



Corso di laurea in Economia e Management

Cattedra: Statistica applicata ed econometria

Il ruolo dei fondamentali nella analisi  
della performance nella pallavolo.  
Un'applicazione dei modelli di  
regressione per dati panel.

Prof.ssa Riviuccio Giorgia

Sarcina Daniele 276391

---

RELATORE

---

CANDIDATO

A handwritten signature in black ink, appearing to read "Daniele Sarcina", is written over a horizontal line.

Anno Accademico 2023/2024

## Sommario

<b>Introduzione</b> .....	<b>4</b>
<b>1 Rilevazione statistica e ruolo dello scoutman</b> .....	<b>5</b>
1.1 Principi di rilevazione.....	7
1.2 Performance <i>analysis</i> e match <i>analysis</i> .....	8
<b>2 Variabili fondamentali nella pallavolo e rilevazione statistica</b> .....	<b>13</b>
2.1 Positività, efficacia ed efficienza.....	15
2.2 Correlazione dell'efficienza.....	17
<b>3 Esplorazione dei dati</b> .....	<b>18</b>
3.1 Analisi della multicollinearità.....	19
3.2 Analisi statistiche descrittive.....	23
<b>4 Metodologia di analisi</b> .....	<b>32</b>
4.1 Principali modelli di regressione panel.....	33
4.1.1 Modello a effetti fissi.....	33
4.1.2 Modello a effetti casuali.....	34
4.2 Test di Hausman.....	34
<b>5 Analisi e sviluppo dei modelli</b> .....	<b>35</b>
5.1 Analisi alternativa.....	39
5.2 Scelta delle variabili.....	39
<b>6 Conclusioni</b> .....	<b>44</b>
<b>Fonti</b> .....	<b>46</b>

## **Indice figure**

<i>Figura 1: Boxplot dell'efficienza in battuta di tutte le squadre di tutti i campionati .....</i>	<i>25</i>
<i>Figura 3: Boxplot dell'efficienza in alzata di tutte le squadre di tutti i campionati. ....</i>	<i>28</i>
<i>Figura 4: Boxplot dell'efficienza in attacco di tutte le squadre di tutti i campionati. ....</i>	<i>29</i>
<i>Figura 5 :Boxplot dell'efficienza a muro di tutte le squadre di tutti i campionati. ....</i>	<i>30</i>
<i>Figura 6: Boxplot dell'efficienza in difesa di tutte le squadre di tutti i campionati. ....</i>	<i>32</i>
<i>Figura 7: Evoluzione degli effetti fissi nel tempo.....</i>	<i>38</i>

## **Indice tabelle**

<i>Tabella 1: Effetti vincenti ed effetti perdenti per ogni fondamentale.....</i>	<i>16</i>
<i>Tabella 2: Correlazione della valutazione dei i fondamentali.....</i>	<i>17</i>
<i>Tabella 3: Matrice di correlazione per i fondamentali rilevati nelle squadre di serie A1.....</i>	<i>20</i>
<i>Tabella 4: Matrice di correlazione per i fondamentali rilevati nelle squadre di serie A2.....</i>	<i>20</i>
<i>Tabella 5: Matrice di correlazione per i fondamentali rilevati nelle squadre di serie A3.....</i>	<i>21</i>
<i>Tabella 6: Matrice di correlazione per i fondamentali rilevati nelle squadre di serie B.....</i>	<i>21</i>
<i>Tabella 7: Matrice di correlazione per i fondamentali rilevati nelle squadre dell'intero dataset .....</i>	<i>23</i>
<i>Tabella 8: statistiche descrittive della differenza punti per ogni squadra .....</i>	<i>24</i>
<i>Tabella 9: statistiche descrittive dell'efficienza percentuale in battuta per ogni squadra .....</i>	<i>25</i>
<i>Tabella 10: statistiche descrittive dell'efficienza percentuale in ricezione per ogni squadra .....</i>	<i>26</i>
<i>Tabella 11: statistiche descrittive dell'efficienza percentuale in alzata per ogni squadra. ....</i>	<i>27</i>
<i>Tabella 12: statistiche descrittive dell'efficienza percentuale in attacco per ogni squadra ....</i>	<i>29</i>
<i>Tabella 13: statistiche descrittive dell'efficienza percentuale a muro per ogni squadra .....</i>	<i>30</i>
<i>Tabella 14: statistiche descrittive dell'efficienza percentuale in difesa per ogni squadra .....</i>	<i>31</i>
<i>Tabella 15: Riepilogo dell'output di regressione dei modelli a effetti fissi e ad effetti variabili .....</i>	<i>37</i>
<i>Tabella 16: Riepilogo degli effetti fissi rispetto al tempo.....</i>	<i>37</i>
<i>Tabella 17: riepilogo dell'output del test di Hausmann.....</i>	<i>38</i>
<i>Tabella 18: Riepilogo dell'output di regressione per le squadre di fascia bassa.....</i>	<i>40</i>
<i>Tabella 19: Riepilogo dell'output di regressione per le squadre di fascia media.....</i>	<i>41</i>
<i>Tabella 20: Riepilogo dell'output di regressione per le squadre di fascia alta .....</i>	<i>42</i>
<i>Tabella 21: Effetti fissi per campionato e per fascia di classifica .....</i>	<i>43</i>
<i>Tabella 22: Test di Hausmann per le squadre di alta classifica.....</i>	<i>43</i>
<i>Tabella 23: Test di Hausmann per le squadre di alta classifica.....</i>	<i>43</i>
<i>Tabella 24: Test di Hausmann per le squadre di alta classifica.....</i>	<i>44</i>

## Introduzione

Nella società odierna, l'avvento delle nuove tecnologie ha trasformato radicalmente l'approccio alla competizione nello sport di alto livello. Uno dei motivi di questa trasformazione è rappresentato dall'evoluzione dei metodi di raccolta e analisi dei dati di gioco, insieme all'introduzione di figure professionali specializzate nell'utilizzo di questi dati. Questi analisti di dati sportivi sono diventati una componente essenziale per le squadre che aspirano al successo nelle competizioni di alto livello, in qualunque sport.

La raccolta dei dati in tempo reale permette di acquisire informazioni dettagliate su una vasta gamma di variabili, dalle prestazioni fisiche, come velocità e resistenza, ai parametri tecnici, come la precisione dei passaggi o la potenza del tiro. Questi dati vengono poi utilizzati per individuare tendenze, punti di forza, di debolezza, sia a livello individuale che di squadra. L'obiettivo ultimo di questa analisi è rendere gli atleti consapevoli dei propri limiti e potenzialità, indirizzando anche le scelte strategiche durante gli allenamenti e le competizioni.

Esempi di questo tipo oggi si possono osservare in tutti gli sport: *in primis* nel calcio, dove le squadre di vertice impiegano analisti che studiano ogni dettaglio delle partite, dalla posizione in campo dei giocatori ai movimenti senza palla, fino alle tattiche degli avversari. Grazie a queste informazioni, gli allenatori possono compiere scelte più consapevoli, che possono spaziare dalla modifica della formazione di base, allo sfruttamento di alcune specifiche debolezze degli avversari in determinate situazioni di gioco.

In tutti gli sport è diventato indispensabile adottare delle decisioni *data-driven*: i dati raccolti durante allenamenti e partite possono ridurre lo stress fisico, diminuendo gli infortuni, aumentando parallelamente le prestazioni degli atleti con programmi di allenamento personalizzati.

L'integrazione della scienza dei dati all'interno del processo decisionale in ambito sportivo non è, peraltro, priva di sfide. È fondamentale, infatti, che le squadre sappiano interpretare correttamente i dati e che gli allenatori siano disposti a fidarsi delle analisi fornite dagli esperti. Per competere ad alti livelli è dunque ormai indispensabile adottare questo tipo di approccio guidato dai dati: tuttavia ciò non deve eclissare l'importanza dell'aspetto umano nella

competizione. Un'eccessiva dipendenza dai dati potrebbe infatti svilire la componente agonistica, intuitiva ed emotiva che caratterizza i grandi campioni. Queste variabili rappresentano aspetti qualitativi che i dati non riescono a misurare e che nessuna analisi è in grado di prevedere.

## **1 Rilevazione statistica e ruolo dello scoutman**

All'interno del mondo dell'analisi dei dati sportivi, la figura preposta alla rilevazione, all'analisi e alla comunicazione dei dati è chiamata *scoutman*. Il significato letterale del termine *scoutman* è quello di esploratore: colui che perlustra ed effettua una ricognizione su qualcosa, raccogliendo informazioni preziose che possono essere utilizzate per orientare le decisioni strategiche. In particolare, lo *scoutman* rappresenta una figura tecnica di grande importanza facente parte dello staff di una squadra sportiva. Il suo ruolo è fondamentale perché, attraverso il suo lavoro, mette in condizione l'intera squadra, allenatori compresi, di compiere le scelte giuste e di individuare con precisione punti di forza e di debolezza. Questo permette di indirizzare correttamente il lavoro in palestra, ottimizzare le strategie di gioco e migliorare le performance complessive della squadra.

Lo *scoutman* opera principalmente dietro le quinte, lontano dai riflettori che solitamente illuminano il lavoro in palestra o durante le partite. È una figura che lavora in modo discreto ma estremamente efficace, analizzando una quantità significativa di dati e informazioni che vengono raccolti durante gli allenamenti e le competizioni. Questi dati possono includere statistiche dettagliate sulle prestazioni dei singoli giocatori, sull'efficacia delle diverse tattiche utilizzate, e su come la squadra risponde alle strategie degli avversari. Grazie alla sua analisi, lo *scoutman* è in grado di fornire indicazioni precise e mirate che aiutano gli allenatori a prendere decisioni informate e strategiche.

Esistono due tipologie principali di *scoutman*, ciascuna con compiti specifici, che contribuiscono in maniera diversa ma complementare al successo della squadra.

La prima tipologia è lo *scoutman* rilevatore. Questo professionista si occupa di eseguire correttamente la rilevazione statistica. Il suo compito è quello di osservare e

registrare accuratamente tutti gli eventi rilevanti che si verificano durante una partita o un allenamento. La rilevazione statistica include la registrazione dei punti, degli errori, delle azioni di gioco e di altre metriche fondamentali che forniscono una base dati dettagliata e oggettiva. La precisione e l'accuratezza nella raccolta dei dati sono cruciali per garantire che le analisi successive siano affidabili e utili.

La seconda tipologia è lo *scoutman* analista. Questo professionista si occupa dell'interpretazione dei dati raccolti. Il suo compito è tradurre queste informazioni grezze in concrete indicazioni tecnico-tattiche che possono essere utilizzate dagli allenatori e dai giocatori per migliorare le performance. Lo *scoutman* analista esamina i dati con l'obiettivo di identificare schemi, tendenze e aree di miglioramento. Fornisce *feedback* dettagliati sugli aspetti tecnici e tattici del gioco, suggerendo modifiche e adattamenti che possono essere implementati durante gli allenamenti o le partite. Questo processo di analisi richiede una profonda conoscenza del gioco, delle strategie e delle dinamiche di squadra, nonché competenze avanzate nell'uso di *software* e strumenti di analisi dei dati.

L'adozione di una di queste figure all'interno di una squadra spesso dipende dall'allenatore, il quale ha la responsabilità di decidere quale approccio sia più utile per le esigenze specifiche del *team*. Gli allenatori, infatti, potrebbero avere preferenze diverse in base alla propria filosofia di gioco, alla loro esperienza e al contesto in cui operano. Alcuni allenatori potrebbero preferire avere dei suggerimenti più distaccati e strettamente legati ai dati raccolti, affidandosi prevalentemente allo *scoutman* rilevatore. Questo approccio consente agli allenatori di ottenere una visione oggettiva e non influenzata dei vari aspetti del gioco, basata esclusivamente su metriche quantitative. La rilevazione statistica fornita dallo *scoutman* rilevatore diventa quindi uno strumento fondamentale per prendere decisioni strategiche, poiché offre una rappresentazione chiara e precisa delle performance della squadra e degli avversari.

D'altro canto, altri allenatori potrebbero optare per avvalersi principalmente delle competenze dello *scoutman* analista, che non solo raccoglie i dati, ma li interpreta e li traduce in indicazioni tecnico-tattiche. Questo approccio permette agli allenatori di ricevere una sintesi elaborata delle informazioni, corredata da suggerimenti e consigli specifici su come migliorare le strategie di gioco. L'analisi dei dati effettuata dallo *scoutman* analista

può fornire *insight* preziosi sulle tendenze emergenti, sugli schemi di gioco degli avversari e sulle aree di miglioramento della propria squadra. Gli allenatori possono quindi utilizzare queste informazioni per sviluppare piani di allenamento mirati e per prendere decisioni più informate durante le partite.

È importante sottolineare che, indipendentemente dall'approccio scelto, la comunicazione tra l'allenatore e gli scoutman è cruciale. Una comunicazione efficace assicura che le informazioni raccolte e le analisi effettuate siano comprese correttamente e utilizzate in modo appropriato. Gli allenatori devono essere in grado di interpretare i dati e le analisi fornite dagli scoutman per tradurle in azioni concrete sul campo. Allo stesso tempo, gli scoutman devono comprendere le esigenze e le aspettative degli allenatori per poter fornire informazioni rilevanti e utili.

## **1.1 Principi di rilevazione**

Lo *scoutman* è dunque la figura professionale che si occupa di raccogliere, analizzare e interpretare i dati delle prestazioni, utilizzando vari strumenti e metodologie. La valutazione delle prestazioni attraverso lo *scoutman* deve rispettare una serie di requisiti fondamentali per garantire l'accuratezza e l'utilità dei dati raccolti. Questi requisiti sono essenziali per ottenere una rappresentazione fedele delle prestazioni reali degli atleti durante le competizioni o gli allenamenti. Di seguito vengono illustrati questi requisiti in dettaglio:

- **Affidabilità:** È imprescindibile che la valutazione delle prestazioni sia effettuata in modo completo, catturando accuratamente tutti gli aspetti osservabili delle performance degli atleti. Ciò implica che la raccolta dei dati deve essere coerente con la realtà empirica delle situazioni di gioco, evitando di essere influenzata dalle intenzioni tecniche e tattiche dei giocatori o da interpretazioni soggettive. L'affidabilità garantisce che i dati raccolti siano forniscano una base solida per l'analisi delle prestazioni.
- **Condivisibilità dei parametri:** È fondamentale che i parametri utilizzati per effettuare la valutazione delle prestazioni siano codificati in modo chiaro e

comprensibile per tutti i membri del *team* di lavoro. Questo significa che deve esistere un linguaggio comune e un set di criteri condivisi che permettano a tutti di interpretare i dati nello stesso modo. La condivisione dei parametri assicura che tutte le persone coinvolte nel processo di analisi lavorino per lo stesso obiettivo, facilitando la comunicazione e la collaborazione all'interno del gruppo.

- Utilizzo di parametri non derivanti da variabili esterne: È importante che i parametri utilizzati per la valutazione non siano influenzati da variabili esterne che possono distorcere i risultati. Non devono essere rilevate variabili qualitative legate, ad esempio, alla spettacolarità del gesto tecnico o alle caratteristiche fisiche individuali degli atleti, come l'altezza o la forza muscolare. L'attenzione deve essere rivolta esclusivamente a parametri oggettivi e misurabili che riflettano la reale performance in campo. Questo approccio assicura che la valutazione sia equa e basata su dati tangibili, evitando *bias* che potrebbero compromettere l'oggettività dell'analisi.

## **1.2 Performance analysis e match analysis**

Come detto l'analisi e la rilevazione statistica dei dati stanno diventando sempre più cruciali nello studio e nelle decisioni tecnico-tattiche relative al lavoro di squadra. La raccolta dati nel suo complesso è definibile come *performance analysis*, che fornisce informazioni dettagliate e preziose per migliorare le performance degli atleti e delle squadre.

In questa pratica è possibile distinguere in due diversi segmenti, ciascuno caratterizzato da un *focus* specifico e una metodologia particolare.

- *Notational analysis*: questo segmento comprende un'oggettiva e accurata registrazione degli eventi che si susseguono in gara. Questa registrazione dettagliata darà vita ai dati grezzi che saranno successivamente analizzati. La *notational analysis* si concentra su una rappresentazione sistematica e strutturata delle azioni e delle sequenze di gioco, utilizzando strumenti e software dedicati per garantire precisione e affidabilità. Vengono annotate varie situazioni di gioco e momenti chiave che caratterizzano la

performance della squadra e degli avversari. L'obiettivo è quello di creare un database esaustivo che possa essere utilizzato per identificare schemi, tendenze e aree di miglioramento. Ad esempio, si possono analizzare le percentuali di successo dei vari tipi di attacco o le zone del campo più frequentemente utilizzate per determinati tipi di gioco.

- *Motion analysis*: In questo segmento, rientrano tutti quegli aspetti legati alla meccanica e al movimento degli atleti. Questa analisi è volta a stimare il tasso di fatica degli atleti durante l'allenamento o la gara, tramite vari indicatori come il numero di salti effettuati, la velocità di corsa, il calcolo della distanza percorsa e altre metriche biomeccaniche. La *motion analysis* utilizza tecnologie avanzate come sensori indossabili, videocamere ad alta velocità e software di analisi del movimento per monitorare e valutare i movimenti degli atleti in tempo reale. Questi dati permettono di comprendere meglio come gli atleti utilizzano il loro corpo durante la competizione, identificando potenziali problemi di tecnica o efficienza. Ad esempio, l'analisi può rivelare se un giocatore sta eseguendo movimenti che aumentano il rischio di infortunio o se c'è un calo di prestazione legato alla stanchezza muscolare. Inoltre, la *motion analysis* può essere utilizzata per ottimizzare le strategie di allenamento, personalizzando i programmi di esercizio per migliorare specifiche abilità fisiche e ridurre il rischio di infortuni.

Un'altra distinzione all'interno della *Performance analysis* può essere data da:

- *Analisi quantitativa*: Questo tipo di studio si concentra sull'analisi di una grande quantità di dati raccolti durante le performance. L'obiettivo principale è quello di esaminare e interpretare i dati numerici e statistici. È importante sottolineare che questa forma di analisi non prende in considerazione l'aspetto qualitativo dei dati rilevati. Ad esempio, non valuta la straordinarietà di un punto segnato o la correttezza e la precisione di un gesto tecnico eseguito dall'atleta. L'analisi quantitativa è quindi orientata verso una comprensione generale e oggettiva delle performance, basandosi su numeri e statistiche che possono essere facilmente misurati e confrontati.

- **Analisi qualitativa:** Questo tipo di studio, al contrario, si concentra sulla qualità degli eventi e delle azioni registrate durante le performance. L'analisi qualitativa è profondamente interessata agli aspetti tecnico-tattici delle singole fasi di gioco, cercando di comprendere e valutare la qualità e l'efficacia delle azioni compiute dagli atleti. In questo contesto, viene dato rilievo alla precisione di un gesto tecnico, alla strategia e alla tattica utilizzate in determinate situazioni di gioco. L'analisi qualitativa può includere l'osservazione dettagliata di movimenti specifici, la valutazione della coordinazione tra i giocatori e la comprensione delle scelte tattiche fatte durante una partita. Questo tipo di analisi richiede una visione più soggettiva, basata su competenze tecniche e tattiche specifiche, e mira a migliorare la qualità complessiva della performance attraverso un'attenzione particolare ai dettagli e alle sfumature del gioco.

Differente dalla *performance analysis* è la *match analysis*: nel primo caso intendiamo un'analisi d'insieme della squadra e del suo ciclo di lavoro anche durante gli allenamenti, mentre nel secondo caso consideriamo solo ed esclusivamente la prestazione dei singoli atleti durante la gara (come suggerisce il termine *match analysis*). Quest'ultima rappresenta quindi uno strumento attraverso il quale poter migliorare le capacità individuali.

Nel contesto della *match analysis*, particolare rilievo assume il concetto dell'analisi SWOT, che rappresenta uno strumento fondamentale per la preparazione di una gara. Questo tipo di analisi, che ha origine nel mondo aziendale dove è frequentemente utilizzata per pianificare in modo strategico le attività dell'organizzazione produttiva, è stata adattata con successo anche nel contesto sportivo.

L'analisi SWOT mira a fare emergere quattro elementi chiave:

1. *Strengths* (punti di forza)
2. *Weaknesses* (punti di debolezza)
3. *Opportunities* (opportunità)

#### 4. *Threats* (minacce)

Questi elementi vengono esaminati in dettaglio per fornire una visione completa e strutturata della situazione attuale e delle prospettive future. Nell'ambito della *match analysis*, il metodo SWOT svolge un ruolo cruciale, poiché analizza e mette in comparazione i fattori favorevoli e sfavorevoli di entrambe le squadre. Questo processo è estremamente utile per il raggiungimento dell'obiettivo principale, che è quello di migliorare le performance e ottenere risultati positivi.

L'analisi dei punti di forza (punto 1) consente di identificare le caratteristiche e le competenze che danno un vantaggio competitivo alla squadra, come ad esempio l'abilità tattica, una solidità tecnica o una buona costanza.

L'analisi dei punti di debolezza (punto 2) permette, invece, di individuare gli aspetti che necessitano di miglioramento, come ad esempio carenze tecniche o lacune nella preparazione fisica.

Le opportunità (punto 3) rappresentano gli elementi esterni che la squadra può sfruttare a proprio vantaggio, come ad esempio, errori tattici degli avversari o una solida preparazione mentale.

Le minacce (punto 4), infine, sono i fattori esterni che potrebbero ostacolare il raggiungimento degli obiettivi, come ad esempio gesti tecnici in cui l'avversario è particolarmente forte, individualità avversarie spiccate, o strategie estremamente efficaci degli avversari.

L'analisi SWOT, inoltre, consente al proprio team di valutare i livelli di performance individuale e collettiva, mettendoli in confronto diretto con quelli degli avversari, consentendo in tal modo di identificare non solo le aree di eccellenza e quelle che necessitano di sviluppo, ma anche di pianificare le strategie di gioco in modo più efficace. Ad esempio, se l'analisi evidenzia che una squadra ha un punto di forza nell'attacco mentre l'altra è debole in difesa, la strategia potrà essere adattata per sfruttare questa differenza a proprio favore.

La *match analysis* si articola in cinque fasi:

1. Osservazione e analisi delle gare precedenti;
2. Elaborazione delle informazioni ottenute dall'analisi;
3. Verifica in gara di quanto osservato nelle fasi precedenti;
4. Applicazione in gara della propria tattica di gioco sviluppata;
5. Analisi post-gara

Nelle prime due fasi definite "fasi pre-gara" viene effettuata una comparazione tra le caratteristiche proprie e quelle dell'avversario, al fine di sviluppare la migliore tattica possibile.

Nelle successive due fasi, definite "fasi di gara", l'attenzione viene posta su entrambe le squadre, monitorando la prestazione dei singoli atleti e quella complessiva, con un *focus* particolare sullo scostamento tra la prestazione effettiva e quanto precedentemente studiato.

Nell'ultima fase, definita "fase post-gara", l'analisi si concentrerà più sulla propria squadra, cercando di effettuare uno studio degli elementi fondamentali che hanno influenzato la vittoria o la sconfitta, anche a fine di indirizzare correttamente il lavoro per gli allenamenti successivi.

Ciascuna di queste fasi è supportata da uno studio:

- a. della tecnica, che, tramite l'utilizzo di analisi video, permette di comprendere meglio i sistemi di gioco avversari ma soprattutto di migliorare il bagaglio tecnico individuale.
- b. della tattica, che si concretizza in analisi quantitative, volte a individuare tendenze e debolezze di ciascun gesto tecnico analizzato.

I dati utilizzati durante tutte le fasi della *match analysis* devono essere correttamente

interpretati, cercando di filtrare le informazioni utili da quelle inutili. Durante le fasi di *match analysis*, infatti, numerosi possono essere gli errori commessi. Un errore molto comune riguarda un'eccessiva attenzione al gioco avversario: un metodo di gioco che tenga in considerazione solo la tattica avversaria, potrebbe portare al mancato sfruttamento dei punti di forza della propria squadra.

Un altro errore comune è quello di credere che i numeri rappresentino verità assolute: essi forniscono un'accurata idea dell'andamento di gioco, ma devono essere letti e utilizzati mettendoli in relazione a specifiche situazioni di gioco non rilevabili dalla tecnologia, come la concentrazione, la motivazione o la psiche di un atleta. Proprio a causa di questo errato modo di pensare, molti allenatori si avvicinano alla gara riponendo estrema fiducia nei dati e nelle analisi effettuate, non tenendo invece conto dell'analogo processo di analisi effettuato dalla squadra avversaria e le conseguenti strategie che ne derivano.

## **2 Variabili fondamentali nella pallavolo e rilevazione statistica**

La rilevazione statistica è definita come il complesso di azioni che permette la raccolta di dati utili alla valutazione e all'analisi di situazioni di gioco, e più in generale, della prestazione nel suo complesso.

Nel contesto della pallavolo, essa assume un ruolo particolarmente importante essendo uno sport caratterizzato da un'alta intensità e da azioni rapide e dinamiche, in cui sono minori le possibilità per i giocatori di improvvisate soluzioni di gioco adeguate al contesto di gara.

La rilevazione statistica può avvenire manualmente, per cui sarà necessario solamente utilizzare carta e penna, o, alternativamente, può avvenire tramite supporti informatici, che permettono una più rapida rilevazione ed elaborazione dei dati.

La raccolta dei dati attraverso strumenti informatici presuppone la codifica alfanumerica di ogni situazione di gioco, per ogni gesto tecnico fondamentale. Ciascun codice alfanumerico sarà composto in modo tale da incorporare al suo interno ogni possibile

situazione, e, quanto più è complesso, maggiori saranno il numero di variabili registrate. I fondamentali che vengono codificati, e che vengono presi in considerazione nella fase di analisi sono:

1. battuta: azione di mettere la palla in gioco colpendola dal fondo del campo;
2. ricezione: azione di ricevere la battuta controllando la palla in modo che possa essere giocata dai propri compagni. Ci sono vari tipi di battuta a seconda della tecnica utilizzata nella sua esecuzione, ad esempio battuta "salto spin" o "salto float";
3. alzata: tecnica utilizzata al fine di passare la palla in modo preciso, per possibile l'attacco di uno dei compagni;
4. attacco: azione di colpire la palla con forza e precisione verso il campo avversario allo scopo di far toccare la palla per terra sul campo avversario. Include schiacciate, pallonetti e altri tipi di colpi offensivi;
5. muro: azione difensiva di saltare e posizionare le mani sopra la rete per bloccare o deviare l'attacco avversario. Viene eseguito da uno o più giocatori;
6. difesa: insieme di tecniche utilizzate per evitare che la palla tocchi il proprio campo dopo un attacco avversario. Include infilate, tuffi e recuperi vari.

Ognuno di questi fondamentali può essere valutato in base all'esito dello stesso, e la valutazione può essere:

- # (doppio più): il fondamentale ha sortito l'esito migliore possibile, come ad esempio un punto diretto nel caso dell'attacco;
- + (più): il fondamentale ha sortito un esito positivo, ma non il migliore. Potrebbe quindi aver messo in seria difficoltà la squadra avversaria;
- ! (esclamativo): il fondamentale ha sortito un effetto non classificabile né come negativo né come positivo, ma la natura di tale valutazione può tendere verso uno degli estremi a seconda del fondamentale in esame;

- = (doppio meno): il fondamentale ha sortito l'effetto peggiore possibile, come ad esempio un errore diretto;
- - (meno): il fondamentale ha sortito un effetto negativo per la squadra, ma non il peggiore. Ad esempio, un attacco con esito "meno" ha permesso alla squadra avversaria di difendere perfettamente la palla;
- \ (slash): il fondamentale sortisce un effetto che pone in una situazione svantaggiosa (eccezion fatta per la battuta).

Si specifica che ogni valutazione assume un significato leggermente differente a seconda del fondamentale a cui viene associata. La corretta identificazione dell'esito del fondamentale è utile per non avere situazioni di ambiguità nell'analisi.

Due esempi di rilevazione con i codici alfanumerici è il seguente<sup>1</sup>:

- \*15A#: attacco punto del giocatore numero 15, della propria squadra, dove \* sta per squadra interna e 15 è il numero del giocatore. A è il fondamentale dell'attacco, # è l'esito.
- a8SQ19=: battuta spin errore del giocatore numero 8 della squadra avversaria, dove "a" sta per avversario, 8 è il numero del giocatore, S è il codice per la battuta Q identifica la tipologia di battuta spin, 1 è la zona di partenza del colpo, 9 è la zona di arrivo del colpo, = è la valutazione.

## 2.1 Positività, efficacia ed efficienza

I concetti di positività, efficacia ed efficienza spesso vengono confusi tra di loro, nonostante rappresentino aspetti ben distinti dell'analisi statistica dei fondamentali. Fatta eccezione per punto ed errore, esistono molti modi per intendere l'esito di un fondamentale. Di seguito viene riportata una tabella di effetti vincenti ed effetti perdenti, condivisi nella comunità pallavolistica, di ogni fondamentale:

---

<sup>1</sup> Annalisa Pinto, 2020.

Fondamentale	Effetti vincenti	Effetti perdenti
Battuta	#+!\	=
Attacco	#+	= \
Muro	#	= \
Difesa	#	= \
Alzata	#	= \

*Tabella 1: Effetti vincenti ed effetti perdenti per ogni fondamentale*

Data questa suddivisione di effetti vincenti ed effetti perdenti, possiamo definire la positività come la somma di tutti i valori positivi associati a uno specifico fondamentale di gioco. La positività, dunque, rappresenta una misura complessiva di tutte le azioni che hanno contribuito in modo favorevole al successo della squadra. Questo concetto è particolarmente rilevante e di grande importanza per la ricezione, poiché questo fondamentale è il punto di partenza di ogni azione di gioco. La ricezione, infatti, tende a generare una serie di effetti positivi a catena che influenzano l'intera sequenza di gioco. Una buona ricezione permette di impostare l'azione offensiva in modo ottimale, creando opportunità per attacchi efficaci e punti vincenti. Pertanto, la positività della ricezione non solo incide direttamente sulla qualità della fase di gioco successiva, ma ha anche un impatto significativo sull'andamento complessivo della partita.

Definiamo poi l'efficacia come la capacità di raggiungere un obiettivo prefissato. Nel contesto della pallavolo, questo può significare il raggiungimento di un certo numero di muri vincenti o di difese positive durante un match. L'efficacia, quindi, si misura in termini di successo nel conseguire i risultati desiderati. Ad esempio, se una squadra si pone l'obiettivo di effettuare un certo numero di muri vincenti per limitare l'attacco avversario, l'efficacia sarà determinata dal numero di muri vincenti effettivamente realizzati rispetto all'obiettivo prefissato. Analogamente, l'efficacia nella difesa si valuta in base alla capacità della squadra di eseguire difese che trasformano gli attacchi avversari in opportunità di contrattacco.

Infine, l'efficienza è definita come la capacità di raggiungere un determinato obiettivo commettendo il minor numero possibile di errori. Questo concetto implica non solo il raggiungimento del risultato desiderato, ma anche la minimizzazione delle risorse spese, in termini di errori commessi. Ad esempio, una squadra che riesce a mantenere un alto livello

di performance difensiva con pochissimi errori di posizionamento o di esecuzione sarà considerata altamente efficiente. L'efficienza è quindi una misura della qualità dell'esecuzione tecnica e tattica, riflettendo la capacità della squadra di operare in modo preciso e senza sprechi di opportunità. In altre parole, una squadra efficiente è in grado di massimizzare il rendimento riducendo al minimo le azioni fallimentari che possono concedere punti agli avversari. L'efficienza può essere calcolata tramite una formula:

$$Efficienza = \frac{(effetti\ vincenti) - (effetti\ perdenti)}{Totale\ colpi}$$

## 2.2 Correlazione dell'efficienza

Come è logico e naturale pensare, quando si verifica un effetto positivo per una squadra, tale situazione comporta inevitabilmente, in modo diretto e immediato, un effetto negativo corrispondente per la squadra avversaria. Questo perché nel contesto di una partita, ogni azione che avvantaggia una delle due formazioni, sia essa un punto segnato, una buona difesa o una tattica efficace, si traduce parallelamente in una situazione sfavorevole o svantaggiosa per l'altra squadra. Pertanto, l'equilibrio competitivo tra le squadre viene influenzato in modo reciproco e opposto, creando un'interdipendenza degli esiti tra i due contendenti. Di seguito viene riportata una tabella che spiega come avviene la correlazione tra i fondamentali.

Attacco	Muro		Attacco	Difesa		Battuta	Ricezione
=			=			=	
\	#		\			\	\
-	+		-	#		-	#+
!	!		!			!	!
+	-		+	-		+	-
#	=		#	=		#	=

Tabella 2: Correlazione della valutazione dei fondamentali.

### 3 Esplorazione dei dati

Il dataset in esame raccoglie le performance di 12 squadre appartenenti a 4 diversi campionati di pallavolo: Serie A1, Serie A2, Serie A3 e Serie B. Per ciascuno di questi campionati, sono state selezionate tre squadre rappresentative, ciascuna appartenente a diverse fasce di posizionamento in classifica al termine della stagione. Più specificamente, sono state scelte squadre di alta classifica, di media classifica e di bassa classifica per garantire una rappresentazione equilibrata delle diverse performance e dinamiche di gioco all'interno di ciascun campionato. Il dataset a nostra disposizione è composto dai risultati dettagliati di diverse squadre di pallavolo che partecipano ai campionati italiani di serie A1, serie A2, serie A3 e serie B. Per ogni squadra selezionata, il dataset include tutte le partite disputate durante l'intero campionato. Questo significa che abbiamo a disposizione una serie completa di dati che copre ogni singolo incontro giocato dalle squadre scelte. Le squadre di Serie A hanno giocato 22 partite ciascuna, mentre le squadre degli altri campionati (Serie A2, Serie A3 e Serie B) hanno giocato 26 partite ciascuna. Questa differenza è dovuta al diverso numero di squadre che compongono ciascun campionato. Complessivamente, il numero di entità osservate nel dataset analizzato è  $n=12$  (4 campionati con 3 squadre ciascuno), per un numero di periodi che varia, in base al campionato di riferimento, da  $T=22$  a  $T=26$ . Il numero totale di osservazioni nel dataset è pari a 300. Per ciascuna partita, il dataset riporta una serie di parametri specifici che sono cruciali per analizzare le performance di squadra. Questi parametri includono:

- differenza punti (negativa o positiva)
- variabile dicotomica indicante vittoria o sconfitta
- efficienza percentuale della battuta
- efficienza percentuale della ricezione della battuta
- efficienza percentuale dell'attacco
- efficienza percentuale del muro
- efficienza percentuale della difesa
- efficienza percentuale dell'alzata

Ogni valore che indica l'efficienza di un fondamentale è espresso in percentuale, offrendo così un'indicazione chiara e quantitativa della performance della squadra in ciascuno di questi aspetti del gioco.

### **3.1 Analisi della multicollinearità**

Prima di intraprendere qualsiasi tipo di analisi statistica sui dati a nostra disposizione, è essenziale esaminare attentamente se le variabili all'interno del dataset presentano correlazioni significative tra di loro. Questo passaggio preliminare è fondamentale perché la presenza di multicollinearità potrebbe compromettere i risultati delle nostre analisi. La multicollinearità si verifica quando due o più variabili indipendenti in un modello di regressione sono altamente correlate, cioè, forniscono informazioni sovrapposte. Questo fenomeno può rendere difficile stimare con precisione i coefficienti delle variabili nel modello, poiché l'effetto di una variabile può essere confuso con quello di un'altra.

La multicollinearità riduce la potenza del modello di individuare le relazioni significative tra le variabili indipendenti e la variabile dipendente. In altre parole, anche se una variabile è effettivamente importante per spiegare la variabile dipendente, la multicollinearità può mascherarne l'importanza.

Per evitare la multicollinearità, sono state calcolate le matrici di correlazione per campionato, tenendo quindi in considerazione tutte le squadre che partecipano ad uno stesso campionato di una categoria. Questo approccio permette di osservare le relazioni tra variabili all'interno di ciascun campionato e di identificare eventuali correlazioni significative che potrebbero influenzare ulteriori analisi. Per effettuare il calcolo di queste matrici di correlazione si è tenuto conto dell'ordine temporale delle partite disputate dalle varie squadre.

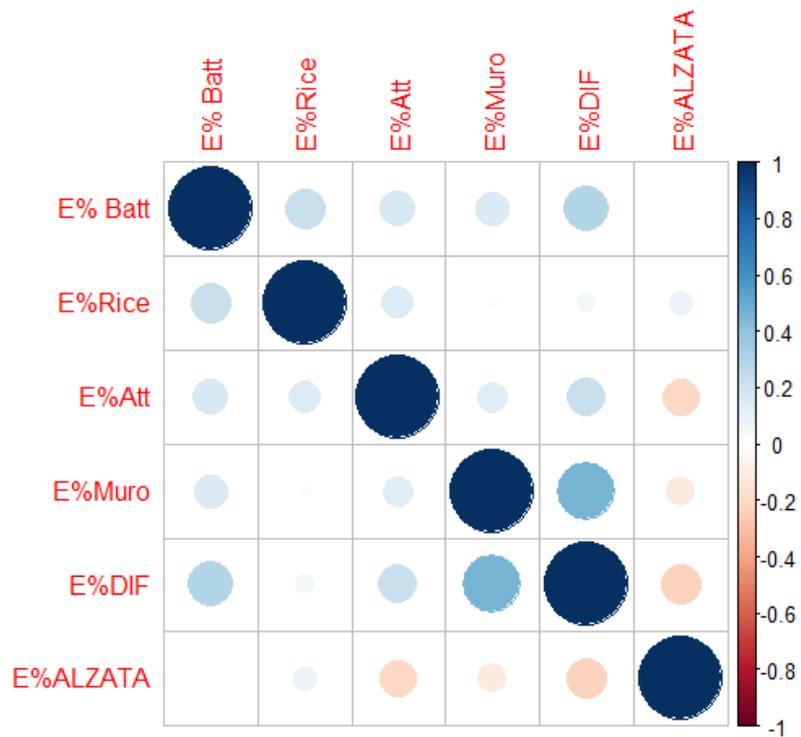


Tabella 3: Matrice di correlazione per i fondamentali rilevati nelle squadre di serie A1

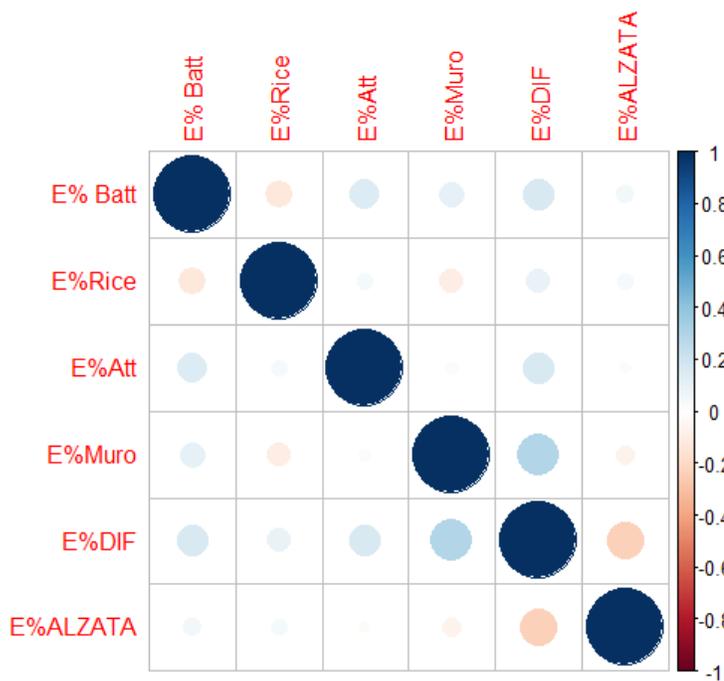


Tabella 4: Matrice di correlazione per i fondamentali rilevati nelle squadre di serie A2

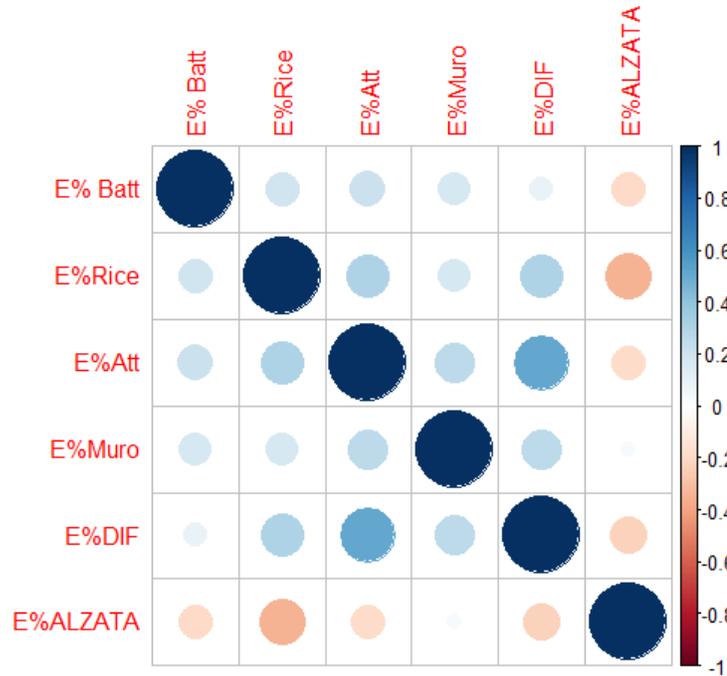


Tabella 5: Matrice di correlazione per i fondamentali rilevati nelle squadre di serie A3

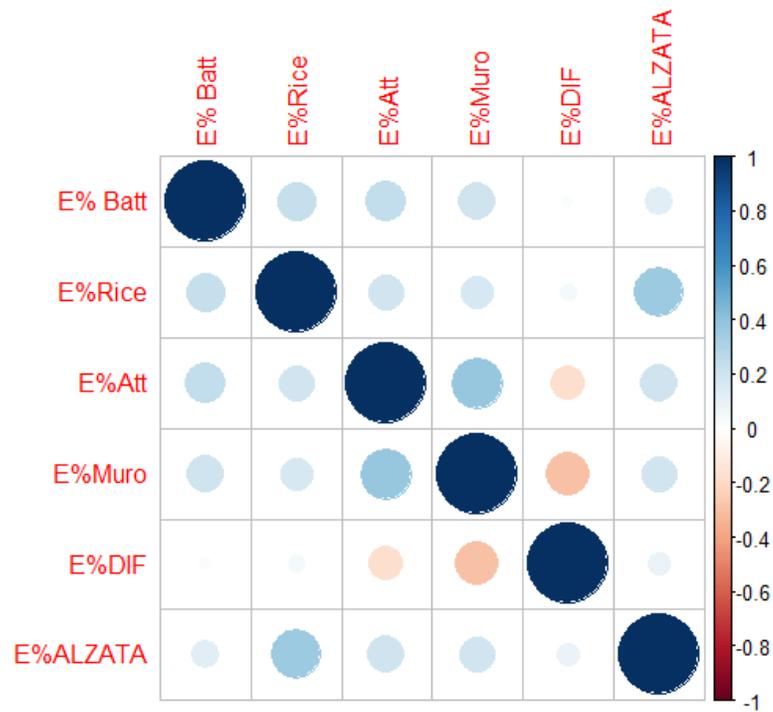


Tabella 6: Matrice di correlazione per i fondamentali rilevati nelle squadre di serie B

Per quanto riguarda la serie A1 osserviamo che le variabili più strettamente correlate

sono l'efficienza d'alzata e l'efficienza della ricezione, con un coefficiente di correlazione di 0.4688, e l'efficienza di muro con l'efficienza d'attacco con un coefficiente di 0.297.

La matrice di correlazione di serie A2 non presenta coefficienti di correlazione particolarmente rilevanti.

La serie A3 ha una matrice di correlazione che evidenzia un coefficiente positivo di correlazione tra l'efficienza in difesa e l'efficienza in attacco di 0.516, e un coefficiente di correlazione negativo tra l'efficienza d'alzata e l'efficienza in ricezione di -0.347.

In ultimo, la matrice di correlazione della serie B presenta un coefficiente positivo di correlazione di 0.366 tra l'efficienza in alzata e quella in ricezione, e una correlazione sempre positiva tra l'efficienza in attacco e quella di muro di 0.389.

Si specifica che un coefficiente di correlazione positivo o negativo elevato non rappresenta necessariamente una correlazione di tipo causale tra le variabili. Tuttavia, un'alta correlazione può rappresentare un problema nel momento in cui si devono condurre analisi più approfondite, come la costruzione di un modello di regressione, poiché può introdurre multicollinearità, che distorce le stime dei coefficienti del modello.

Per queste analisi è stato utilizzato il coefficiente di correlazione di Pearson, un indice che misura la forza e la direzione di una relazione lineare tra due variabili. Questo coefficiente varia tra -1 e 1. Un valore di 1 indica una perfetta correlazione positiva, un valore di -1 indica una perfetta correlazione negativa, mentre un valore di 0 indica l'assenza di una correlazione lineare. Il coefficiente di Pearson è calcolato come la covarianza delle due variabili divisa per il prodotto delle loro deviazioni standard<sup>2</sup>:

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

dove  $x_i$  e  $y_i$  sono i valori delle due variabili, e  $\bar{x}$  e  $\bar{y}$  sono le loro medie.

Per completezza viene riportata anche la matrice di correlazione completa, che tiene in considerazione tutti e quattro i campionati. Calcolando una matrice di correlazione congiunta, non si rilevano coefficienti di correlazione significativi. Questo può indicare che, a livello aggregato, le variabili non mostrano relazioni forti tra loro, probabilmente a causa

---

<sup>2</sup> Stock, James H., and Mark W. Watson, 2005.

delle differenze tra campionati e squadre.

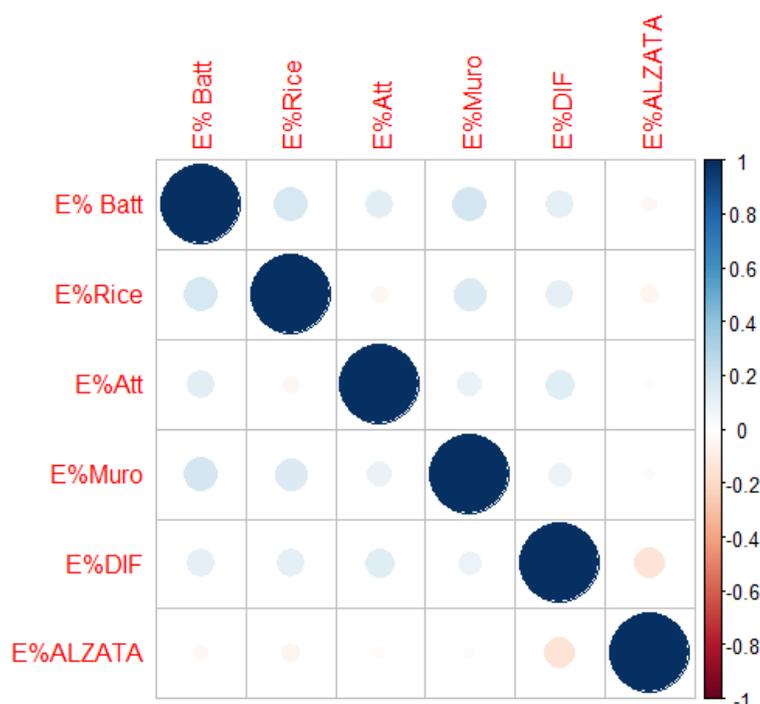


Tabella 7: Matrice di correlazione per i fondamentali rilevati nelle squadre dell'intero dataset

Nei grafici soprastanti, la grandezza di ogni cerchio sta ad indicare l'intensità della correlazione tra due variabili. Il colore del cerchio invece indica se la correlazione è negativa (rosso) o positiva (blu).

### 3.2 Analisi statistiche descrittive

Vengono riportate di seguito le principali statistiche descrittive con una suddivisione per fondamentali e per squadra:

#### Differenza punti

Squadra	Pos	Campionato	Min	Max	Q <sub>1</sub>	Mediana	Media	Q <sub>3</sub>	Dev. std
Perugia		A1	6	25	11	14	15,05	18,75	5,41
Piacenza		A1	-19	22	-6,25	4,5	1,682	10,25	11,02

Taranto		A1	-20	17	13,75	-8	-5	1,75	11,1
Vibo Valentia		A2	-16	26	7,5	12,5	9,885	17	11,66
Castellana		A2	-19	19	-8	2	2,65	13	11,86
Motta di Livenza		A2	-18	14	-12,5	-7,5	-5	1,5	9,2
Pineto		A3	-20	34	10,25	14,5	12,62	19,75	12,08
Tuscania		A3	-22	26	-10	7	3,73	16	15,09
Brugherio		A3	-34	24	-13,75	-10	-6,54	-2	13,62
Sarroch		B	1	35	15	17	17,81	21,50	7,61
Lazio		B	-22	38	-11,25	9,5	6,04	19,75	17,51
Fenice		B	-35	25	-19,75	-11,5	-10,27	1,75	15,03

Tabella 8: statistiche descrittive della differenza punti per ogni squadra

Legenda:



= Posizionamento in alta classifica

= Posizionamento a metà classifica

= posizionamento in bassa classifica

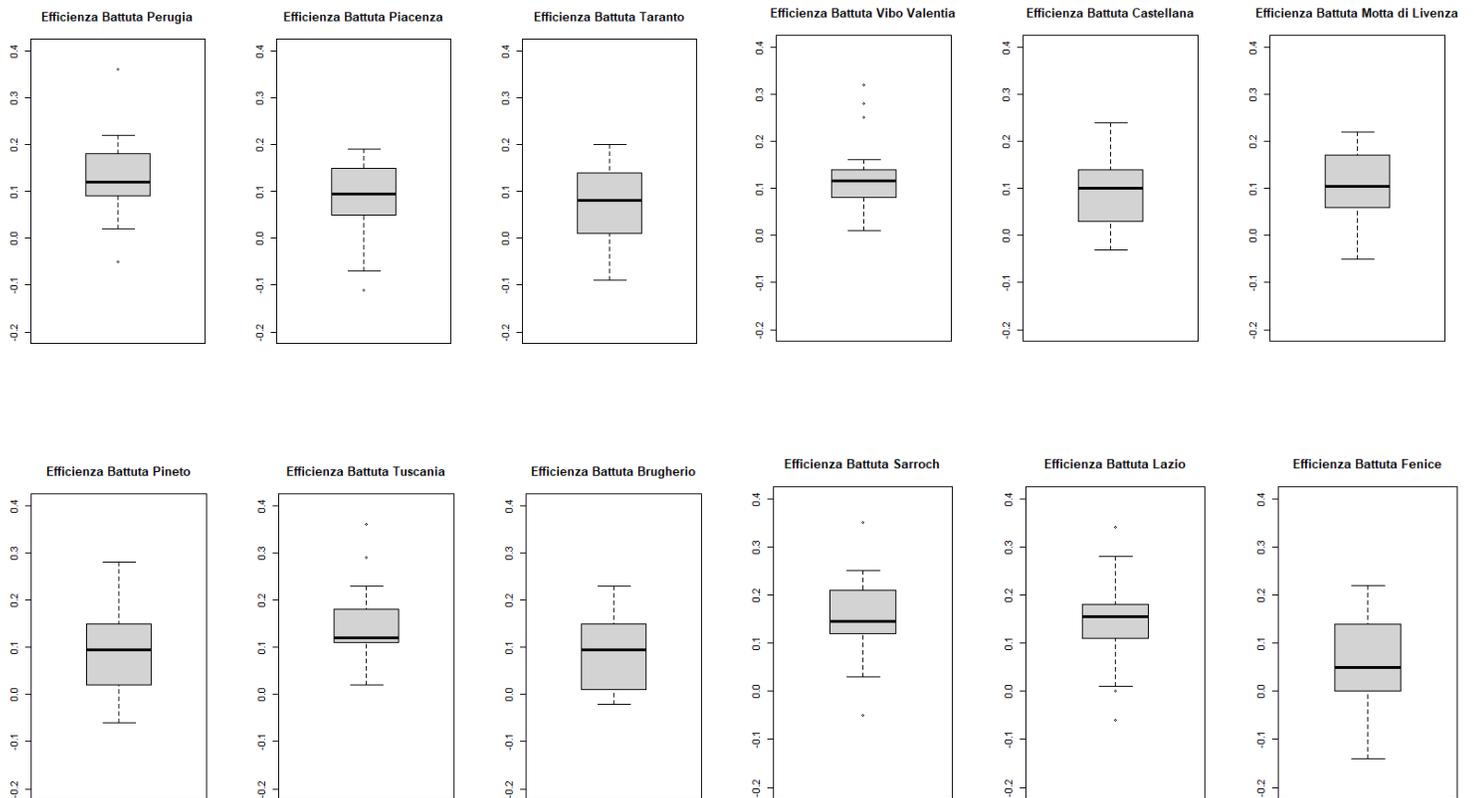
### Efficienza percentuale battuta

Squadra	Pos	Campionato	Min	Max	Q <sub>1</sub>	Mediana	Media	Q <sub>3</sub>	Dev. std
Perugia		A1	-0,05	0,36	0,092	0,12	0,1268	0,17	0,085
Piacenza		A1	-0,11	0,19	0,05	0,095	0,09	0,15	0,084
Taranto		A1	-0,09	0,2	0,0125	0,08	0,0677	0,14	0,086
Vibo Valentia		A2	0,01	0,32	0,0825	0,115	0,1223	0,14	0,07
Castellana		A2	-0,03	0,24	0,032	0,1	0,0942	0,14	0,08

Motta di Livenza		A2	-0,05	0,22	0,063	0,105	0,105	0,1625	0,071
Pineto		A3	-0,06	0,28	0,023	0,095	0,0923	0,15	0,085
Tuscania		A3	0,02	0,36	0,11	0,12	0,144	0,178	0,074
Brugherio		A3	-0,02	0,23	0,125	0,095	0,091	0,143	0,075
Sarroch		B	-0,05	0,35	0,12	0,145	0,158	0,208	0,077
Lazio		B	-0,06	0,34	0,113	0,155	0,147	0,18	0,085
Fenice		B	-0,14	0,22	0,003	0,05	0,061	0,128	0,099

Tabella 9: statistiche descrittive dell'efficienza percentuale in battuta per ogni squadra

Figura 1: Boxplot dell'efficienza in battuta di tutte le squadre di tutti i campionati



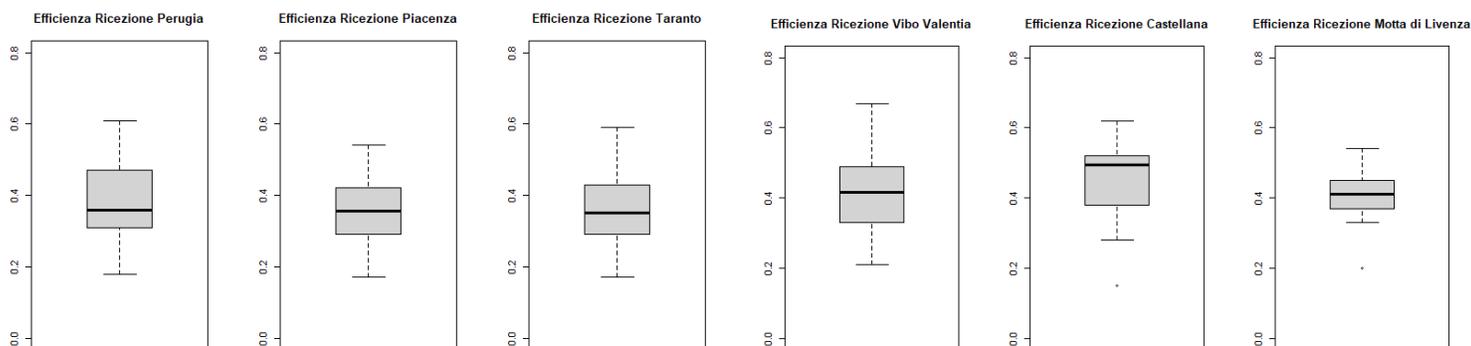
Fonte: Ns. elaborazione dati

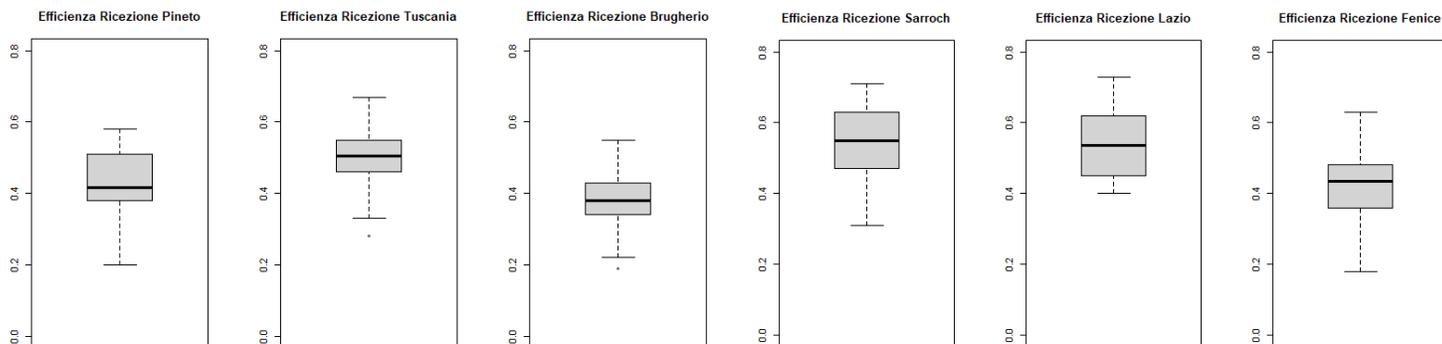
### Efficienza percentuale ricezione

Squadra	Pos	Campionato	Min	Max	Q <sub>1</sub>	Mediana	Media	Q <sub>3</sub>	Dev. std
Perugia	Green	A1	0,18	0,61	0,31	0,36	0,379	0,465	0,107
Piacenza	Orange	A1	0,17	0,54	0,293	0,355	0,35	0,413	0,101
Taranto	Red	A1	0,18	0,59	0,293	0,35	0,369	0,43	0,105
Vibo Valentia	Green	A2	0,21	0,67	0,33	0,415	0,412	0,488	0,126
Castellana	Orange	A2	0,15	0,62	0,383	0,495	0,453	0,52	0,113
Motta di Livenza	Red	A2	0,21	0,54	0,373	0,41	0,411	0,45	0,077
Pineto	Green	A3	0,20	0,58	0,383	0,429	0,429	0,508	0,097
Tuscania	Orange	A3	0,28	0,67	0,468	0,498	0,498	0,548	0,088
Brugherio	Red	A3	0,19	0,55	0,345	0,381	0,381	0,428	0,085
Sarroch	Green	B	0,31	0,71	0,47	0,537	0,537	0,628	0,104
Lazio	Orange	B	0,40	0,73	0,455	0,545	0,55	0,62	0,103
Fenice	Red	B	0,18	0,63	0,363	0,429	0,429	0,48	0,102

Tabella 10: statistiche descrittive dell'efficienza percentuale in ricezione per ogni squadra

Figura 2: Boxplot dell'efficienza in ricezione di tutte le squadre di tutti i campionati





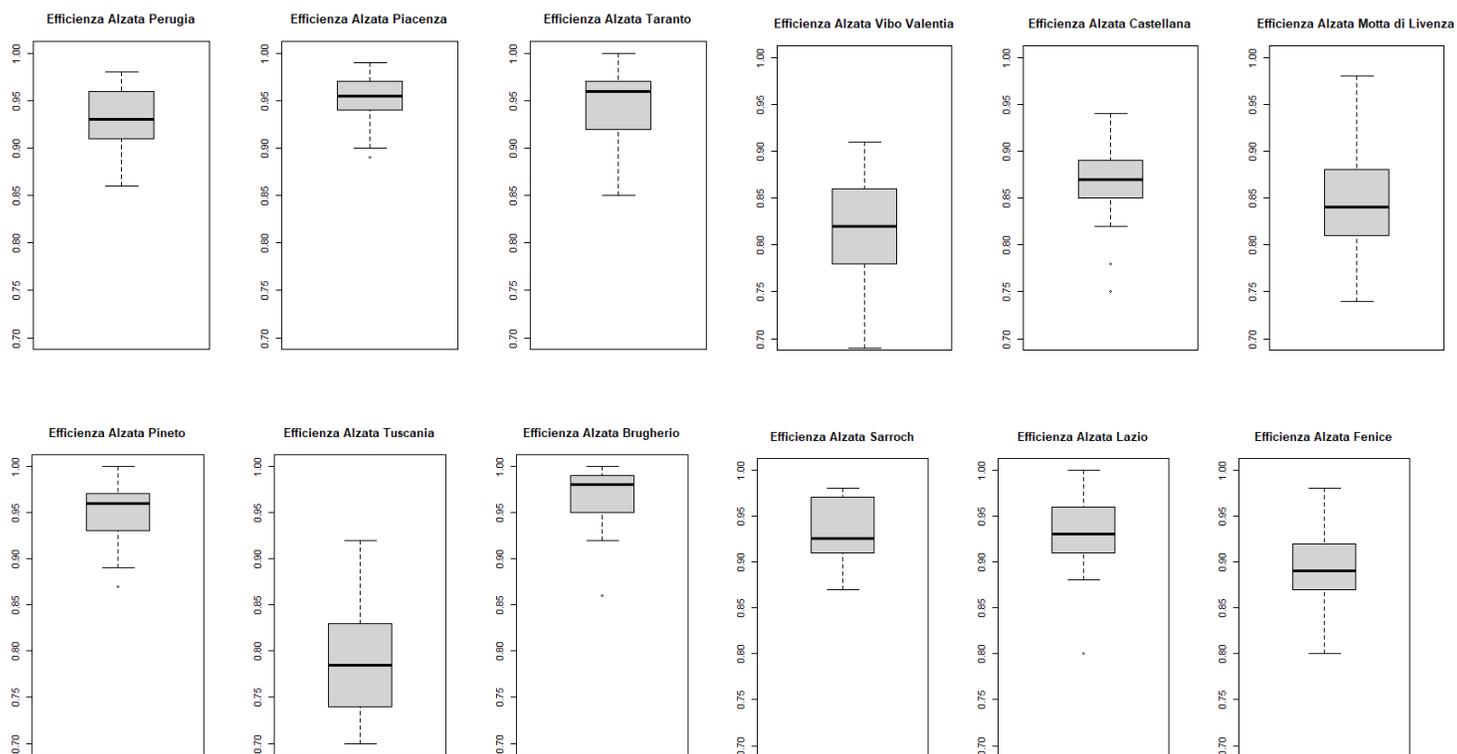
Fonte: Ns. elaborazione dati

### Efficienza percentuale alzata

Squadra	Pos	Campionato	Min	Max	Q <sub>1</sub>	Mediana	Media	Q <sub>3</sub>	Dev. std
Perugia		A1	0.860	0.980	0.912	0.930	0.931	0.958	0.029
Piacenza		A1	0.890	0.990	0.942	0.955	0.954	0.970	0.028
Taranto		A1	0.850	1.000	0.923	0.960	0.950	0.970	0.037
Vibo Valentia		A2	0.690	0.910	0.782	0.820	0.817	0.860	0.057
Castellana		A2	0.750	0.940	0.850	0.870	0.868	0.890	0.042
Motta di Livenza		A2	0.740	0.980	0.812	0.840	0.846	0.878	0.059
Pineto		A3	0.870	1.000	0.932	0.960	0.953	0.970	0.034
Tuscania		A3	0.700	0.920	0.742	0.785	0.793	0.827	0.064
Brugherio		A3	0.860	1.000	0.950	0.980	0.969	0.990	0.033
Sarroch		B	0.870	0.980	0.910	0.925	0.934	0.968	0.031
Lazio		B	0.800	1.000	0.910	0.930	0.933	0.960	0.044
Fenice		B	0.800	0.980	0.872	0.890	0.891	0.920	0.043

Tabella 11: statistiche descrittive dell'efficienza percentuale in alzata per ogni squadra.

Figura 2: Boxplot dell'efficienza in alzata di tutte le squadre di tutti i campionati.



Fonte: Ns. elaborazione dati

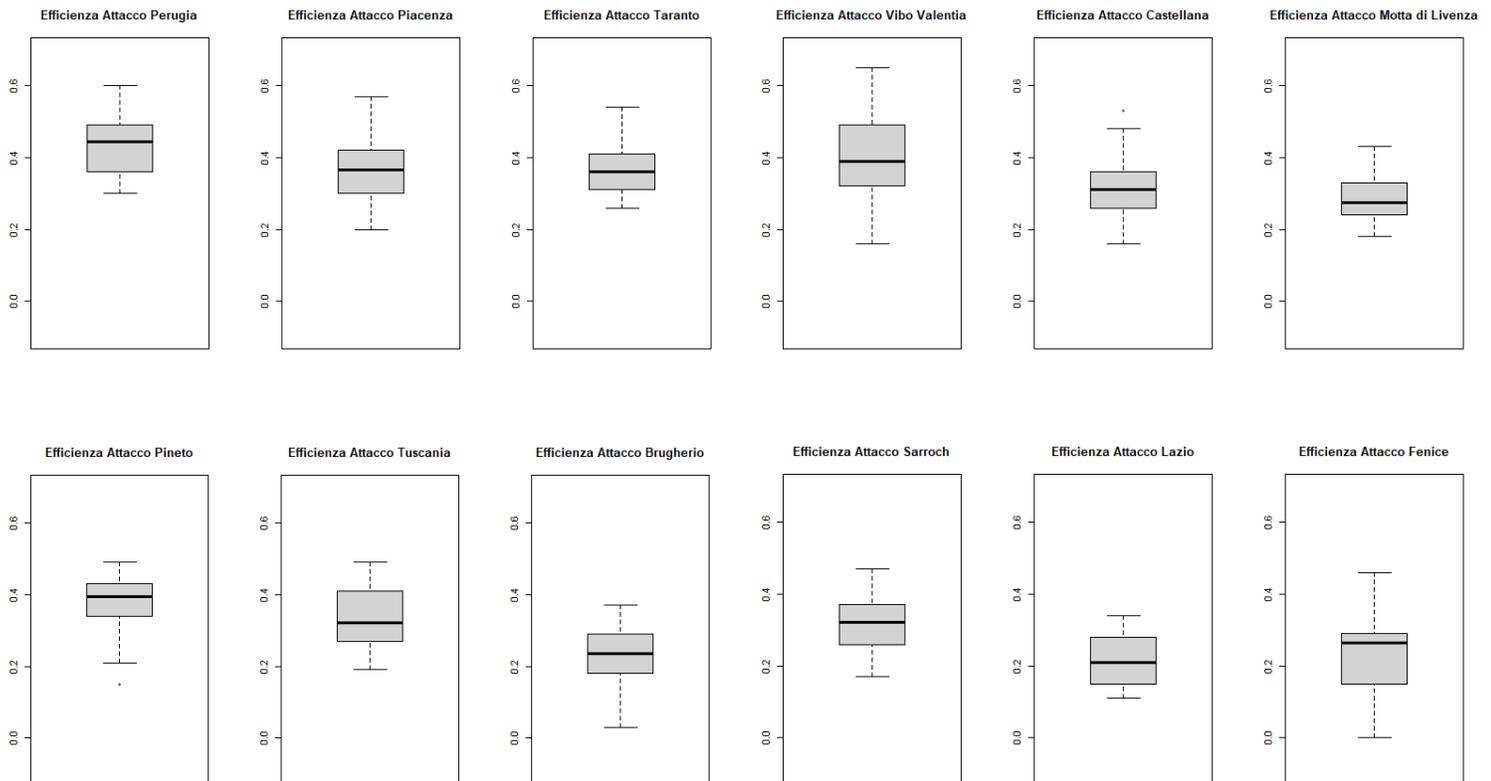
### Efficienza percentuale attacco

Squadra	Pos	Campionato	Min	Max	Q <sub>1</sub>	Mediana	Media	Q <sub>3</sub>	Dev. std
Perugia		A1	0,30	0,60	0,365	0,445	0,434	0,49	0,082
Piacenza		A1	0,20	0,57	0,305	0,365	0,368	0,413	0,080
Taranto		A1	0,26	0,54	0,360	0,360	0,365	0,405	0,069
Vibo Valentia		A2	0,16	0,65	0,325	0,390	0,398	0,480	0,129
Castellana		A2	0,16	0,53	0,26	0,31	0,314	0,360	0,085
Motta di Livenza		A2	0,18	0,43	0,24	0,275	0,283	0,328	0,060

Pineto		A3	0,15	0,49	0,343	0,395	0,377	0,430	0,081
Tuscania		A3	0,19	0,49	0,273	0,320	0,338	0,405	0,082
Brugherio		A3	0,03	0,37	0,185	0,235	0,240	0,290	0,078
Sarroch		B	0,17	0,47	0,260	0,320	0,315	0,365	0,072
Lazio		B	0,11	0,34	0,152	0,210	0,219	0,278	0,071
Fenice		B	0,00	0,46	0,153	0,265	0,231	0,290	0,105

Tabella 12: statistiche descrittive dell'efficienza percentuale in attacco per ogni squadra

Figura 3: Boxplot dell'efficienza in attacco di tutte le squadre di tutti i campionati.



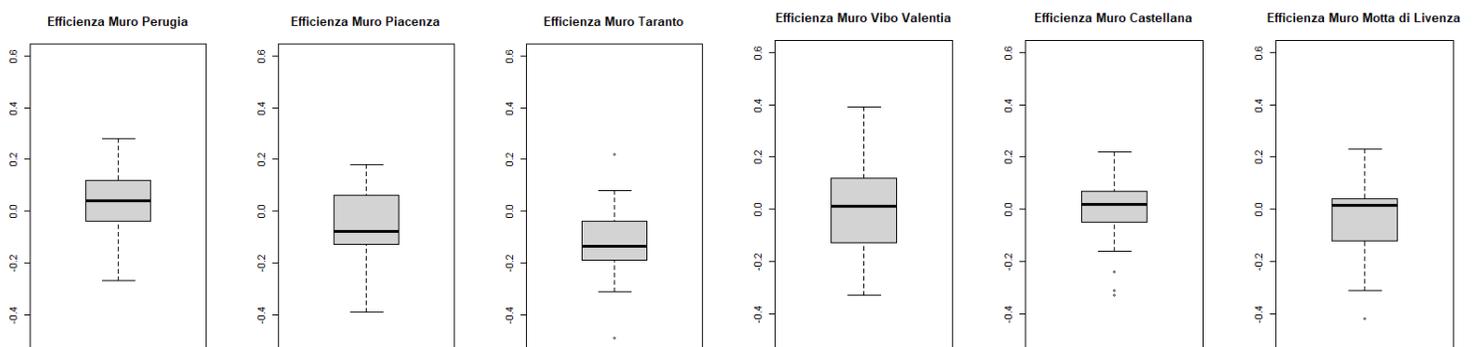
Fonte: Ns. elaborazione dati

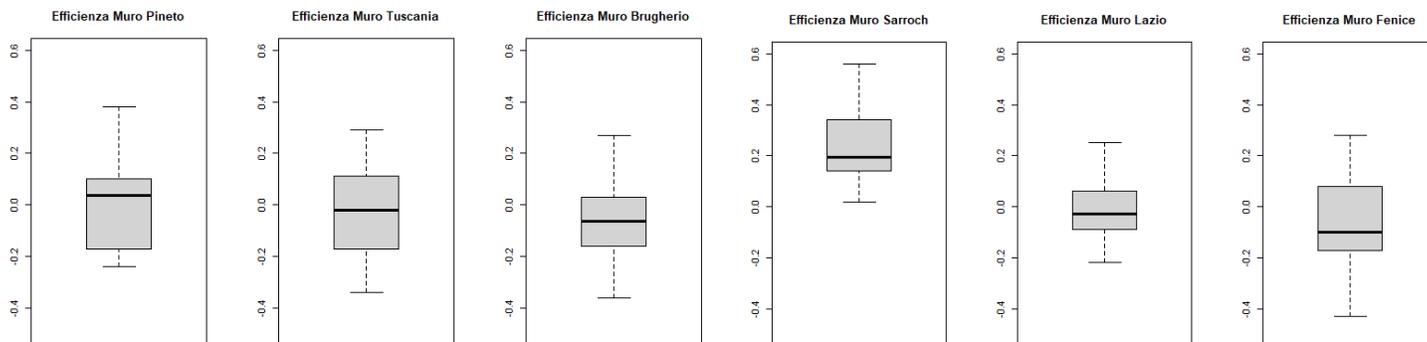
### Efficienza percentuale muro

Squadra	Pos	Campionato	Min	Max	Q <sub>1</sub>	Mediana	Media	Q <sub>3</sub>	Dev. std
Perugia		A1	-0.270	0.280	-0.038	0.040	0.025	0.115	0.132
Piacenza		A1	-0.390	0.180	-0.125	-0.080	-0.067	0.045	0.143
Taranto		A1	-0.490	0.220	-0.183	-0.135	-0.115	-0.048	0.149
Vibo Valentia		A2	-0.330	0.390	-0.122	0.010	-0.002	0.120	0.174
Castellana		A2	-0.330	0.220	-0.045	0.020	-0.005	0.068	0.137
Motta di Livenza		A2	-0.420	0.230	-0.115	0.015	-0.036	0.040	0.147
Pineto		A3	-0.240	0.380	-0.145	0.035	0.018	0.098	0.172
Tuscania		A3	-0.340	0.290	-0.165	-0.020	-0.020	0.108	0.168
Brugherio		A3	-0.360	0.270	-0.158	-0.065	-0.060	0.028	0.151
Sarroch		B	0.020	0.560	0.145	0.195	0.243	0.332	0.142
Lazio		B	-0.220	0.250	-0.090	-0.030	-0.006	0.055	0.123
Fenice		B	-0.430	0.280	-0.170	-0.100	-0.070	0.065	0.181

Tabella 13: statistiche descrittive dell'efficienza percentuale a muro per ogni squadra

Figura 4 :Boxplot dell'efficienza a muro di tutte le squadre di tutti i campionati.





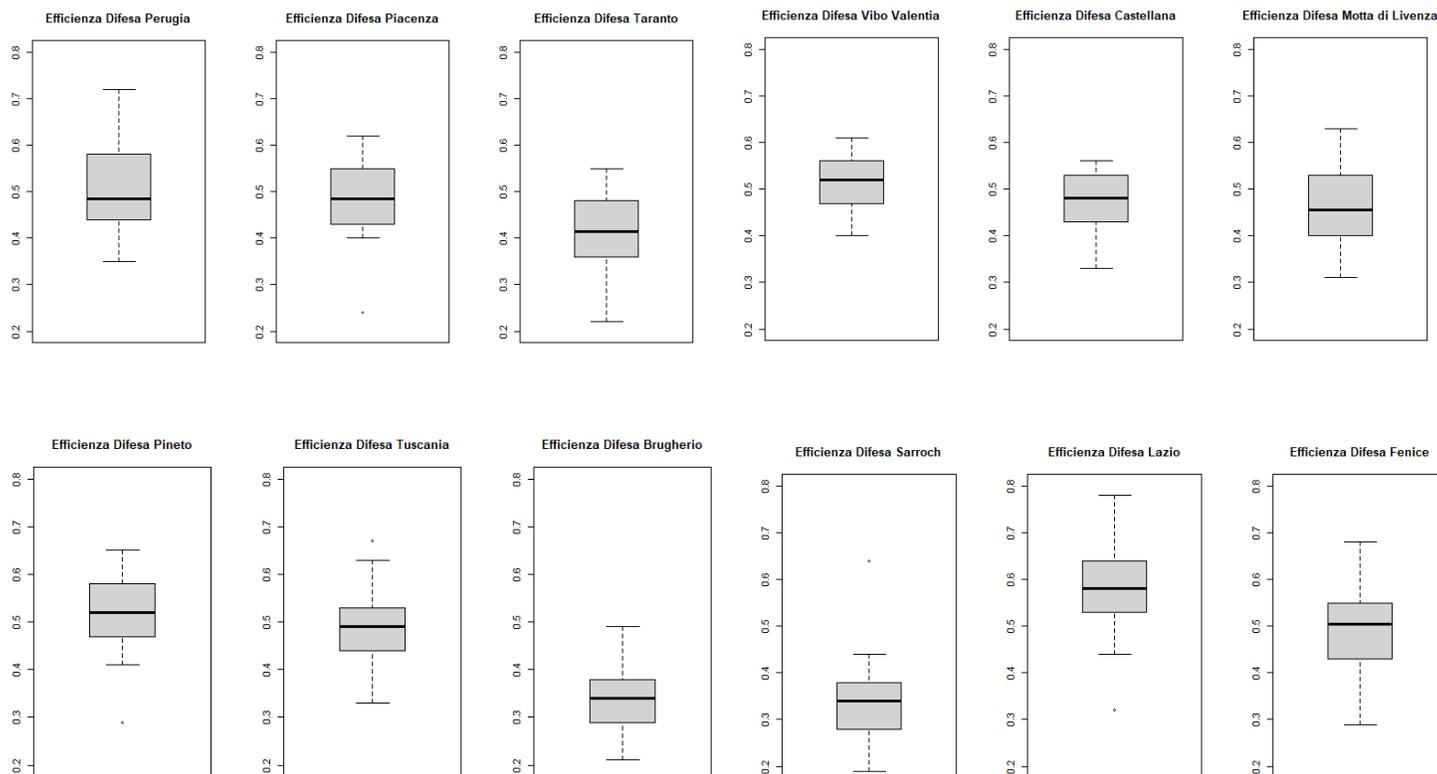
Fonte: Ns. elaborazione dati

### Efficienza percentuale difesa

Squadra	Pos	Campionato	Min	Max	Q <sub>1</sub>	Mediana	Media	Q <sub>3</sub>	Dev. std
Perugia		A1	0.350	0.720	0.440	0.485	0.505	0.573	0.097
Piacenza		A1	0.240	0.620	0.432	0.480	0.480	0.545	0.086
Taranto		A1	0.220	0.550	0.365	0.415	0.415	0.478	0.09
Vibo Valentia		A2	0.400	0.610	0.470	0.520	0.510	0.558	0.058
Castellana		A2	0.330	0.560	0.435	0.480	0.479	0.528	0.059
Motta di Livenza		A2	0.310	0.630	0.402	0.455	0.457	0.525	0.088
Pineto		A3	0.290	0.650	0.472	0.520	0.522	0.580	0.085
Tuscania		A3	0.330	0.670	0.442	0.490	0.486	0.528	0.08
Brugherio		A3	0.210	0.490	0.292	0.340	0.339	0.380	0.07
Sarroch		B	0.190	0.640	0.283	0.340	0.338	0.378	0.089
Lazio		B	0.320	0.780	0.530	0.580	0.585	0.637	0.096
Fenice		B	0.290	0.680	0.432	0.505	0.491	0.548	0.09

Tabella 14: statistiche descrittive dell'efficienza percentuale in difesa per ogni squadra

Figura 5: Boxplot dell'efficienza in difesa di tutte le squadre di tutti i campionati.



Fonte: Ns. elaborazione dati

## 4 Metodologia di analisi

La composizione del dataset che include l'osservazione di più unità (squadre) su differenti istanti temporali, ci permette di sviluppare dei modelli di regressione panel, particolarmente adatti all'analisi di dati per i quali l'analisi *cross-section* viene combinata con quella *time-series*. I dati panel, detti anche dati longitudinali, offrono l'opportunità di osservare e analizzare il comportamento delle singole unità nel tempo, permettendo di controllare per variabili non osservabili che possono influenzare le variabili di interesse. [2]

Uno dei principali vantaggi dell'analisi panel è la capacità di gestire l'eterogeneità individuale. Ogni squadra ha delle caratteristiche intrinseche che potrebbero influenzarne le performance e queste caratteristiche possono essere costanti nel tempo. Utilizzando dati panel,

possiamo controllare per queste differenze non osservabili tra le squadre.[3] Questo è particolarmente importante nel nostro caso, dove le squadre appartengono a diversi campionati e potrebbero avere diverse caratteristiche che influiscono sulle loro performance.

I dati panel consistono di osservazioni sulle stesse entità in più periodi. Se i dati contengono osservazioni sulle variabili X e Y, allora essi si indicano con

$$(X_{it}, Y_{it}), \quad \text{con } i = 1, \dots, n \quad \text{e } t = 1, \dots, T,$$

dove il primo pedice,  $i$ , fa riferimento alla singola unità che è oggetto dell'osservazione, mentre il secondo pedice,  $t$ , fa riferimento al periodo in cui questa è osservata.

## 4.1 Principali modelli di regressione panel

### 4.1.1 Modello a effetti fissi

Il modello ad effetti fissi tiene conto delle differenze non osservabili tra le unità (in questo caso, il tempo) che possono influenzare le variabili dipendenti. Queste differenze sono considerate come costanti (fisse) nel tempo e variano tra le unità. Questo modello è particolarmente utile quando si sospetta che queste differenze non osservabili siano correlate con le variabili esplicative.

L'equazione di un modello a effetti fissi è data da:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_i + \mu_{it}^3$$

In questa equazione la variabile  $Z_i$  rappresenta una variabile inosservata che varia da una categoria all'altra ma è invariata nel tempo. Il modello può essere interpretato come avente  $n$  intercette, una per ogni categoria, dove  $Z_i$  è variabile tra categorie ma costante nel tempo.

Se  $\alpha_i = \beta_0 + \beta_2 Z_i$ , allora l'equazione del modello di regressione a effetti fissi sarà:

$$Y_{it} = \alpha_i + \beta_1 X_{it} + \mu_{it}$$

L'errore  $\mu_{it}$  può essere scomposto in:

- $\mu_i$  che rappresenta è l'effetto non osservabile che varia tra individui;
- $\epsilon_{it}$  rappresenta il termine di errore idiosincratico, che cattura quindi caratteristiche non

---

<sup>3</sup> Stock, James H., and Mark W. 2005.

osservabili che influenzano  $Y_{it}$ . [4]

La pendenza della retta di regressione data da  $\beta_1$  sarà uguale per tutti le categorie, ma la sua intercetta varierà tra le diverse categorie.

#### 4.1.2 Modello a effetti casuali

A differenza del modello a effetti fissi, questo tipo di modello ipotizza che gli effetti non osservabili, che precedentemente sono stati indicati con  $\alpha_i$ , non siano inclusi all'interno dell'intercetta, bensì all'interno del termine di errore. Le differenze non osservabili tra le osservazioni si ipotizza quindi siano casuali, e non correlate con le variabili esplicative. Il modello a effetti casuali può essere scritto come:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \mu_{it}$$

dove  $\mu_{it}$  è composto da:

- $\mu_i$  che rappresenta lo specifico effetto casuale dell'unità  $i$
- $\epsilon_{it}$  che rappresenta l'errore idiosincratico, ovvero l'errore specifico per l'unità  $i$  e il tempo  $t$ .

## 4.2 Test di Hausman

Per scegliere il modello adatto per l'analisi dei dati panel, si esegue un test statistico chiamato test di Hausman<sup>4</sup>, che consente di determinare se sia preferibile utilizzare un modello a effetti fissi o un modello a effetti casuali. Il test di Hausman confronta le stime dei coefficienti ottenute dai due modelli per verificare la presenza di correlazione tra gli effetti casuali e le variabili esplicative. Le ipotesi del test di Hausman sono le seguenti:

- $H_0$  (ipotesi nulla): gli effetti casuali non sono correlati con le variabili esplicative. In questo caso, il modello a effetti casuali è da preferire perché si assume che le differenze tra le unità siano puramente casuali e non influenzino le variabili indipendenti del modello. Se l'ipotesi nulla è vera, le stime del modello a effetti casuali sono più

---

<sup>4</sup> Hausman J., 1978.

efficienti, in quanto riducono la varianza degli stimatori e migliorano la precisione delle stime.

- **H1 (Ipotesi Alternativa):** gli effetti casuali sono correlati con le variabili esplicative. Se questa ipotesi viene accettata, significa che esiste una correlazione significativa tra le variabili indipendenti e gli effetti specifici delle unità, suggerendo che il modello a effetti casuali potrebbe produrre stime distorte e inconsistenti. In questo caso, respingendo l'ipotesi nulla, si preferisce il modello a effetti fissi, poiché questo modello controlla meglio per le differenze specifiche delle unità che possono essere correlate con le variabili esplicative, fornendo stime più affidabili.

Il test di Hausman viene eseguito calcolando una statistica test che segue una distribuzione chi-quadrato con gradi di libertà pari al numero di variabili esplicative nel modello. Se il valore della statistica test supera una certa soglia critica (determinata dal livello di significatività scelto, solitamente 5%), si respinge l'ipotesi nulla a favore dell'ipotesi alternativa.

## 5 Analisi e sviluppo dei modelli

Un tentativo di analisi può essere effettuato considerando gli effetti che si manifestano nel tempo. In altre parole, anziché esaminare se i risultati delle squadre (espressi come differenza punti) siano influenzati esclusivamente dalle caratteristiche specifiche delle squadre stesse, si tenterà di comprendere se e come queste prestazioni variano nel corso del tempo.

Per questa analisi verranno prese in considerazione tutte le variabili, al fine di intercettare una più ampia porzione di variabilità della Y. Di seguito vengono riportati i risultati dei due modelli a effetti fissi e ad effetti variabili.

	<i>Effetti fissi</i>	<i>Effetti casuali</i>
<i>Intercetta</i>		-63.21 *** (6.18)

<i>X1</i>	47.07 *** (5.24)	45.52 *** (5.04)
<i>X2</i>	29.43 *** (3.81)	29,41 *** (3.69)
<i>X3</i>	70.69 *** (4.07)	69.69 *** (3.98)
<i>X4</i>	38.03 *** (2.57)	37.84 *** (2.47)
<i>X5</i>	13.30 ** (4.04)	16.17 *** (3.97)
<i>X6</i>	22.01 *** (5.90)	21.96 *** (5.86)
<i>Livello</i>	-2.98 (2.93)	0.47 (2.93)
<i>S. Idios</i>		7.11
<i>S. Ind</i>		0.00
<i>R<sup>2</sup></i>	0.79	0.78
<i>R<sup>2</sup> adj.</i>	0.76	0.77
<i>Num. Obs</i>	300	300

Note: \*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05

*Tabella 15: Riepilogo dell'output di regressione dei modelli a effetti fissi e ad effetti variabili*

Si noti come, all'interno di questo output di regressione, siano significativi tutti i coefficienti, con un livello di significatività maggiore dell'1%. L'unica variabile che non presenta alcun livello di significatività è quella del livello: questo ci suggerisce che un'analisi legata al livello quella squadra non produrrebbe dei risultati significativi.

Di seguito vengono riportati gli effetti temporali del modello a effetti fissi

<b>TEMPO</b>	<b>EFFETTO RELATIVO AL TEMPO</b>	<b>TEMPO</b>	<b>EFFETTO RELATIVO AL TEMPO</b>
<b>1</b>	-48.695	14	-54.601
<b>2</b>	-48.543	15	-59.382
<b>3</b>	-51.191	16	-57.653
<b>4</b>	-50.911	17	-57.129
<b>5</b>	-49.134	18	-56.999
<b>6</b>	-52.182	19	-56.633
<b>7</b>	-56.522	20	-58.314
<b>8</b>	-50.732	21	-62.025
<b>9</b>	-51.456	22	-59.563
<b>10</b>	-54.655	23	-56.074
<b>11</b>	-54.860	24	-64.941
<b>12</b>	-57.650	25	-64.714
<b>13</b>	-55.801	26	-61.838

*Tabella 16: Riepilogo degli effetti fissi rispetto al tempo*

Per comprendere meglio la distribuzione degli effetti fissi, vengono rappresentati graficamente.

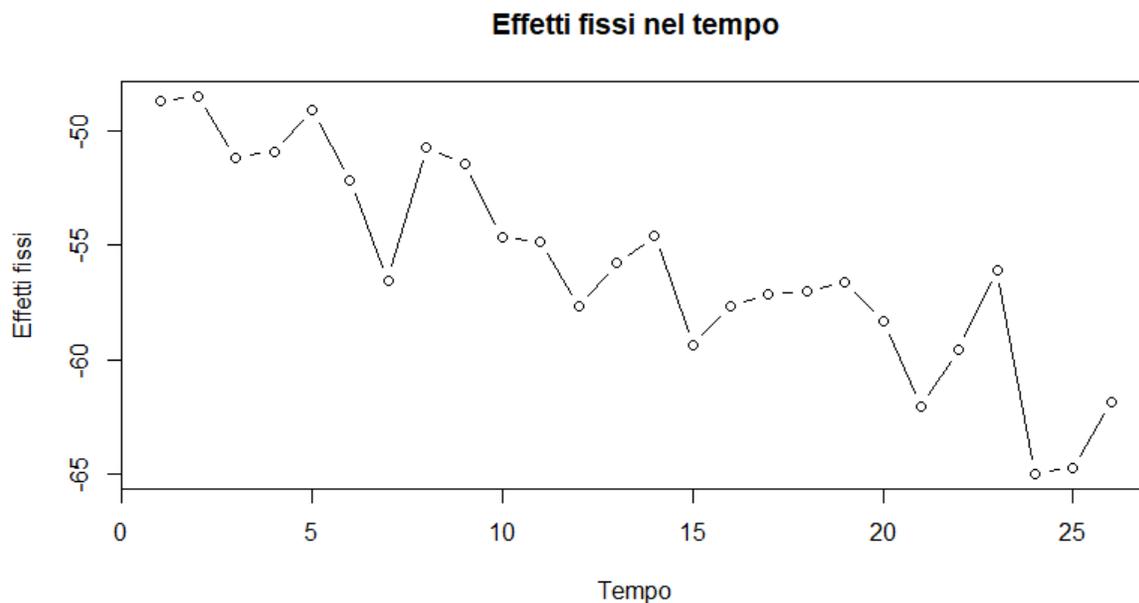


Figura 6: Evoluzione degli effetti fissi nel tempo

Fonte: Ns. Elaborazione

Si noti come applicando il modello a effetti fissi sembra ci sia un trend decrescente: le intercette delle diverse rette di regressione quindi avranno valori decrescenti. Per interpretare correttamente questo dato possiamo dire che all'aumentare del tempo, sarà necessaria un'efficienza più alta nei fondamentali per ottenere una stessa differenza punti.

Viene effettuato il test di Hausman per capire quale dei due modelli è più efficiente.

<b>Statistica chi quadrato</b>	<b>P-valore</b>	<b>Ipotesi accettata</b>
21.163	0.003537	H1

Tabella 17: riepilogo dell'output del test di Hausmann

Accettando l'ipotesi alternativa del test di Hausman, concludiamo che il modello migliore è quello a effetti fissi, che ha come ipotesi che gli effetti casuali del modello sono correlati con le variabili esplicative.

## 5.1 Analisi alternativa

Nel caso si volesse studiare in modo più approfondito e preciso gli effetti non catturati dalle variabili all'interno di ciascun campionato, il dataset originario può essere suddiviso in tre subset distinti. Questa suddivisione mira a creare gruppi omogenei di squadre in base al posizionamento in classifica che hanno conseguito. In pratica, ciascun subset conterrà squadre che si sono posizionate in fasce di classifica simili: alte, medie o basse.

Questa metodologia consente di isolare e studiare gli effetti costanti nel tempo di ogni squadra, consentendo un'analisi più dettagliata e comparativa delle prestazioni delle squadre che occupano posizioni simili in classifica, indipendentemente dal campionato di appartenenza.

## 5.2 Scelta delle variabili

Come per l'analisi precedentemente effettuata, si prenderanno in considerazione tutte le variabili al fine di intercettare una più ampia porzione di variabilità della variabile Y.

Di seguito vengono riportati i risultati dei tre modelli che sono stati sviluppati.

	<i>Effetti fissi</i>	<i>Effetti casuali</i>
<i>Intercetta</i>		-34.44 *** (10.00)
<i>X1</i>	35.55 *** (7.69)	33.65 *** (7.64)
<i>X2</i>	15.06 ** (5.60)	17.55 *** (5.31)
<i>X3</i>	51.81 *** (6.51)	51.01 *** (6.31)

<i>X4</i>	25.92 *** (3.99)	27.69 *** (3.49)
<i>X5</i>	7.04 (7.74)	6.91 (6.12)
<i>X6</i>	-18.46 (15.60)	12.88 (9.14)
<i>S. Idios</i>		5.80
<i>S. Id</i>		1.18
$R^2$	0.65	0.67
$R^2$ <i>adj.</i>	0.62	0.64
<i>Num. obs</i>	100	100

Note: \*\*\*  $p < 0.001$ ; \*\*  $p < 0.01$ ; \*  $p < 0.05$

Tabella 18: Riepilogo dell'output di regressione per le squadre di fascia bassa.

	<i>Effetti fissi</i>	<i>Effetti casuali</i>
<i>Intercetta</i>		-69.45*** (10.11)
<i>X1</i>	51.04 *** (9,12)	50.10 *** (9.20)

X2	37.97 ** (7.58)	34.57 *** (6.70)
X3	79.28 *** (9.62)	72.79*** (8.02)
X4	42.32 *** (5.36)	44.45 *** (5.36)
X5	15.56 (10.21)	17.90 (9.70)
X6	-8.72 (17.43)	23.16* (10.46)
<i>S. Idios</i>		7.43
<i>S. Id</i>		0.00
$R^2$	0.77	0.75
$R^2$ adj.	0.74	0.73
<i>Num. obs</i>	100	100

Note: \*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05

Tabella 19: Riepilogo dell'output di regressione per le squadre di fascia media

	<i>Effetti fissi</i>	<i>Effetti casuali</i>
<i>Intercetta</i>		-64.46*** (12.46)
<i>X1</i>	38.58 *** (8.67)	44.44 *** (8.26)
<i>X2</i>	29.16 ** (7.64)	29.79*** (7.48)
<i>X3</i>	72.35 *** (9.47)	63.87 *** (7.52)
<i>X4</i>	32.68 *** (4.58)	35.27 *** (4.46)
<i>X5</i>	16.70 (8.89)	7.40 (7.70)
<i>X6</i>	20.02 (16.22)	26.10* (11.55)
<i>S. Idios</i>		6.38
<i>S. Id</i>		1.99
<i>R<sup>2</sup></i>	0.73	0.73
<i>R<sup>2</sup> adj.</i>	0.71	0.71
<i>Num. obs</i>	100	100

Note: \*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05

Tabella 20: Riepilogo dell'output di regressione per le squadre di fascia alta

Vengono riportate, di seguito, delle tabelle riassuntive degli effetti fissi dei diversi modelli.

	Serie A	Serie A2	Serie A3	Serie B
Fascia alta	-4.6606	-9.7552	-3.1921	-3.6490
Fascia media	-41.418	-43.895	-49.053	-40.231
Fascia bassa	-66.966	-64.891	-61.559	-65.596

Tabella 21: Effetti fissi per campionato e per fascia di classifica

### Test di Hausman

Come detto in precedenza, per capire quale dei due modelli (se a effetti fissi o casuali) è più adatto ai dati in analisi, si effettua il test di Hausman, approfondito nel capitolo 4.2. Di seguito vengono riportati i risultati di questo test effettuato su tutti i modelli di regressione precedentemente illustrati.

<i>Squadre di alta classifica</i>		
<i>Statistica chi quadrato</i>	P-valore	Ipotesi accettata
5.8766	0.4372	H0

Tabella 22: Test di Hausman per le squadre di alta classifica

<i>Squadre di media classifica</i>		
<i>Statistica chi quadrato</i>	P-valore	Ipotesi accettata
23.752	0.0005802	H1

Tabella 23: Test di Hausman per le squadre di media classifica

<i>Squadre di bassa classifica</i>		
<i>Statistica chi quadrato</i>	P-valore	Ipotesi accettata
6.0216	0.4208	H0

*Tabella 24: Test di Hausman per le squadre di bassa classifica*

I test di Hausman ci suggeriscono che per le squadre di alta classifica e bassa classifica le stime del modello a effetti casuali sono più efficienti e precisi, mentre per le squadre di media classifica possiamo concludere che gli effetti casuali sono correlati con le variabili esplicative, preferendo quindi il modello a effetti fissi.

Una conclusione che può essere tratta osservando i modelli, è che una suddivisione per livelli non è particolarmente efficace, come d'altronde era stato anticipato in seguito all'analisi del primo modello.

## **6 Conclusioni**

In questo lavoro è stato affrontato il tema della rilevazione statistica nel mondo della pallavolo e sono state applicate metodologie avanzate di analisi dei dati attraverso modelli di regressione panel.

L'obiettivo principale è stato quello di comprendere se le prestazioni delle squadre possano essere influenzate non solo dalle caratteristiche tecniche e tattiche individuali e di squadra, ma anche da variabili non immediatamente osservabili, come il fattore tempo o le specificità di ogni campionato.

Nella prima parte, l'importanza della raccolta dati è stata evidenziata attraverso l'analisi del ruolo dello scoutman e delle tecniche di rilevazione e analisi delle performance. La rilevazione statistica nel contesto sportivo, e in particolare nella pallavolo, si è dimostrata uno strumento indispensabile per supportare le decisioni tattiche e strategiche delle squadre. La precisione dei dati raccolti consente, infatti, di misurare con esattezza l'efficacia e l'efficienza dei fondamentali di gioco, permettendo alle squadre di ottimizzare le prestazioni durante la gara.

La seconda parte dell'analisi è incentrata sull'introduzione di modelli di regressione panel che, sulla base dei dati raccolti, consentono un'analisi più mirata delle performance delle squadre. Tramite l'uso dei modelli a effetti fissi e casuali, è stato possibile osservare come variabili nascoste, come l'appartenenza a determinati campionati o il trascorrere del tempo, possano influire sulle prestazioni delle squadre. I risultati ottenuti hanno mostrato che, sebbene sussistano differenze tra squadre di diverse fasce e campionati, esiste una correlazione significativa tra l'avanzare delle giornate, e l'efficienza nei fondamentali di gioco, evidenziando come con l'avanzare del campionato è necessaria una maggiore efficienza per mantenere o migliorare la differenza punti.

## Fonti

- [1] Annalisa Pinto. *Rilevazione statistica e match analysis nella pallavolo*. Calzetti&Mariucci editori
- [2] Stock, James H., and Mark W. Watson. *Introduzione All'econometria*. Pearson Education Italia, 2005.
- [3] Das, and Ghosh. *Econometrics in Theory and Practice*. Springer Singapore, 2019.
- [4] Wooldridge, J. M. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, 2010
- [5] Miguel Silva, Rui Marcelino, Daniel Lacerda, Paulo Vincente Joao. *Match Analysis in volleyball, a systematic review*, 2016
- [6] Kasper Vourinen. *Modern Volleyball Analysis and Training Periodization*, 2018
- [7] Hausman J., 1978. Specification tests in econometrics. *Econometrica* 46, 1251-1271