di minuti giocati, con il Gruppo 1 che ha una distribuzione più ampia o con minuti mediamente superiori rispetto al Gruppo 2.

### 3.5.2 gruppo3 vs gruppo 2: Test a una coda

```
Two-sample Kolmogorov-Smirnov test
data:  gruppo3 and gruppo2
D^- = 0.74242, p-value = 0.001788
```

## **IPOTESI**

- Ipotesi Nulla (H0): La distribuzione dei minuti giocati nel "Gruppo 3" è inferiore o uguale a quella del "Gruppo 2".
- Ipotesi Alternativa (H1): La distribuzione dei minuti giocati nel "Gruppo 3" è maggiore di quella del "Gruppo 2".

Ho condotto un test di Kolmogorov-Smirnov unilaterale a coda sinistra per verificare questa ipotesi.

## **RISULTATO**

Data: "gruppo3" e "gruppo2" rappresentano i minuti giocati nei due gruppi.

$D^- = 0.74242$ : Questo è il valore statisticamente calcolato di  $D^-$ , la massima differenza tra le due funzioni cumulative di distribuzione (CDF).

$p\text{-value} = 0.001788$ : Il  $p\text{-value}$  rappresenta la probabilità che l'osservazione di  $D^-$  sia casuale. In questo caso, il  $p\text{-value}$  è 0.001788, che è molto inferiore al livello di significatività comune del 5% (0.05).

## **CONCLUSIONI**

Sono presenti evidenze statistiche a favore del rifiuto dell'ipotesi nulla (H0).

Sulla base dei risultati del test di Kolmogorov-Smirnov, si può affermare con evidenza statistica che la distribuzione dei minuti giocati nel "Gruppo 3" è significativamente superiore a quella del "Gruppo 2" a un livello di significatività del 5%. Questo risultato suggerisce che esiste una differenza sostanziale tra i due gruppi in termini di minuti giocati, con il "Gruppo 3" che presenta una distribuzione dei minuti giocati più ampia o valori mediamente superiori rispetto al "Gruppo 2".

Queste conclusioni rafforzano l'ipotesi alternativa (H1) che indicava che il "Gruppo 3" avrebbe una distribuzione superiore a quella del "Gruppo 2". L'analisi dei dati fornisce un supporto statistico significativo a questa affermazione, suggerendo che i minuti giocati nel "Gruppo 3" sono in effetti più elevati rispetto al "Gruppo 2" con un alto grado di certezza.

### 3.5.3 gruppo3 vs gruppo 1: Test a due code

```
Two-sample Kolmogorov-Smirnov test  
  
data:  gruppo1 and gruppo3  
D = 0.57576, p-value = 0.02346  
  
alternative hypothesis: two-sided
```

#### **IPOTESI**

- Ipotesi Nulla (H0): L'ipotesi nulla afferma che le due distribuzioni di dati sono identiche, indicando l'assenza di differenze significative tra le distribuzioni dei due gruppi.
- Ipotesi Alternativa (H1): L'ipotesi alternativa afferma che le due distribuzioni non sono identiche, suggerendo la presenza di differenze significative tra le distribuzioni.

Ho condotto un test di Kolmogorov-Smirnov bilaterale per verificare queste ipotesi. Questo significa che sto cercando differenze in entrambe le direzioni (superiore e inferiore). Pertanto, l'evidenza di differenze significative potrebbe indicare che uno dei gruppi ha una distribuzione diversa dall'altro, indipendentemente dalla direzione specifica delle differenze.

## **RISULTATO**

Data: "gruppo3" e "gruppo1" rappresentano i minuti giocati nei due gruppi.

$D = 0.57576$ : Questo valore rappresenta la statistica  $D$ , che misura la massima differenza tra le funzioni cumulative di distribuzione dei due gruppi.

$p\text{-value} = 0.02346$ : Il  $p\text{-value}$  rappresenta la probabilità che l'osservazione di  $D$  sia casuale. In questo caso, il  $p\text{-value}$  è  $0.02346$ , che è inferiore al livello di significatività comune del 5% (0.05).

## **CONCLUSIONI**

Sulla base dei risultati del test di Kolmogorov-Smirnov, si può affermare che ci sono evidenze statistiche a favore del rifiuto dell'ipotesi nulla ( $H_0$ ). Ciò indica che esistono differenze significative tra le distribuzioni dei minuti giocati nel "Gruppo 3" e nel "Gruppo 1". Il valore di  $D$ , che rappresenta la massima differenza tra le CDF, è  $0.57576$ , suggerendo che le due distribuzioni sono notevolmente diverse.

Questi risultati supportano l'ipotesi alternativa ( $H_1$ ) che suggerisce la presenza di differenze significative tra i due gruppi. L'analisi statistica fornisce un sostegno solido a questa affermazione, indicando che le distribuzioni del tempo trascorso nei due gruppi sono significativamente diverse in entrambe le direzioni.

### **3.6 Esito e interpretazione risultati Test Kolmogorov-Smirnov**

In sintesi, i risultati complessivi suggeriscono che i "Gruppo 1" e "Gruppo 3" hanno distribuzioni dei minuti giocati diverse tra loro e che entrambi sono significativamente superiori al "Gruppo 2".

## **IV Discussione**

### **4.1 Limiti dell'approccio**

Il presente capitolo si propone di esaminare criticamente gli aspetti limitanti dell'approccio metodologico adottato nell'analisi del minutaggio degli *under 21* nelle stagioni di Serie A.

La suddivisione in tre gruppi decennali, la scelta del periodo di studio, l'omissione della distinzione basata sulla nazionalità dei giocatori e la decisione di non cercare correlazioni con la sentenza Bosman rappresentano le fondamenta su cui si basa l'analisi, ma si ritiene cruciale comprendere i confini intrinseci di questo approccio. Attraverso una riflessione critica su tali limitazioni, si mira a fornire una prospettiva completa sull'applicabilità e sulla generalizzabilità dei risultati ottenuti, nonché a delineare le eventuali considerazioni da tenere presente nell'interpretazione dei dati. La consapevolezza di questi limiti è essenziale per una comprensione accurata delle dinamiche del minutaggio degli *under 21* in Serie A e contribuirà a una discussione più approfondita dei risultati emersi dall'analisi.

#### **4.1.1 Divisione in soli 3 gruppi**

La decisione di suddividere le circa 30 stagioni in tre gruppi decennali è stata adottata strategicamente al fine di implementare un test con tre classi di campioni. Questa scelta metodologica è stata motivata dalla volontà di condurre un'analisi più dettagliata delle dinamiche temporali, consentendo così di esaminare con precisione i cambiamenti significativi nel corso del tempo. Tale approccio ha garantito una cornice temporale sufficientemente ampia per catturare le tendenze a lungo termine, mentre la specifica divisione in tre classi ha fornito una base robusta per la valutazione delle differenze tra i periodi.



Inoltre, la decisione di analizzare due decenni alla volta attraverso il test di Kolmogorov-Smirnov è stata guidata dalla necessità di mantenere coerenza nelle dimensioni temporali dei confronti. Questo ha consentito un'analisi più dettagliata delle disparità tra i due periodi senza compromettere l'integrità statistica dell'indagine. L'adozione di questa strategia ha, quindi, facilitato una concentrazione più mirata sui confronti temporali, contribuendo in modo significativo ad una comprensione approfondita delle variazioni nel minutaggio degli *under 21* nel corso delle diverse stagioni.

#### **4.1.2 Scelta del periodo di studio**

La decisione di esaminare il periodo delle stagioni sportive di Serie A dal 1988/89 è stata guidata principalmente dall'obiettivo di condurre un'analisi approfondita degli ultimi 30 anni. In questo arco temporale, si è osservato un notevole impiego di giovani talenti da parte di numerose squadre di Serie A, rendendo il periodo particolarmente ricco di spunti per indagare sulle dinamiche di sviluppo dei giocatori emergenti.

In aggiunta, la scelta di concentrarsi su questo intervallo temporale è stata motivata dalla disponibilità di dati più affidabili e dettagliati. Questo ha consentito di garantire una base solida per l'analisi, contribuendo a una valutazione accurata delle tendenze nel minutaggio degli *under 21* e alla comprensione delle variazioni nel corso delle diverse stagioni.

#### **4.1.3 Scelta di non distinguere la nazionalità dei giocatori**

La decisione di non distinguere le nazionalità degli *under 21* nel contesto della Serie A è stata orientata dalla necessità di semplificare l'analisi.

L'introduzione di una suddivisione basata sulle nazionalità avrebbe aggiunto complessità all'elaborazione dei dati, considerando che numerosi giocatori siano in possesso della doppia cittadinanza. Due esempi emblematici sono Andrea Cambiaso e Matias Soulé.

Andrea Cambiaso, terzino sinistro della Juventus, ha scelto di rappresentare l'Italia a livello internazionale, nonostante abbia ereditato la cittadinanza argentina dal padre. D'altra parte, Matias Soulé, attualmente in prestito dalla Juventus al Frosinone, possiede la cittadinanza italiana grazie alle sue origini, ma ha optato per la rappresentanza della nazionale “Albiceleste” a livello internazionale.

Queste scelte individuali riflettono la complessità delle decisioni legate alla cittadinanza nei contesti sportivi ed evidenziano l'importanza di una visione d'insieme nell'analisi del minutaggio degli *under 21* in Serie A.

La categorizzazione basata sulla nazionalità avrebbe generato un contesto ambiguo e intricato, rendendo ardua l'interpretazione dei risultati. Attraverso la scelta di un approccio più generale, l'obiettivo è stato ottenere una panoramica del minutaggio degli *under 21*, concentrandosi sulle opportunità di gioco in Serie A indipendentemente dalla provenienza nazionale.

Questa decisione, oltre a semplificare l'analisi, mira a fornire una visione più chiara e accessibile sull'esperienza di gioco dei giovani talenti nel contesto calcistico italiano.

#### **4.1.4 Non aver cercato correlazioni con la sentenza Bosman**

Nella redazione di questo elaborato, ho scelto di non approfondire specificamente la correlazione tra la sentenza Bosman<sup>13</sup> e la nazionalità dei giocatori. Questa decisione è dovuta ad una serie di ragioni.

In primo luogo, tale pronuncia ha principalmente influenzato la libertà di movimento dei giocatori all'interno dell'Unione Europea, eliminando le restrizioni per i giocatori dell'UE nelle leghe nazionali. Poiché la mia analisi si è

---

<sup>13</sup> V. Corte di Giustizia dell'Unione Europea, sentenza 15 dicembre 1995, C-415/93

concentrata sul minutaggio degli *under 21* in Serie A, ho ritenuto che esaminare dettagliatamente la correlazione con la sentenza Bosman non fosse cruciale, considerando che i giocatori dell'UE già godono di libertà di movimento in questa competizione.

In secondo luogo, gli effetti più diretti della medesima pronuncia si sono manifestati sulla competitività delle leghe nazionali e sulla dinamica delle competizioni europee come la Champions League. Data la focalizzazione del mio studio sulle opportunità di gioco per gli *under 21* in Serie A, ho preferito concentrarmi su fattori più specifici a questa competizione, quali il minutaggio e le dinamiche di gioco.

Infine, la correlazione tra la sentenza Bosman e la nazionalità dei giocatori può risultare complessa e influenzata da numerosi fattori, rendendo difficile trarre conclusioni dirette. La presenza di giocatori con doppia cittadinanza può essere influenzata da vari elementi, tra cui le origini personali del giocatore, le opportunità di carriera e le regole di eleggibilità della FIFA. Pertanto, ho preferito mantenere la mia analisi più focalizzata su aspetti direttamente collegati agli obiettivi della ricerca, semplificando così l'analisi e migliorando la chiarezza dei risultati ottenuti.

## **4.2 Fattori che influenzano il minutaggio degli under 21**

Il minutaggio degli *under 21* è influenzato da una serie di dinamiche complesse che riflettono la natura multifattoriale delle decisioni di schieramento nel calcio professionistico.

Sono stati esaminati alcuni dei principali fattori che contribuiscono a plasmare il tempo di gioco assegnato ai giovani talenti e l'intreccio di questi elementi sottolinea la complessità delle decisioni di schieramento nel contesto calcistico di massimo livello.

### **4.2.1 Livello di fiducia del club**

Il sostegno e la fiducia di un club nei confronti dei giovani talenti emergenti possono determinare significativamente il loro minutaggio.

Ad esempio, nella stagione 2021-22, il Barcellona è riuscito a distinguersi concedendo il maggior numero di minuti in campo ai giocatori *under 21* tra tutte le squadre delle cinque principali leghe europee, evidenziando la differenza nelle politiche di squadra<sup>xvi</sup>.

### **4.2.2 Maturità fisica e psicologica**

Lo stadio di sviluppo fisico e mentale di un giocatore *under 21* può influenzare notevolmente le decisioni dell'allenatore. I giocatori più maturi fisicamente possono avere un vantaggio in termini di forza, resistenza e capacità di gestire la pressione del gioco ad alto livello, rendendoli più propensi a ottenere un maggiore minutaggio<sup>xvii</sup>.

### **4.2.3 Posizione in campo**

La posizione occupata in campo è un elemento cruciale nella definizione delle opportunità di gioco per gli *under 21*. Studi indicano che l'effetto dell'età relativa è più pronunciato in alcune posizioni, come difensori e portieri, suggerendo un'influenza delle caratteristiche fisiche e di maturazione specifiche della posizione<sup>xix</sup>.

### **4.2.4 Esperienza di gioco**

L'esperienza di gioco a livello *senior* emerge come un fattore discriminante nelle decisioni di minutaggio. I giocatori con esperienza competitiva superiore possono essere considerati maggiormente pronti per le sfide della Serie A, guadagnandosi la fiducia degli allenatori nella gestione delle partite<sup>xx</sup>.

#### **4.2.5 Prestazioni durante l'allenamento**

Le prestazioni durante le sessioni di allenamento giocano un ruolo chiave nelle decisioni dell'allenatore sul concedere più tempo di gioco. I giocatori che dimostrano impegno, miglioramenti e adattabilità durante le sessioni di preparazione possono essere premiati con un aumento delle opportunità di gioco<sup>xxi</sup>.

#### **4.2.6 Influenza dei dati statistici**

Con l'uso sempre crescente di *analytics*<sup>14</sup> nel calcio, i dati statistici hanno un impatto significativo sulle decisioni di schieramento. I giocatori che mostrano prestazioni superiori in termini di passaggi, tiri in porta, intercettazioni e altri indicatori chiave possono beneficiare di un maggior minutaggio.

### **4.3 Confronto risultati ottenuti e affermazioni comuni**

Il confronto dettagliato delle affermazioni comuni riguardanti la partecipazione dei giovani calciatori in Serie A offre un'analisi approfondita del panorama calcistico italiano. Comunemente si ritiene che l'Italia sia un Paese che predilige giocatori più anziani, con una media d'età elevata.

I risultati emersi da questa ricerca evidenziando una tendenza decrescente nel minutaggio per i giovani giocatori caratterizzata da un picco negativo nel decennio di mezzo. Inoltre, l'analisi si confronta con le preoccupazioni espresse da figure autorevoli nel mondo del calcio italiano. Paolo Nicolato, allenatore dell'Under 21, ha manifestato la sua inquietudine riguardo al limitato tempo di gioco per i giovani calciatori, soprattutto in reparti critici come l'attacco<sup>xxii</sup>. Le sue affermazioni

---

<sup>14</sup> Analytics: analisi dei dati nel contesto calcistico. L'approccio analitico guida decisioni tattiche informate e ottimizza le strategie di gioco

aggiungono un elemento di urgenza al dibattito sulla partecipazione dei giovani talenti nella Serie A, sottolineando il possibile impatto sulle prestazioni della nazionale italiana.

Tuttavia, i risultati dell'analisi condotta indicano chiaramente una positiva inversione di tendenza verso la stagione 2011/2012, con un aumento significativo nell'opportunità e nel minutaggio offerti ai giovani calciatori nel calcio italiano nell'ultimo decennio. Tale fenomeno contraddice le percezioni comuni e riflette un cambiamento sostanziale nella dinamica della partecipazione giovanile in Serie A. Questo trend positivo è di notevole importanza per il futuro dello sviluppo calcistico in Italia, suggerendo una maggiore valorizzazione dei giovani talenti e una crescente apertura a offrire opportunità significative a una nuova generazione di giocatori.

#### **4.4 Implicazioni per lo sviluppo dei giovani talenti**

Il debutto storico di giovani talenti come Francesco Camarda e Wisdom Amey nella Serie A assume un ruolo cruciale nel contesto del minutaggio degli *under 21* nel calcio italiano.

L'eccezionalità di questi esordi rappresenta non solo un momento significativo nella carriera dei singoli giocatori, ma anche un segnale positivo per la promozione di giovani prospetti nel panorama calcistico professionistico. Francesco Camarda<sup>xxiii</sup> ha fatto il suo esordio il 25 novembre 2023 all'età di 15 anni e 260 giorni, superando il record di Wisdom Amey, che aveva debuttato per il Bologna a 15 anni e 277 giorni<sup>xxiv</sup> nel 2021. L'eccezionalità di questi esordi rappresenta non solo un momento significativo nella carriera dei singoli giocatori, ma anche un segnale positivo per la promozione di giovani prospetti nel panorama calcistico professionistico. Questi eventi offrono un'opportunità unica di acquisire esperienza in un ambiente altamente competitivo, fattore da ritenere determinante nel processo di sviluppo.

Inoltre, essi fungono da fonte d'ispirazione per altri talenti emergenti, dimostrando come sia possibile raggiungere livelli di eccellenza nella Serie A ad un'età straordinariamente giovane.

Tuttavia, è imperativo sottolineare che il debutto in Serie A costituisce solo un punto di partenza in un percorso di sviluppo più ampio. La continuità nell'impegno, il costante affinamento delle abilità e una gestione oculata da parte dei club e degli allenatori diventano elementi cruciali per garantire il mantenimento e l'ulteriore crescita di questi giovani talenti. Questo approccio attento è essenziale per evitare sovraccarichi che potrebbero mettere a rischio il benessere dei giocatori e aumentare la possibilità di infortuni.

L'articolo de *La Repubblica* riflette sull'unicità di tale debutto, descrivendo il momento di Camarda come la realizzazione di un "sogno". Questi momenti non solo segnano importanti traguardi individuali ma fungono anche da catalizzatori per una cultura che promuove attivamente l'inclusione di giovani talenti.

La maturità dimostrata da Amey, nonostante la sua giovane età, è considerata un aspetto positivo, evidenziando come l'esperienza acquisita in Serie A possa contribuire in modo significativo al suo sviluppo futuro.

In conclusione, il debutto di giovani talenti come Camarda e Amey in Serie A non solo contribuisce al loro sviluppo personale, ma ha un impatto più ampio sull'approccio del calcio italiano verso la promozione di nuovi talenti. Questi momenti rappresentano una preziosa opportunità per i giocatori di crescere e apprendere a livelli di eccellenza, diventando contemporaneamente una fonte di ispirazione per le generazioni future di talenti emergenti.

La gestione attenta di questi giovani prospetti da parte dei club è fondamentale per garantire uno sviluppo sostenibile nel competitivo scenario della Serie A.





## Elenco figure

Figura 1: caricamento librerie e inizializzazione dataset .....	19
Figura 2: db_analisi.....	21
Figura 3: nuovo_db .....	22
Figura 4: nuovo_db aggiornato .....	24
Figura 5: grafico 1 .....	25
7Figura 6: grafico 2 .....	27
Figura 7: grafico 3 .....	28
Figura 8: stagioni con i relativi gruppi e minutaggio normalizzato .....	30
Figura 9: istogramma somma aggregata minuti normalizzati (grafico 4).....	31
Figura 10: boxplot (grafico 5) .....	32
Figura 11: valori boxplot .....	33
Figura 11: media dei minuti per gruppo (grafico 6).....	34

## **Ringraziamenti**

*Alla fine di questo mio percorso ritengo fondamentale dedicare queste pagine a chi ha reso possibile la realizzazione di questa tesi e di questo traguardo.*

*Ringrazio il Professore Marco Rocchetti per accettarmi da relatore, e per la pazienza e l'interesse che mi ha dedicato nella realizzazione di questo progetto che racchiude le mie passioni.*

*Simone, grazie per essere stato il mio punto di riferimento. Fratello e amico che spesso hai portato su di te i miei pesi, grazie per avermi sopportata e supportata in qualsiasi mia decisione. Anche se non te lo dimostro mai abbastanza sei il regalo più grande che mamma e papà potessero farmi.*

*Mamma, Ti ringrazio perché in questi tre anni non hai mai smesso di credere in me, anche quando nemmeno io lo facevo. Grazie per avermi sempre ricordato le mie potenzialità, quando io nemmeno sapevo di averle. Mi hai dato amore, ma anche tenacia e resilienza quando volevo mollare e non mi sentivo all'altezza. So quanto sia stato difficile lasciarmi al mondo e tenere solo un piccolo pezzetto per te, con la consapevolezza che per molti momenti, belli e brutti, saremmo state distanti. Momenti in cui avresti voluto starmi vicino per tirarmi su, per darmi un abbraccio e dirmi che non ero sola e che mi capivi. Il tuo sostegno, con le tue chiamate giornaliere, non è mai mancato e mi ha dato la forza per continuare il mio percorso e costruire il mio futuro.*

*Mi hai sempre ripetuto “pensa alla meta, non vedo l'ora di metterti la corona”, mamma ho raggiunto il traguardo.*

*Papà, Ti ringrazio per essere presente in modo silenzioso, i tuoi gesti e i tuoi occhi parlano più delle parole.*

*Penso alle volte in cui torno a casa e mi vieni a prendere in aeroporto: i tuoi occhi e la tua voce emozionata trasmettono la gioia di avermi accanto, e ogni volta questo momento riesce a emozionarmi profondamente.*

*Giorno dopo giorno, hai sostenuto ogni mio passo, sia nei momenti di successo che in quelli di difficoltà. La tua presenza è stata una fonte costante di conforto e incoraggiamento. So quanto il lavoro ti stanchi, eppure, nonostante tutto, trovi sempre il modo di aiutarmi. Grazie.*

*Mamma e papà vi ringrazio per tutti i vostri sacrifici e fatiche, non mi avete mai fatto mancare nulla e anche se le parole non sono abbastanza, questo è il mio modo di rendervi fieri e orgogliosi di me perchè io lo sono di voi. La vostra forza e il vostro amore incondizionato sono tesori preziosi che custodisco nel cuore.*

*Gaia, grazie perché in questi 3 anni sei stata un'amica e un esempio per migliorarmi. Con la tua presenza costante anche a distanza mi hai dato un equilibrio e un punto fisso su cui poter sempre contare. Grazie per esserci.*

*Giuseppe, Ti ringrazio per aver condiviso con me questo percorso accademico. Il nostro incontro è avvenuto nel momento più impegnativo del mio cammino universitario, e la tua presenza è stata un sostegno prezioso. Mi hai stimolata a far meglio e pretendere di più da me stessa. Sono contenta di condividere proprio con te questo giorno e di essere diventata tua amica. Grazie per esserci e ascoltarmi sempre.*

*Ringrazio Luca M. perchè mi ha permesso di riuscire ad arrivare a questo traguardo.*

*Ringrazio il calcio che mi ha permesso di incontrare persone fantastiche, e grazie al quale ho tenuto un equilibrio con lo studio portando avanti due passioni.*

*Ringrazio i momenti di difficoltà e i momenti bui in cui sembra di toccare il fondo perchè mi hanno permesso di crescere e diventare la persona che sono oggi. E' grazie a questi che ho potuto apprezzare e gioire di cuore di ogni singolo momento di felicità e dare la giusta importanza a ogni tipo di problema.*

*Infine ringrazio chi in questi anni è andato via, ma che veglia su di me da lassù. Nonno e Vittoria, dedico questo traguardo anche a voi.*

## Bibliografia e Sitografia

---

<sup>i</sup> S. LEALOS, P. HEYSSEL, *Moneyball True Story: How Accurate The Baseball Movie Is*, disponibile su screenrant.com, 1 novembre 2023

<sup>ii</sup> D. JOHNSON, *The True Story That Inspired Moneyball*, disponibile su grunge.com, 18 febbraio 2022

<sup>iii</sup> *Moneyball - Das sind die wahren Hintergründe des Films mit Brad Pitt*, disponibile su moviepilot.de

<sup>iv</sup> L. DEL FRATE, *Il metodo Moneyball: il film, le scommesse e la storia di Billy Beane*, disponibile su assopoker.com, 23 agosto 2021

<sup>v</sup> <https://www.profootballnetwork.com/the-use-of-moneyball-in-the-nfl-and-beyond/>

<sup>vi</sup> <https://sportsvictor.com/moneyball-nfl/>

<sup>vii</sup> J. BUCKLEY, *How Has Moneyball Influenced the NBA?*, Sports Victor, 24 giugno 2023

<sup>viii</sup> A. F. GIUDICE, *Milan da record, effetto Moneyball*, disponibile su corrieredellosport.it, 11 settembre 2023.

<sup>x</sup> M. TRABUCCHI, *Data analytics negli sport, un mercato da 3 miliardi di dollari*, Il Sole 24 ORE. 18 luglio 2022

- 
- <sup>xi</sup> R CORE TEAM, *stats: The R Stats Package*, disponibile su <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html>
- <sup>xii</sup> H. WICKAM, *ggplot2: Elegant Graphics for Data Analysis*, disponibile su <https://ggplot2.tidyverse.org>
- <sup>xiii</sup> H. WICKAM, R. FRANÇOIS, L. HENRY, K. MÜLLER, *dplyr: A Grammar of Data Manipulation*, disponibile su <https://dplyr.tidyverse.org>
- <sup>xiv</sup> A. WALKER, *openxlsx: Read, Write and Edit XLSX Files*, disponibile su <https://github.com/awalker89/openxlsx>
- <sup>xv</sup> H. WICKAM, J. BRYAN, R. FRANÇOIS, *readxl: Read Excel Files*, disponibile su <https://readxl.tidyverse.org>
- <sup>xvi</sup> FURNISS, Matt. "Which Clubs Trusted Youth the Most in 2021-22?" *The Analyst*, JUN 9, 2022, disponibile su: <https://theanalyst.com/eu/2022/06/which-clubs-trusted-youth-the-most-in-2021-22/>
- <sup>xvii</sup> IVARSSON, Andreas et al. "Psychological factors and future performance of football players: A systematic review with meta-analysis." *Journal of Science and Medicine in Sport*, vol. 23, no. 4, 2020, pp. 415-420. DOI: 10.1016/j.jsams.2019.10.021, disponibile su: <https://pubmed.ncbi.nlm.nih.gov/31753742/>
- <sup>xviii</sup> ESKANDARIFARD, Ebrahim et al. "Exploring interactions between maturity status and playing time with fluctuations in physical fitness and hormonal markers in youth soccer players." *Scientific Reports*, vol. 12, Article number: 4463 (2022), 16 March 2022, disponibile su: <https://www.nature.com/articles/s41598-022-08567-5>

---

<sup>xix</sup> SIERRA-DÍAZ, Manuel Jacob et al. "Soccer and Relative Age Effect: A Walk among Elite Players and Young Players." *Sports*, vol. 5, no. 1, 2017, p. 5. DOI: 10.3390/sports5010005, disponibile su:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5969014/>

<sup>xx</sup> PÉREZ-GONZÁLEZ, Benito et al. "The New Generation of Professional Soccer Talent Is Born under the Bias of the RAE: Relative Age Effect in International Male Youth Soccer Championships." *Children*, vol. 8, no. 12, 2021, p. 1117. DOI: 10.3390/children8121117, disponibile su:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8700599/>

<sup>xxi</sup> Vaughan, J., Mallett, C. J., Potrac, P., Woods, C., O'Sullivan, M., & Davids, K. (2022). Social and Cultural Constraints on Football Player Development in Stockholm: Influencing Skill, Learning, and Wellbeing. *Frontiers in Sports and Active Living*, 4, 832111. DOI: 10.3389/fspor.2022.832111, disponibile su:

<https://www.frontiersin.org/articles/10.3389/fspor.2022.832111/full>

<sup>xxii</sup> Corriere dello Sport. (21 marzo 2022). Under 21, Nicolato: "I giovani giocano poco, sono preoccupato."

<sup>xxiii</sup> Carci, F. (2023, 25 novembre). "Francesco Camarda, chi è il giovane del Milan che ha debuttato in Serie A a 15 anni, 8 mesi e 15 giorni segnando un record." *La Repubblica*, disponibile su: [https://www.repubblica.it/sport/calcio/serie-a/milan/2023/11/23/news/camarda\\_milan\\_convocazione-421073461/?ref=RHLFBG-I421211842-P2-S1-T1](https://www.repubblica.it/sport/calcio/serie-a/milan/2023/11/23/news/camarda_milan_convocazione-421073461/?ref=RHLFBG-I421211842-P2-S1-T1)

<sup>xxiv</sup> Gazzetta.it. "Il Record di Esordio in Serie A: 15 anni e 277 giorni." Bologna: tutte le notizie. Serie A: tutte le notizie. Calcio: tutte le notizie. 12 maggio 2021, disponibile su: <https://www.gazzetta.it/Calcio/Serie-A/Bologna/12-05-2021/wisdom->

---

[amey-chi-piu-giovane-esordiente-sempre-seriea-15-anni-274-giorni-410966377037.shtml](http://amey-chi-piu-giovane-esordiente-sempre-seriea-15-anni-274-giorni-410966377037.shtml)