

**UNIVERSITA' DEGLI STUDI DI BRESCIA
CORSO DI LAUREA MAGISTRALE IN MANAGEMENT
CURRICULUM IN MARKETING**



TESI DI LAUREA

**BIG DATA E MACHINE LEARNING:
NUOVI ORIZZONTI PER L'ANALISI DEI DATI NEL BASKET**

**RELATORE:
CHIAR.MA PROF.SSA PAOLA ZUCCOLOTTO**

**CORRELATORE:
CHIAR.MA PROF.SSA MARICA MANISERA**

**LAUREANDO:
DAVIDE VERZELETTI
MATRICOLA N. 84520**

ANNO ACCADEMICO 2015/2016

INDICE:

CAPITOLO	TITOLO	PAG.
1	Introduzione	5
1.1	Definizione del tema	5
1.2	Gli obiettivi e la struttura dell'analisi	6
2	Il dataset	8
2.1	Caratteristiche principali	8
2.1.1	Il web scraping: l'estrazione dei dati	10
2.1.2	Errori iniziali presenti nel dataset	11
2.2	Analisi preliminari svolte per Scaligera Basket Verona	15
3	Gli alberi di decisione	41
3.1	Alberi di classificazione e di regressione	41
3.2	L'algoritmo CART	41
3.2.1	L'indice di impurità: l'indice di Gini	42
3.2.2	La bontà del modello: l'indice AUC	43
4	La performance nel tiro in situazioni di stress	45
4.1	Le situazioni di stress	45
4.2	Il modello elaborato	46
4.3	Nuove misurazioni della performance di tiro	48
5	Analisi sul girone di andata	51
5.1	Rappresentazioni grafiche dei nuovi indici	51
6	Analisi sul girone di ritorno	66
6.1	Variazioni dell'albero di classificazione	66
6.2	Variazioni degli indici e movimenti di classifica	68
7	Conclusioni	82

Indice tabelle	85
Indice figure	85
Ringraziamenti	87
Riferimenti bibliografici	88

1 Introduzione

1.1 Definizione del tema

Negli ultimi decenni si è assistito ad una rapida crescita dell'interesse riguardante l'applicazione di metodi di analisi statistica al mondo dello sport, come documentato dalle numerose pubblicazioni scientifiche sul tema e anche da un vasto insieme di studi statistici condotti su dataset provenienti da diversi sport come football Americano, baseball, basket e hockey su ghiaccio (Albert, Bennet, Cochran, 2005; Albert, Koning, 2007). Basti pensare che nel 2005 è stato lanciato *The Journal of Quantitative Analysis in Sports*, la prima rivista accademica dedicata interamente alle ricerche in questo campo (Albert, Koning, 2008).

Per quanto riguarda il basket, sono state applicate diverse tecniche statistiche con una grande varietà di obiettivi, che spaziano dal semplice raffigurare le caratteristiche principali del gioco utilizzando statistiche descrittive (Kubatko, Oliver, Pelton, Rosenbaum et al., 2007) all'indagine di problemi più complessi, come la previsione dell'esito finale di una partita o di un campionato (West et al., 2008; Loeffelholz, Bednar, Bauer et al., 2009; Brown, Skol et al., 2010; Gupta, 2015; Lopez, Matthews, 2015; Ruiz, Perez-Cruz, 2015; Yuan, Liu, Kaufman, Reece, Bull, Franks, Wangs, Illushin, Bornn, 2015; Manner, 2016), l'analisi della performance dei giocatori (Page, Fellingham, Reese, 2007; Cooper, Ruiz, Sirvent, 2009; Piette, Anand, Zhang, 2010; Fearnhead, Taylor, 2011; Ozmen, 2012; Page, Barney, McGuire, 2013; Deshpande, Jensen, 2016), tracciare e studiare la rete di passaggi dei giocatori al canestro (Skinner, 2010) e il loro posizionamento (Shortridge, Goldsberry, Adams, 2014) o identificare strategie di gioco ottimali (Annis et al., 2006).

Ciononostante, la varietà di possibili tematiche che possono essere affrontate tramite le analisi statistiche è in continua crescita grazie alla disponibilità online di grandi dataset contenenti informazioni play-by-play (ossia la precisa sequenza delle azioni avvenute durante una partita, anche dai campionati di minor livello) e all'aumento di tecniche e strumenti computazionali che permettono l'estrazione dei dati (Hand, Mannila, Smyth, 2001; Han, Pei, Kamber, 2011). In verità, da un punto di vista metodologico il grande ammontare di dati disponibili riguardanti azioni e strategie di gioco, combinato con l'assenza di una provata teoria che spieghi la relazione tra le variabili maggiormente considerate, rende tali tematiche una sfida interessante per gli analisti (Carpita, Sandri, Simonetto, Zuccolotto, 2014).

(Zuccolotto, Manisera, Sandri, 2016).

1.2 Gli obiettivi e la struttura dell'analisi

Questo lavoro è stato svolto in seguito all'esperienza di stage presso la società Scaligera Basket Verona e si inserisce all'interno di BDSports, un progetto sviluppato dal Big&Open Data Innovation Laboratory dell'università di Brescia (bodai.unibs.it/BDSports). Parte dello studio svolto verrà presentato nell'ambito della *9th International Conference of the ERCIM WG on Computational Methodological Statistics* (CMStatistics 2016), conferenza internazionale che si terrà all'università di Siviglia nei giorni 9 – 10 – 11 Dicembre 2016.

L'obiettivo principale consiste nell'analisi della performance di tiro delle singole squadre tenendo conto di alcuni fattori che possono generare situazioni di gioco caratterizzate da forte pressione o stress psicologico.

Dal punto di vista statistico, occorre tenere presente alcune considerazioni. Primo, è necessario un numero molto elevato di dati provenienti dai play-by-play per poter ottenere robuste stime delle probabilità di fare canestro; per questo motivo, si parla di “big data” con riferimento al dataset utilizzato per le analisi svolte. Secondo, il fenomeno analizzato è caratterizzato da relazioni e interazioni complesse esistenti tra le variabili coinvolte, quindi si propone l'uso di tecniche di analisi dei dati multivariate, adatte a tenere conto sia della complessità che del numero elevato di dati: in particolare, si fa riferimento all'algoritmo CART (che verrà spiegato successivamente in questo lavoro), il quale rientra nella categoria delle tecniche di “machine learning”. Terzo, analizzando situazioni di gioco generanti stress e pressione, occorre considerare che, secondo alcuni studi psicologici (ad esempio Taylor, 1987; Madden, Kirkby, McDonald, Summers, Brown, King, 1995), alcuni atleti percepiscono la situazione competitiva come stimolante e altri percepiscono la medesima situazione come stressante e generatrice di ansia. Per questo motivo, potrebbe essere difficile individuare statisticamente delle situazioni di gioco sotto stress partendo da un dataset che include diverse squadre poiché la media generale potrebbe rimanere invariata, come effetto di un miglioramento della performance da parte di alcuni team e di un peggioramento da parte di altri (la media stessa della squadra potrebbe non modificarsi, per effetto del miglioramento e peggioramento dei giocatori al suo interno).

Riassumendo, l'analisi verrà condotta tramite una serie di step consecutivi con i seguenti obiettivi intermedi:

1. sviluppare un modello che descriva l'impatto di determinate situazioni di gioco stressanti sulla probabilità di segnare un canestro. Tale modello dovrà permettere (a) di determinare quali situazioni hanno, in media, un effetto significativo sulla probabilità di fare canestro e

- (b) di fornire nuove misurazioni della performance di tiro, integrando le statistiche di tiro correntemente utilizzate;
2. analizzare le performance di tiro delle singole squadre al termine del girone di andata e individuare quali squadre mostrano un miglioramento della performance nelle situazioni di stress individuate nello step 1, contrapposte a quelle che mostrano un peggioramento;
 3. condurre un'analisi sul solo girone di ritorno (escludendo il girone di andata) e osservare se si siano verificati miglioramenti e/o peggioramenti delle performance di tiro che portino a relativi miglioramenti e/o peggioramenti della posizione in classifica della squadra stessa, evidenziando in questo modo l'eventuale relazione che si verrebbe a creare.

L'analisi si articola nel modo seguente. Nel capitolo 2 viene presentato il dataset utilizzato: le sue caratteristiche principali e le analisi preliminari. Il capitolo 3 propone alcuni richiami e concetti teorici riguardanti uno dei metodi di machine learning utilizzati nei capitoli seguenti. Nel capitolo 4 vengono individuate le situazioni di stress, i loro impatti sulla probabilità di fare canestro e le nuove misurazioni della performance di tiro. Il capitolo 5 espone i risultati dell'analisi condotta al termine del girone di andata, mentre l'analisi sul girone di ritorno è descritta nel capitolo 6. Nel capitolo 7 si illustrano alcune considerazioni finali.

2 Il dataset

2.1 Caratteristiche principali

Il dataset sul quale sono state effettuate le analisi proviene dalle sequenze play-by-play di tutte le partite della stagione regolare disputate durante il campionato Italiano maschile di basket “Serie A2” (che costituisce il secondo gradino della Lega Italiana di basket, appena sotto la categoria più importante “Serie A”). I dati sono disponibili sul sito www.legapallacanestro.com nella sezione “Serie A2”: a seconda del girone (vi sono 2 gironi, “Est” e “Ovest”), si possono osservare i risultati della varie giornate e, cliccando sul risultato della partita di interesse, si accede alla pagina contenente il box score (ossia il tabellino della partita) e il play-by-play.

Una volta estratto dal web, tramite una funzione del programma R che verrà spiegata nel paragrafo 2.1.1, il play-by-play si presenta in questa forma (l’esempio riguarda una partita della prima giornata del campionato, Verona contro Trieste):

	Tempo	Alma-Agenzia per il Lavoro Trieste	Punti	Margine	Tezenis Verona
1	00:00	Fine partita	39-68	-29	
2	00:09		39-68	-29	Marco Spanghero, rimbalzo difensivo
3	00:10	team, tiro da tre punti sbagliato	39-68	-29	
4	00:32		39-68	-29	team, tiro da due punti realizzato
5	00:38	Lorenzo Baldasso, palla persa	39-66	-27	
6	00:39		39-66	-27	team, recupero
7	00:47		39-66	-27	Marco Spanghero, palla persa
8	00:48	Lorenzo Baldasso, recupero	39-66	-27	
9	00:52	team, palla persa trattamento di palla	39-66	-27	
10	00:53		39-66	-27	Marco Spanghero, recupero

Figura 2.1 – Esempio di play-by-play

Come si può notare, risulta poco pratico ai fini di qualsiasi analisi statistica. Per ovviare a questo inconveniente, sono state apportate numerose modifiche alla struttura visiva del play-by-play stesso, tramite centinaia di istruzioni contenute in molteplici codici in R, in modo da renderlo più comprensibile e utilizzabile. Innanzitutto gli eventi sono stati ordinati cronologicamente, dall’inizio alla fine della partita (come si vede nella figura 2.1, inizialmente sono riportati in ordine inverso); dopodiché sono state aggiunte colonne per ogni singola voce contenuta del play-by-play, sono stati corretti alcuni errori presenti inizialmente (vedasi il paragrafo 2.1.2) e sono state aggiunte alcune variabili deducibili dalle informazioni di base presenti.

Si giunge, così, ad un dataset in cui ad ogni riga corrisponde un tiro, per il quale vengono rilevate una serie di informazioni:

- *Match, Quarto, Minuti, Secondi, Tempotot*: indicano in quale partita, in che quarto, in che minuti e a quanti secondi avviene il tiro (*Tempotot* consiste nella rilevazione del tempo in secondi anziché in minuti);
- *PtA, PtB, Margine*: indicano il punteggio delle squadre e il margine tra esse al momento del tiro;
- *PtricA, PtricB, Margineric*: indicano punteggio e margine ricalcolati. In caso di un tiro segnato, infatti, il punteggio viene aggiornato nella medesima riga in cui viene rilevato il tiro; in realtà, al momento del tiro il punteggio è ancora fermo e si aggiorna solo dopo l'esito del tiro. Queste variabili permettono di imputare l'eventuale canestro alla riga successiva;
- *secs* indica i secondi di possesso palla da parte della squadra;
- *change* vale 1 quando cambia la squadra in possesso palla, 0 altrimenti;
- *secs14* indica i secondi di possesso palla della squadra "depurati" dalle azioni aggiuntive di 14 secondi;
- *secs14id* vale 1 quando si tratta di azioni aggiuntive di 14 secondi, 0 altrimenti;
- *secs14count* indica i secondi di possesso palla nelle azioni aggiuntive di 14 secondi;
- *Pl, Event, Team, Opp*: indicano il giocatore, l'evento (tiro da 2 punti, da 3 punti o tiro libero), la squadra di appartenenza (*Team*) e la squadra avversaria (*Opp*);
- *Rank, Score*: indicano posizione e relativo punteggio in classifica della squadra di interesse;
- *Rank.Opp, Score.Opp*: indicano posizione e relativo punteggio in classifica della squadra avversaria;
- *h.a, w.l*: indicano se la partita è stata giocata in casa (home) o in trasferta (away) e se è stata vinta (win) o persa (lose);
- *Diff.Rank, Diff.Score*: indicano la differenza di posizione e punteggio in classifica tra le due squadre;
- *Top.Rank, Top.Score, Top.Rank.Opp, Top.Score.Opp*: indicano la differenza di posizione e punteggio in classifica della squadra oggetto di interesse e dell'avversario rispetto alla squadra che occupa la prima posizione in classifica;
- *Off.Rank, Off.Score, Off.Rank.Opp, Off.Score.Opp*: indicano la differenza di posizione e punteggio in classifica della squadra oggetto di interesse e dell'avversario rispetto alle squadre che occupano le posizioni 7-8-9 in classifica (ossia in lotta per i play-off);

- *Out.Rank, Out.Score, Out.Rank.Opp, Out.Score.Opp*: indicano la differenza di posizione e punteggio in classifica della squadra oggetto di interesse e dell'avversario rispetto alle squadre che occupano le posizioni 11-12-13 in classifica (ossia in lotta per i play-out);
- *Bottom.Score, Bottom.Score.Opp*: indicano la differenza di punteggio in classifica della squadra oggetto di interesse e dell'avversario rispetto alla squadra che occupa l'ultima posizione in classifica;
- *Seq.wl* indica la sequenza di vittorie e/o sconfitte nelle 3 gare precedenti;
- *Made* vale Y quando il tiro è realizzato, N se è sbagliato.

La tabella seguente riassume le principali caratteristiche del dataset:

Dataset	Serie A2ITA
Competizione	Campionato – Stagione regolare
Periodo	04/10/2015 – 23/04/2016
Numero di partite	480
Numero di squadre	32
Numero di giocatori	438
Numero di tiri	69688 (50,6 % realizzati)
Numero di tiri da 2 punti	33682 (48,3 % - 50,9 % realizzati)
Numero di tiri da 3 punti	21163 (30,4 % - 34,1 % realizzati)
Numero di tiri liberi	14843 (21,3 % - 73,5 % realizzati)

Tabella 2.1 – Principali caratteristiche del dataset

2.1.1 Il web scraping: l'estrazione dei dati

I dati presenti nel sito sopra indicato sono stati estratti sfruttando una delle funzionalità del programma R: questa tecnica di estrazione dei dati da un sito web è definita “web scraping” e consente di trasformare dei dati non strutturati presenti in rete, di solito in formato HTML, in metadati che possono essere memorizzati e analizzati in locale in un database. Lo script di R che permette di effettuare tale operazione è il seguente:

```
www <- c("...", "...")
setwd(dir)
for (match in Mdownload) {
```

```

print(match)

webpage <- read_html(www[match])
webpage %>%
  html_nodes("table") %>%
  .[[1]] %>%
  html_table()

tab_web <- webpage %>%
  html_nodes("table") %>%
  .[[1]] %>%
  html_table()

nomefile <- paste("TabWWW",match,".Rdata",sep="")
save(tab_web,file=nomefile)

if(nrow(tab_web) == 0){
warn <- paste("Match ", match, "mancante")
print(warn)
}

```

dove *www* è il vettore contenente tutti gli indirizzi web dei play-by-play di tutte le partite di una singola squadra (che viene cambiata di volta in volta in modo da ottenere tutti i dati di tutte le squadre) e *Mdownload* è il vettore contenente i numeri da 1 a 30 (cioè il numero di partite giocate da ogni squadra).

Una volta scaricata, la sequenza play-by-play viene ordinata cronologicamente per avere gli avvenimenti iniziali nelle prime righe del dataset e, a seguire, i successivi fino ad arrivare al termine della partita. In seguito vengono aggiunte le variabili come il margine, il conteggio dei secondi, il cambio di possesso, le varie posizioni in classifica ecc. (vedasi paragrafo 2.1), che sono deducibili dalle informazioni di base contenute nel play-by-play.

2.1.2 Errori iniziali presenti nel dataset

Come anticipato all'inizio del capitolo, il play-by-play estratto dal web è caratterizzato dalla presenza di alcuni errori. Tali errori sono stati individuati analizzando ogni riga delle prime 25 partite di Verona: ciascun evento è stato osservato con attenzione per notare la presenza di qualsiasi genere di errore. Così, dopo aver esaminato più di 10000 righe, sono stati individuati, classificati e catalogati tutti gli errori presenti.

Di seguito viene riportato un elenco riassuntivo delle principali tipologie di defezioni presenti, classificati in base alla possibile causa (è importante tenere presente che alcuni possono essere corretti, mentre per altri ciò non è possibile):

A) Errori dovuti alla rilevazione e non correggibili

- Tempo costante: in alcuni casi, vi sono sequenze di azioni che sono state rilevate come se fossero accadute nello stesso istante. Si tratta di un probabile errore del rilevatore che ha rilevato le azioni con il cronometro fermo; pertanto, non è possibile ristabilire l'esatta distanza temporale tra le azioni.
- Sequenze di azioni impossibili: vi sono casi in cui si susseguono una serie di azioni che non possono essere accadute nella successione rilevata (ad esempio: tiro della squadra A – palla persa dalla squadra A – rimbalzo della squadra B. E' chiaro che nel mezzo ci sia qualche azione che non è stata rilevata). Anche qui si tratta di un probabile errore o dimenticanza del rilevatore, ma non è possibile ricostruire con sicurezza l'esatta sequenza di avvenimenti.
- Eccesso della durata delle azioni senza che si verifichino eventi che giustifichino un continuo reiterarsi delle azioni da 14 secondi: nel basket un'azione può durare al massimo 24 secondi e, in caso di fallo subito o rimbalzo in attacco, si ha diritto ad un ulteriore possesso della durata massima di 14 secondi (in questi casi i secondi eccessivi sono giustificati dalla variabile *secs14id* che assume valore 1; vedasi paragrafo 2.1). In alcune azioni, tuttavia, si nota un anomalo eccesso di durata delle azioni senza che si verifichino eventi che giustifichino un continuo reiterarsi delle azioni da 14 secondi: quando l'eccesso è nell'ordine di 1 – 5 secondi, è considerato un ritardo tollerabile perché dovuto al tempo materiale di rilevare sul pc la tipologia di azione e i giocatori coinvolti. In altri casi, l'eccesso di secondi è molto superiore: anche qui si può ipotizzare che vi siano delle azioni nel mezzo che non sono state rilevate o che vi sia stato un malfunzionamento temporaneo del programma. Comunque non è possibile correggere questo tipo di errore.
- Presenza di un giocatore in meno durante il gioco: in qualche caso, alcune delle azioni rilevate non presentano il giocatore protagonista di tale azione, ma compare la scritta "team". Può essere dovuto ad un errore nella codifica dei giocatori da parte del programma o ad un errore di rilevazione da parte del rilevatore, ma non si può stabilire con certezza quale giocatore abbia compiuto quella particolare azione. L'evento viene rilevato semplicemente come evento di squadra (nel caso del lavoro svolto, questo tipo particolare di errore non incide sui risultati finali).
- Errori sui cambi dei giocatori: in qualche caso, vi sono delle anomalie nella definizione di quale giocatore sia uscito dal campo e di chi l'abbia sostituito (ad esempio: giocatore A esce – dopo qualche azione si legge "giocatore A esce"). Si tratta di un probabile errore del programma di rilevazione, ma sarebbe correggibile semplicemente osservando quali

giocatori hanno effettuato le azioni seguenti; tuttavia, si è deciso di non correggere questo tipo di defezione per poter avere una procedura di estrazione e sistemazione dei dati che sia “standardizzata” (la correzione avrebbe, infatti, richiesto l’intervento manuale nella sistemazione dei dati).

B) Errori dovuti alla rilevazione e corretti

- Punteggio errato: in alcune partite si nota che il punteggio a inizio partita è errato, salvo poi “aggiustarsi” a metà partita. Questo tipo di errore è stato corretto tramite le nuove variabili di punteggio e margine ricalcolati (vedasi paragrafo 2.1).
- Rimbalzo a inizio del quarto: succede che, in qualche partita, il rimbalzo che avviene in occasione del tiro sulla sirena che sancisce la fine del quarto sia rilevato a tempo ormai scaduto e, quindi, venga rilevato come evento iniziale del quarto successivo. Dal momento che il rimbalzo che avviene dopo il suono della sirena è da considerarsi non valido, si è corretto questo tipo di errore semplicemente andando ad eliminare gli eventuali rimbalzi a inizio quarto.

C) Errori dovuti al sistema (non correggibili)

- Partite mancanti: il play-by-play di alcuni match è mancante, la relativa pagina web è vuota.
- Errori nella rilevazione dei tempi supplementari: in caso di match terminati ai tempi supplementari, gli eventi rilevati nel primo tempo supplementare risultano presenti nel primo quarto del match. Ne consegue un insieme di azioni che, secondo il programma, avvengono nel primo quarto ma che, in realtà, avvengono in parte nel primo quarto ed in parte nel primo over time. Risulta impossibile separare correttamente le azioni con una procedura standardizzata e, pertanto, i match terminati ai supplementari sono stati esclusi dall’analisi per questione di semplicità.

Di seguito viene riportato l’elenco completo dei match esclusi dall’analisi:

Giornata 1	OraSì Ravenna – Bondi Ferrara Bawer Matera – Remer Treviglio	Mancante Mancante
Giornata 2	Andrea Costa Imola – Eternedile Bologna	Supplementari
Giornata 3	Bawer Matera – De’ Longhi Treviso Npc Rieti – Assigeco Casalpusterlengo	Mancante Supplementari
Giornata 4	Remer Treviglio – Basket Recanati	Supplementari
Giornata 5	Acea Roma – Assigeco Casalpusterlengo	Supplementari
Giornata 6	FMC Ferentino – LightHouse Conad Trapani	Supplementari
Giornata 7	Moncada Agrigento – FMC Ferentino	Supplementari

	Orsi Tortona – La Briosa Barcellona	Supplementari
Giornata 12	Givova Scafati – Moncada Agrigento	Mancante
	Paffoni Omegna – BCC Agropoli	Supplementari
	Assigeco Casalpusterlengo – Orsi Tortona	Supplementari
Giornata 13	De' Longhi Treviso – Dinamica Generale Mantova	Supplementari
	OraSì Ravenna – Mec Energy Roseto	Supplementari
	Moncada Agrigento – Bermè Reggio Calabria	Mancante
Giornata 14	LightHouse Conad Trapani – Moncada Agrigento	Supplementari
Giornata 15	BCC Agropoli – Orsi Tortona	Mancante
Giornata 16	Tezenis Verona – Alma Agenzia per il Lavoro Trieste	Mancante
	Benacquista Ass. Latina – Mens Sana Basket 1871 Siena	Supplementari
	Paffoni Omegna – LightHouse Conad Trapani	Supplementari
Giornata 18	Novipiù Casale Monferrato – Mens Sana Basket 1871 Siena	Supplementari
	Moncada Agrigento – Acea Roma	Supplementari
Giornata 19	BCC Agropoli – Assigeco Casalpusterlengo	Mancante
Giornata 20	Alma Agenzia per il Lavoro Trieste – Basket Recanati	Supplementari
	Givova Scafati – Paffoni Omegna	Mancante
	Assigeco Casalpusterlengo – Acea Roma	Supplementari
Giornata 21	De' Longhi Treviso – Centrale del Latte Amica-Natura Brescia	Supplementari
	Proger Chieti – Dinamica Generale Mantova	Supplementari
	Bermè Reggio Calabria – Mens Sana Basket 1871 Siena	Supplementari
Giornata 22	Benacquista Ass. Latina – Givova Scafati	Mancante
Giornata 23	Bondi Ferrara – Alma Agenzia per il Lavoro Trieste	Mancante
Giornata 25	Remer Treviglio – Andrea Costa Imola	Supplementari
Giornata 27	Remer Treviglio – Alma Agenzia per il Lavoro Trieste	Supplementari
Giornata 28	Tezenis Verona – Proger Chieti	Mancante
	Benacquista Ass. Latina – Acea Roma	Mancante

Tabella 2.2 – Partite escluse dall'analisi

In seguito all'individuazione di queste tipologie di errori presenti, si è deciso di condurre l'analisi a livello di squadra e non a livello di giocatore: in questo modo, si dispone di un numero di tiri sufficienti a rendere robuste le stime delle probabilità di fare canestro e a permettere il confronto tra le varie squadre. Concentrandosi sul singolo giocatore, infatti, si rischia di avere un numero troppo esiguo di tiri e, di conseguenza, si dovrebbe limitare l'analisi ai giocatori che hanno effettuato almeno un determinato numero di tiri, come riportato nell'articolo (Zuccolotto, Manisera, Sandri, 2016) che verrà presentato a Siviglia in occasione della conferenza internazionale sulla statistica del 9 – 11 Dicembre 2016.

2.2 Analisi preliminari svolte per Scaligera Basket Verona

L'idea per il lavoro svolto è nata dall'esperienza di tirocinio curriculare svolta presso la società Scaligera Basket Verona (meglio nota come Tezenis Verona, dal nome del primo sponsor) sotto la supervisione del coach Marco Crespi, tecnico di grande esperienza nazionale ed internazionale, già vice allenatore di Luca Banchi a Siena nell'anno dell'ottavo scudetto e del successo in Supercoppa Italiana (stagione 2012 / 2013) e capo allenatore della Mens Sana Siena arrivata a gara 7 della finale scudetto nella stagione seguente.

Su richiesta del coach, al termine di ogni giornata di campionato, venivano effettuate una serie di analisi sulla squadra in modo da monitorare con continuità l'andamento del team in determinate situazioni e le performance dei giocatori.

Di seguito vengono riportate, in sintesi, le principali analisi svolte.

Le informazioni iniziali di interesse per il coach riguardavano l'andamento del margine (cioè del divario di punteggio rispetto agli avversari) in ogni partita giocata dalla squadra. Nel primo grafico riportato nella figura 2.2, ogni box plot rappresenta una partita ed il colore ne indica l'esito: verde coincide con una vittoria, rosso con una sconfitta. Dall'analisi della distribuzione del margine, si possono ottenere alcune indicazioni utili sull'andamento della squadra: generalmente, quando il margine è positivo per la maggior parte del tempo (cioè la squadra è in vantaggio per la maggior parte della partita), l'esito finale coincide con una vittoria. Tuttavia ci sono state partite in cui la squadra ha mantenuto il vantaggio per buona parte del tempo, ma alla fine ha perso il match: si osservino i match 4, 7, 15, 21 e 24. In un solo caso (match 30) è avvenuto il contrario: pur essendo stata in svantaggio per quasi tutta la partita, la squadra ha poi vinto il match. Il secondo grafico riporta, invece, l'andamento del margine all'interno di ogni singola partita; la fascia arancione evidenzia, come richiesto dal coach, la fascia di punteggio compresa tra -4 (cioè 4 punti di svantaggio) e $+4$ (cioè 4 punti di vantaggio).

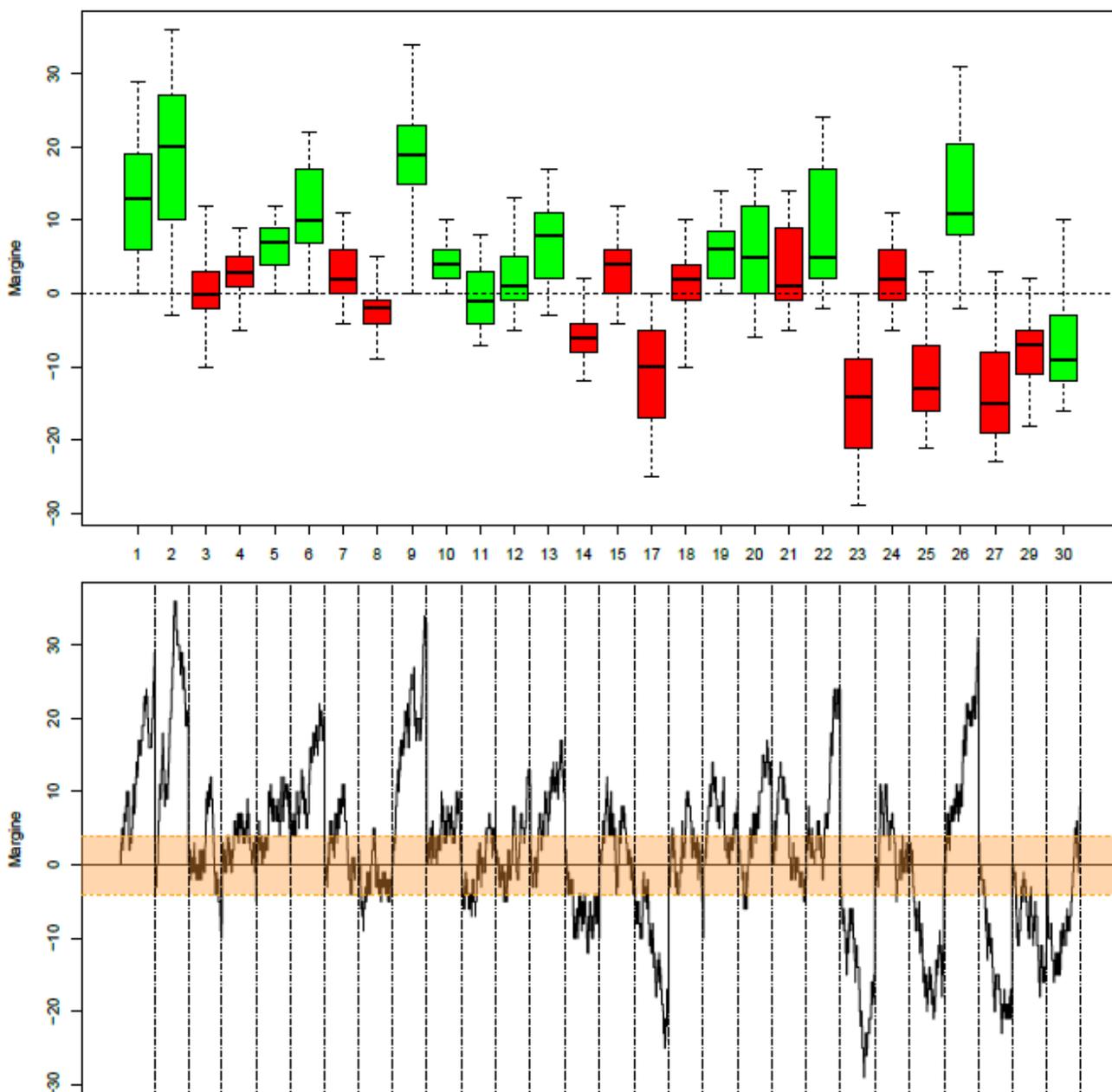


Figura 2.2 – Andamento del margine

Direttamente collegata alle informazioni presenti nella figura precedente, la figura 2.3 mostra il tempo di permanenza nelle fasce di margine. Il grafico a sinistra rappresenta il tempo (in secondi) e la fascia arancione evidenzia la forbice ± 4 . In generale, la squadra ha trascorso il 63% del tempo giocato trovandosi in situazione di parità o in vantaggio, mentre è stata in svantaggio per il 37% del tempo giocato; in particolare, l'andamento della linea blu suggerisce che, per la maggior parte del tempo giocato, sia stata all'interno della forbice o, comunque, a ridosso di essa. A conferma di quanto affermato, il grafico a barre riportato sulla destra stabilisce la percentuale di tempo trascorsa sotto (21,98%), sopra (38,18%) e all'interno della forbice (39,84%).

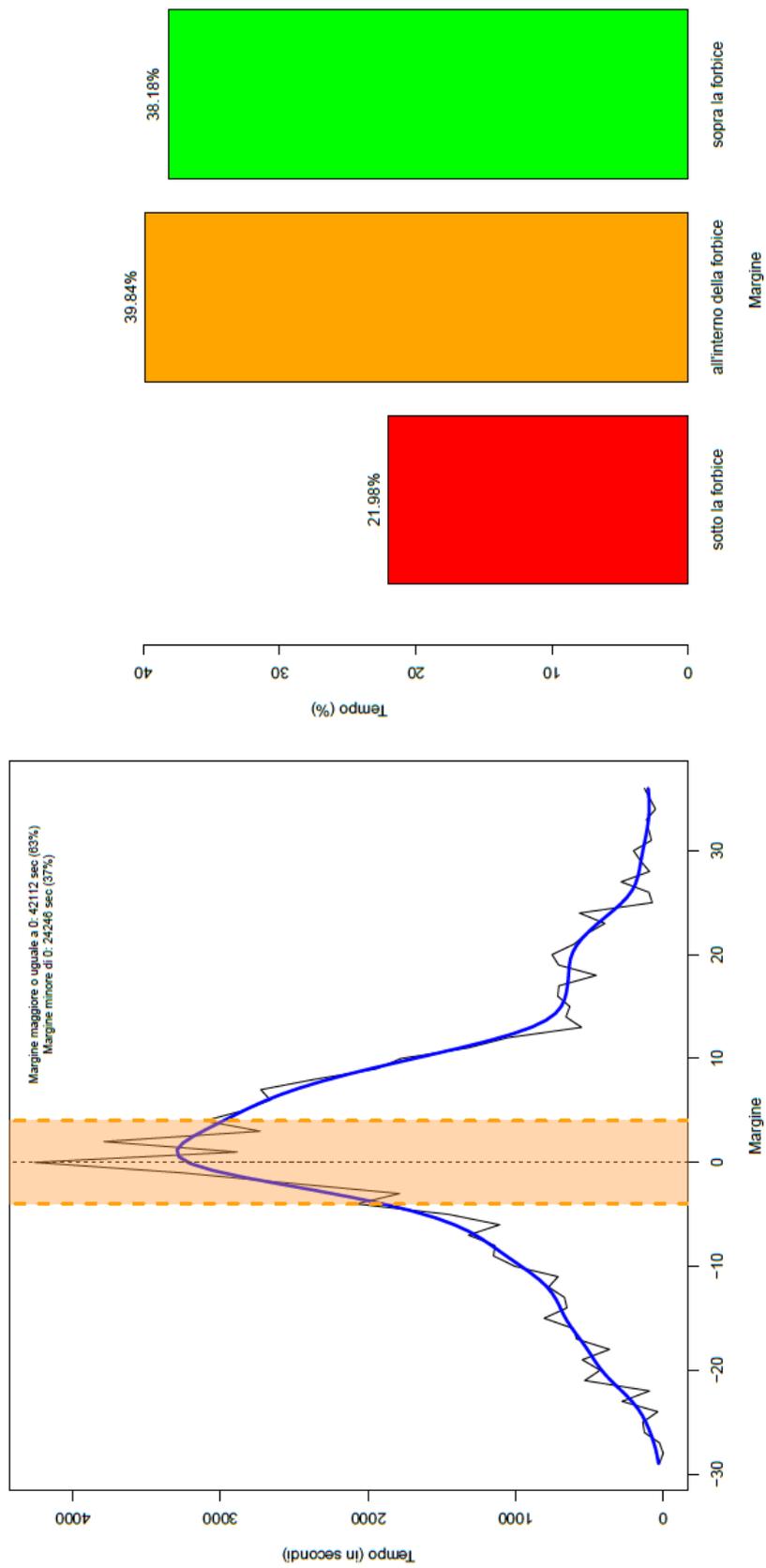


Figura 2.3 – Tempo di permanenza nelle fasce di margine

Analogamente, si osserva anche l'andamento del margine all'interno dei singoli quarti di gioco della partita.

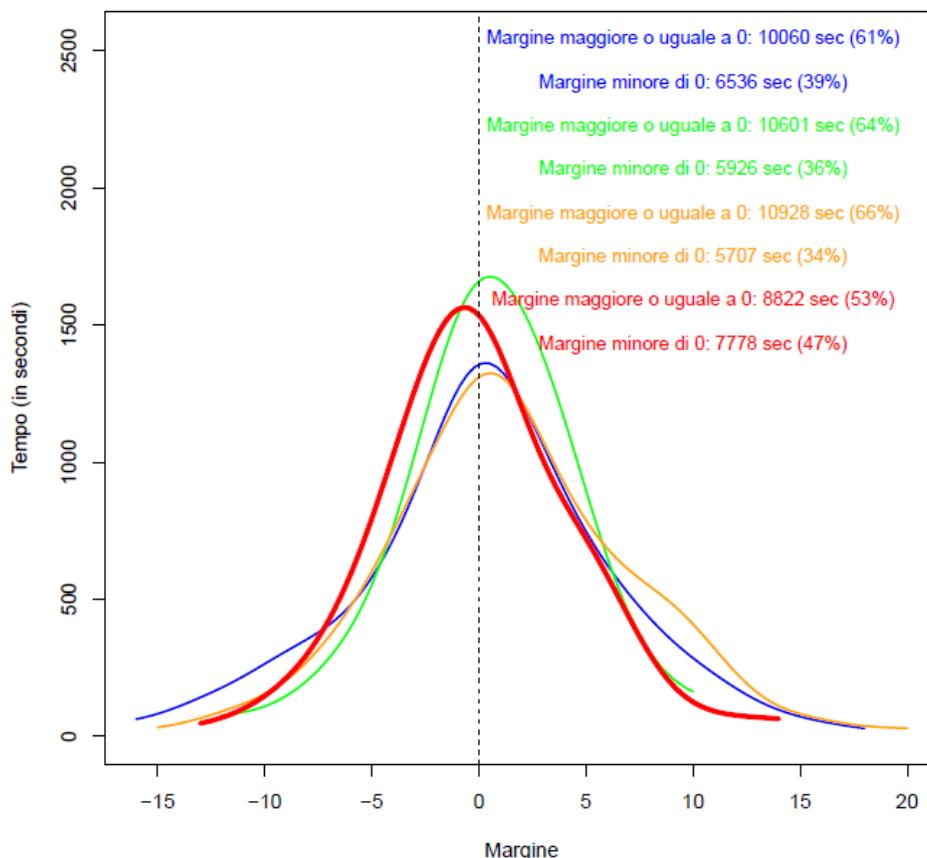


Figura 2.4 – Andamento del margine nei quarti

Il colore blu rappresenta il primo periodo, il verde il secondo periodo, l'arancio il terzo periodo e il rosso il quarto periodo. Come si può facilmente osservare, la squadra tende a mostrare un peggioramento nell'ultimo quarto di gioco: il tempo passato in parità o in vantaggio passa da oltre 60% delle prime tre frazioni di gioco al 53% nell'ultima frazione, come evidenziato anche dal posizionamento più a sinistra della curva rossa rispetto alle altre. Questo è un primo indicatore di un calo che la squadra mostra nel finale di partita e che verrà ulteriormente approfondito più avanti.

In generale, la figura 2.4 rappresenta uno strumento molto utile ai coach per evidenziare eventuali cali di concentrazione / prestazione della propria squadra e per individuare in quale momento della partita essi si manifestano con più evidenza.

Entrando nel dettaglio delle singole azioni di gioco, l'interesse di coach Crespi era orientato verso i seguenti quesiti: dividendo un'azione da 24 secondi in tre fasce da 8 secondi ciascuna, come si distribuiscono i vari eventi all'interno delle fasce? Come si comporta la squadra in ciascuna fascia? E nelle successive azioni da 14 secondi? I seguenti grafici riportati nelle figure 2.5 e 2.6 consentono di dare informazioni in merito a tali questioni.

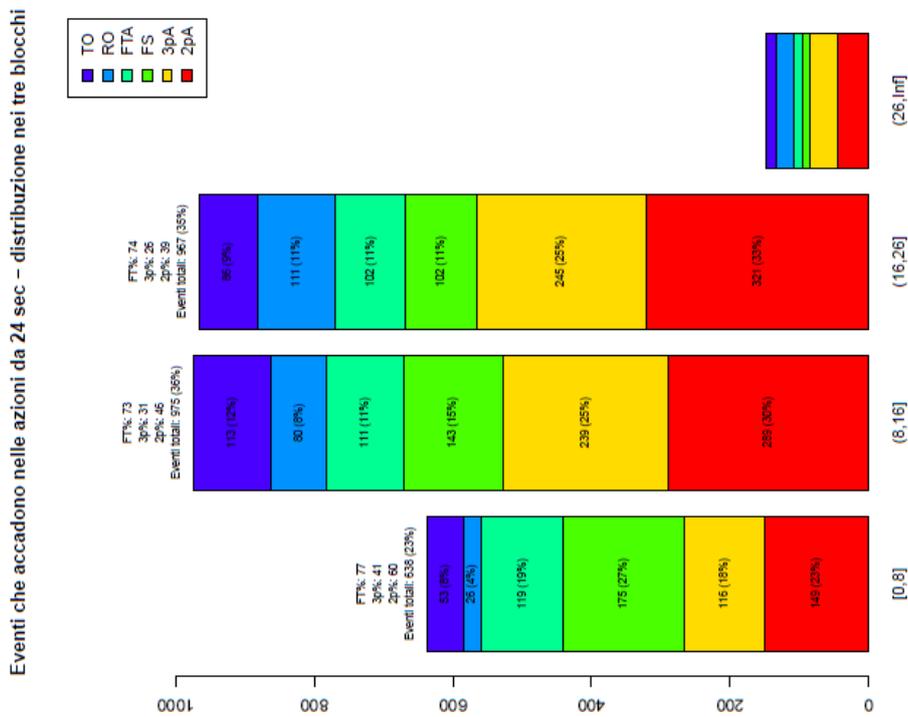
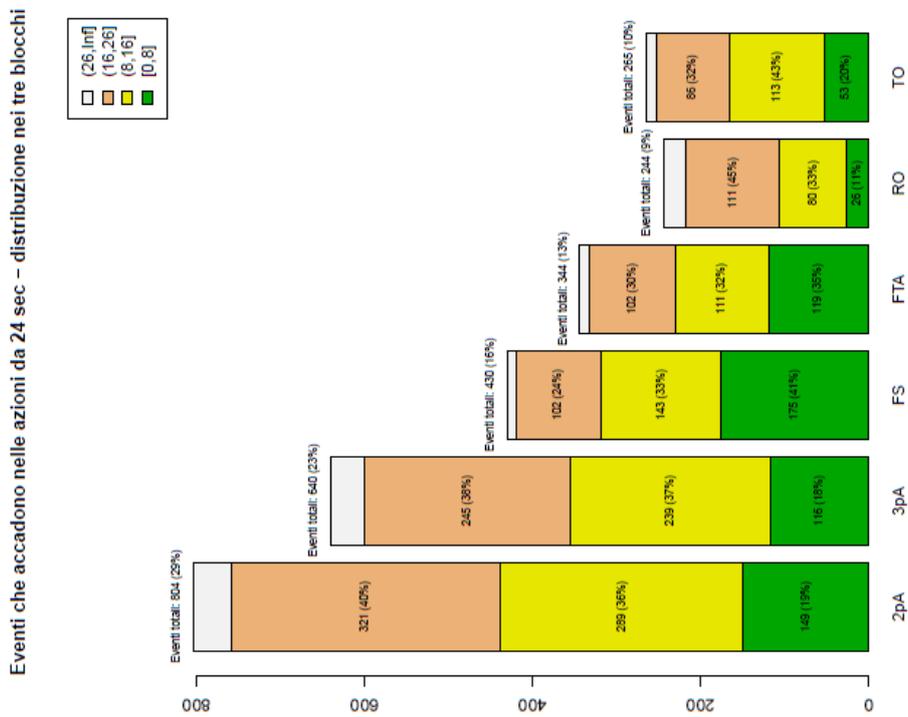


Figura 2.5 – Eventi interni alle azioni da 24 secondi

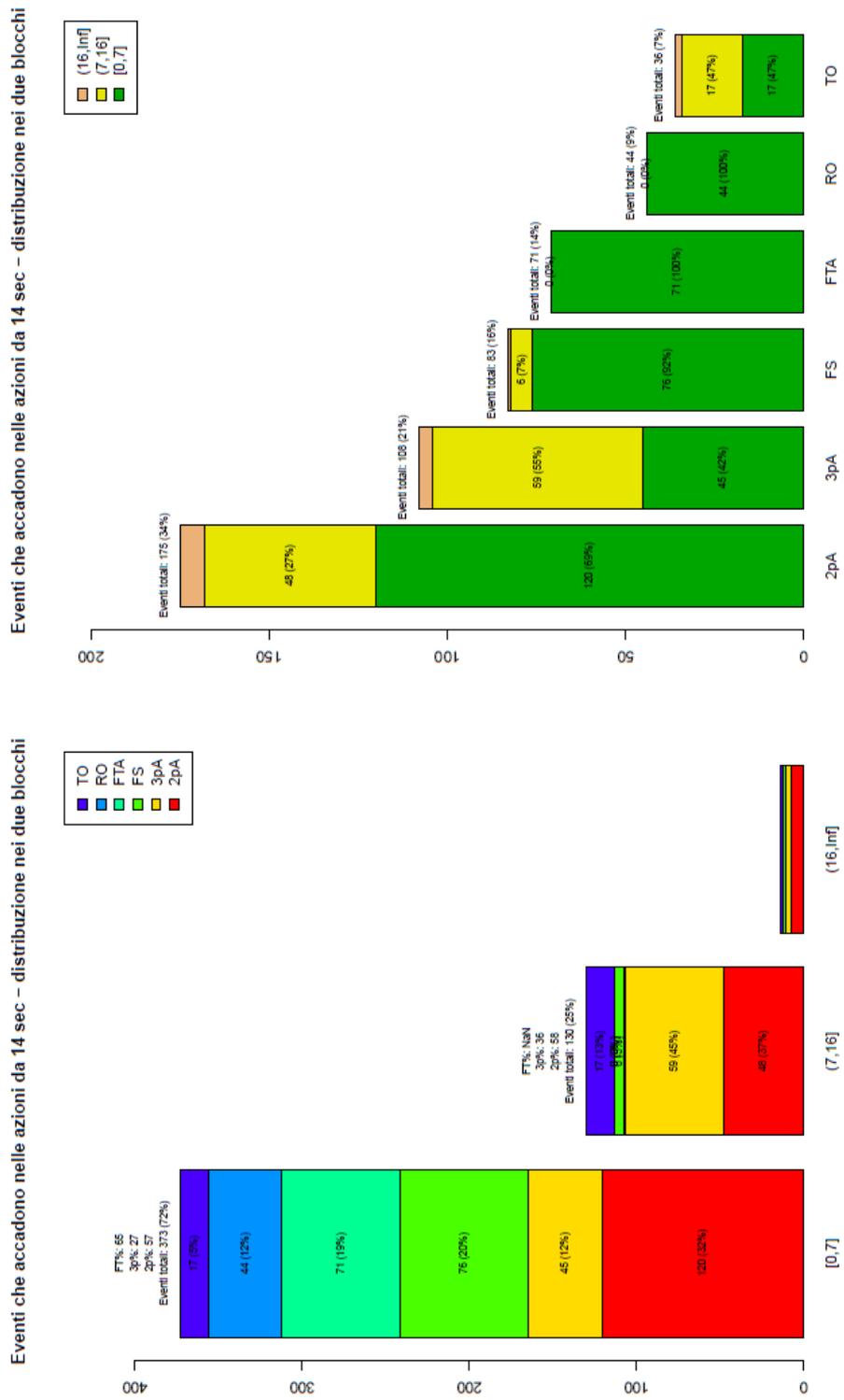


Figura 2.6 – Eventi interni alle azioni da 14 secondi

Sia i grafici a destra che quelli a sinistra raffigurano, in maniera leggermente differente, la distribuzione degli eventi nelle rispettive fasce di secondi (si considerano 2 secondi e una fascia in più per compensare gli errori del dataset; vedasi paragrafo 2.1.2). Sopra ogni barra viene indicato il numero di eventi e, per la parte sinistra, le percentuali di realizzazione al tiro.

Per quanto riguarda le azioni da 24 secondi, si può dire che, nonostante il maggior numero di eventi accada nella seconda e terza fascia (rispettivamente 975 e 967 eventi, rispetto ai 638 della prima fascia), la distribuzione degli eventi all'interno di ciascuna fascia è pressoché simile: i tiri da 2 punti costituiscono il 23% degli eventi della prima fascia, il 30% di quelli in seconda e il 33% di quelli in terza, mentre i tiri da 3 punti costituiscono il 18% delle azioni nella prima fascia, il 25% di quelle in seconda e terza. Da sottolineare come le percentuali al tiro vadano diminuendo con l'aumentare dei secondi: nel tiro da 2 punti si passa dal 60% in prima fascia al 46% in seconda e al 39% in terza, mentre nel tiro da 3 si passa dal 41% in prima al 31% in seconda e al 26% in terza. Va tenuto conto, tuttavia, che buona parte dei tiri effettuati nei primi secondi consistono in tiri in contropiede (quindi con totale assenza, o quasi, della difesa), il che può inevitabilmente portare ad un aumento di percentuale di realizzazione. Dal grafico di destra si evince come i tiri da 2 e da 3 punti siano gli eventi che accadono più spesso, rispettivamente il 29% e il 23% degli eventi totali. Altro dato importante è quello dei falli subiti: nelle azioni della prima fascia, il 27% è rappresentato da questo evento, che tende a diminuire con il passare dei secondi (nella seconda fascia è il 15% delle azioni, per scendere fino al 11% nella seconda). Significativo, inoltre, il dato riguardante i rimbalzi offensivi che emerge dal grafico di destra: solo il 9% degli eventi totali, a sottolineare la scarsa propensione della squadra a catturare rimbalzi in attacco e ottenere, così, seconde opportunità di tiro. La maggior parte di questi rimbalzi (il 45%), comunque, avviene nella terza fascia.

Per quanto riguarda le azioni da 14 secondi, la maggior parte degli eventi si verifica nella prima metà dell'azione (72% contro il 25% della seconda metà). Anche qui, si verificano maggiormente un tiro da 2 punti o un fallo subito, rispettivamente il 32% e il 20% degli eventi della prima fascia. Le percentuali di realizzazione al tiro, tuttavia, sembrano essere più elevate nella seconda metà dell'azione (probabilmente quando la palla circola meglio e la squadra riesce a servire il giocatore che si trova libero in seguito alle rotazioni e agli aiuti difensivi): 58% da 2 punti e 36% da 3 punti, mentre nella prima fascia si osserva un 57% da 2 punti e 27% da 3 punti. Rispecchiata, inoltre, la scarsa propensione al rimbalzo in attacco: il 9% degli eventi totali, tra l'altro tutti avvenuti nei primi 7 secondi. Dal grafico di destra si evince come, per quanto riguarda i tiri da 2 punti (il 34% degli eventi totali), la squadra tende a tirare nella prima fascia (69% dei tiri da 2 avvengono nei primi 7

secondi); per i tiri da 3 punti (il 21% degli eventi totali), invece, la squadra tende a tirare nella seconda fascia (55% dei tiri da 3 avvengono negli ultimi 7 secondi).

Entrando nello specifico, si isolano e analizzano i tiri da 2 e da 3 punti nelle singole fasce di secondi, ottenendo le figure 2.7, 2.8, 2.9 e 2.10.

Come si può vedere in ogni grafico, in ciascuna fascia di secondi sono riportate le informazioni di maggiore interesse: la percentuale di tiri da 2 e 3 punti rispetto alle azioni totali, rispetto ai tiri totali da 2 e 3 punti e le relative percentuali di realizzazione. Ad esempio, con riferimento al primo grafico, si può osservare che l'8% delle azioni totali consiste in un tiro da 2 punti nel primo blocco delle azioni da 24 secondi, che il 19% dei tiri da 2 punti è avvenuto nel primo blocco e che la percentuale di realizzazione è del 60%. Analoga chiave di lettura per gli altri blocchi di secondi e per gli altri grafici. Interessante notare come, per quanto riguarda i tiri da 2 punti nei 24 secondi, si possa intravedere nelle percentuali al tiro, "l'effetto contropiede" citato in precedenza: la percentuale di realizzazione è del 60%, mentre scende al 49% nella seconda fascia e al 39% nella terza fascia. Tale effetto è visibile anche nei tiri da 3 punti, nonostante nella prima fascia la squadra tiri leggermente di meno (18% dei tiri da 3 sono avvenuti nella prima fascia, contro il 19% dei tiri da 2): la percentuale di realizzazione è del 41% nella prima fascia e scende progressivamente fino al 31% e poi al 29%. Per quanto riguarda i tiri da 2 punti nei 14 secondi, particolarmente significative risultano le percentuali di realizzazione in entrambe le fasce (57% e 58%), di poco inferiori al primo blocco dei 24 secondi ma superiori alle altre due fasce, a sottolineare l'importanza di seconde occasioni in attacco dopo un fallo subito o dopo un rimbalzo offensivo catturato. Per quanto riguarda i tiri da 3 punti nei 14 secondi, risultano più efficaci nella seconda fascia: percentuale di realizzazione pari a 36% (contro il 27% della prima fascia).

Questi grafici risultano importanti per permettere al coach di individuare eventuali momenti durante la singola azione in cui si verificano cali di precisione o per analizzare la tendenza della squadra, se tende a "correre" (cioè effettuare contropiede oppure effettuare tiri affrettati) o tende ad effettuare azioni più lunghe. Queste informazioni diventano ancora più utili se tessute sui singoli giocatori perché permettono al coach di individuare in quale particolare momento dell'azione ogni giocatore è più o meno pericoloso e, di conseguenza, consentono di lavorare sulle scelte delle strategie difensive.

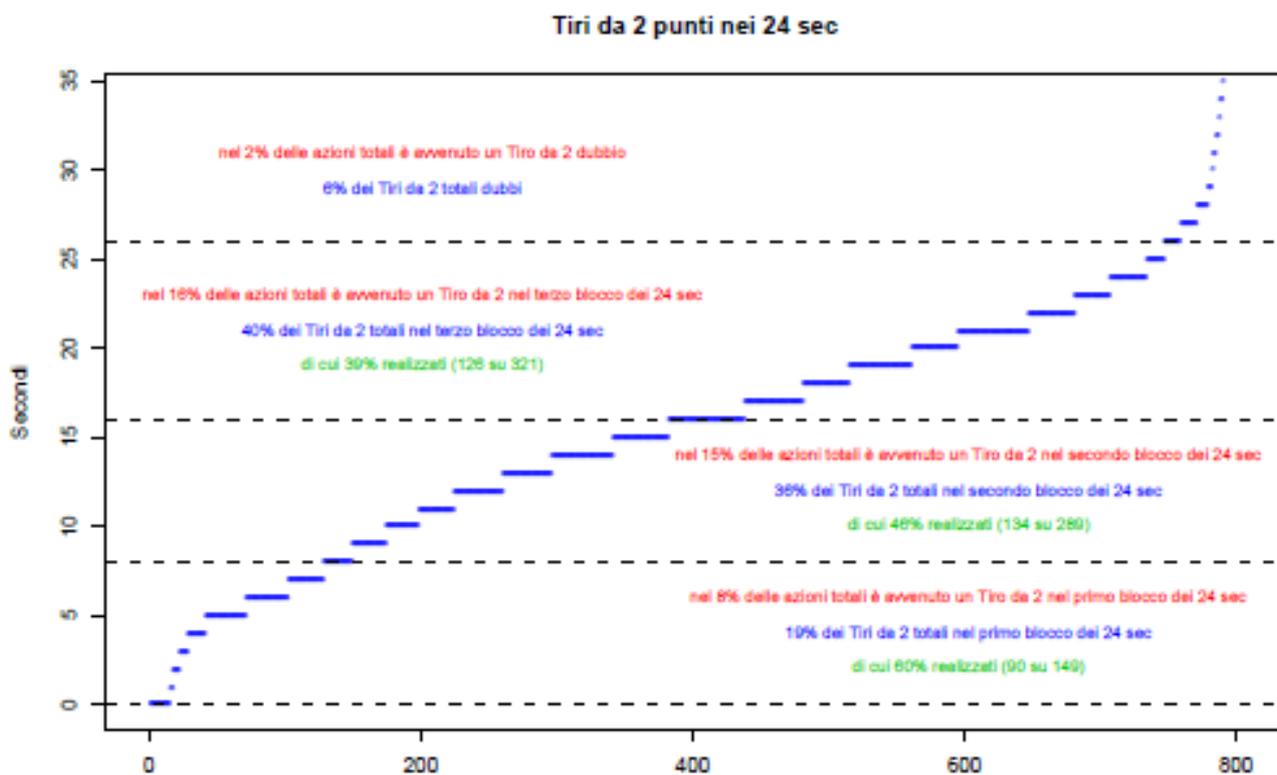


Figura 2.7 – Tiri da 2 punti nei 24 secondi

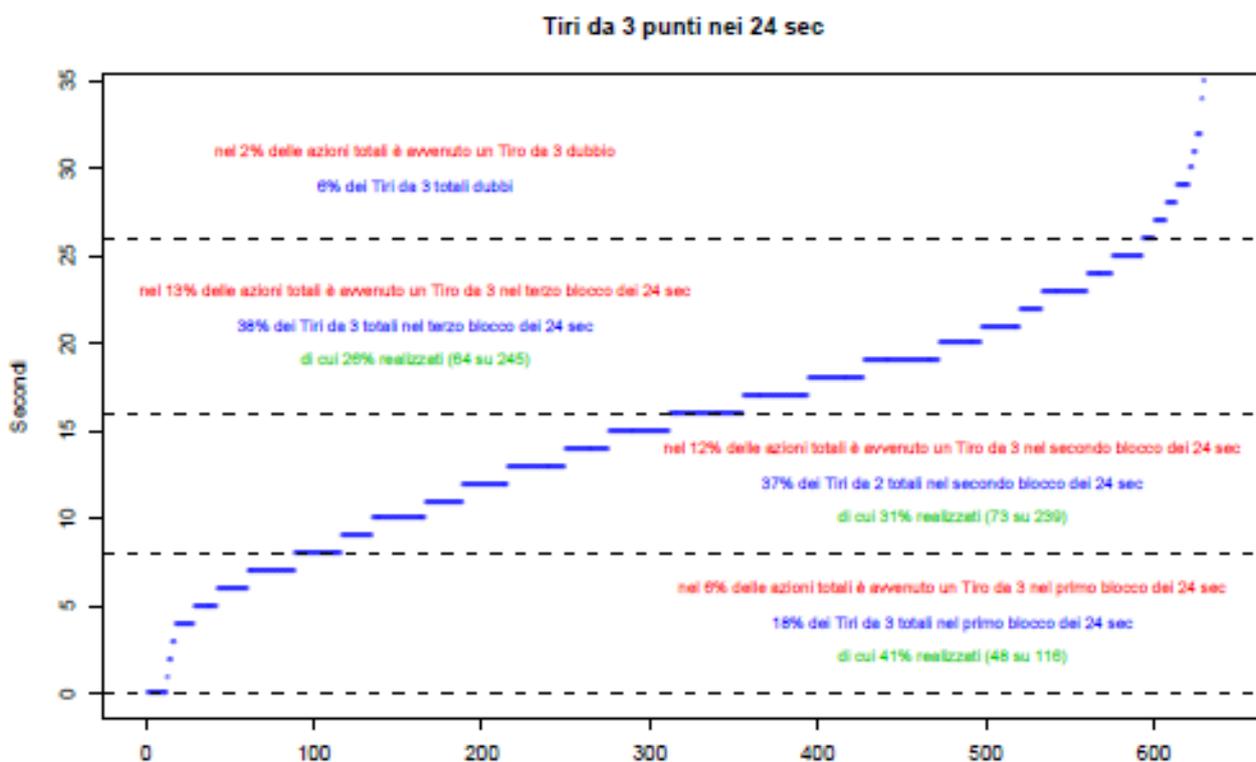


Figura 2.8 – Tiri da 3 punti nei 24 secondi

Tiri da 2 punti nei 14 sec

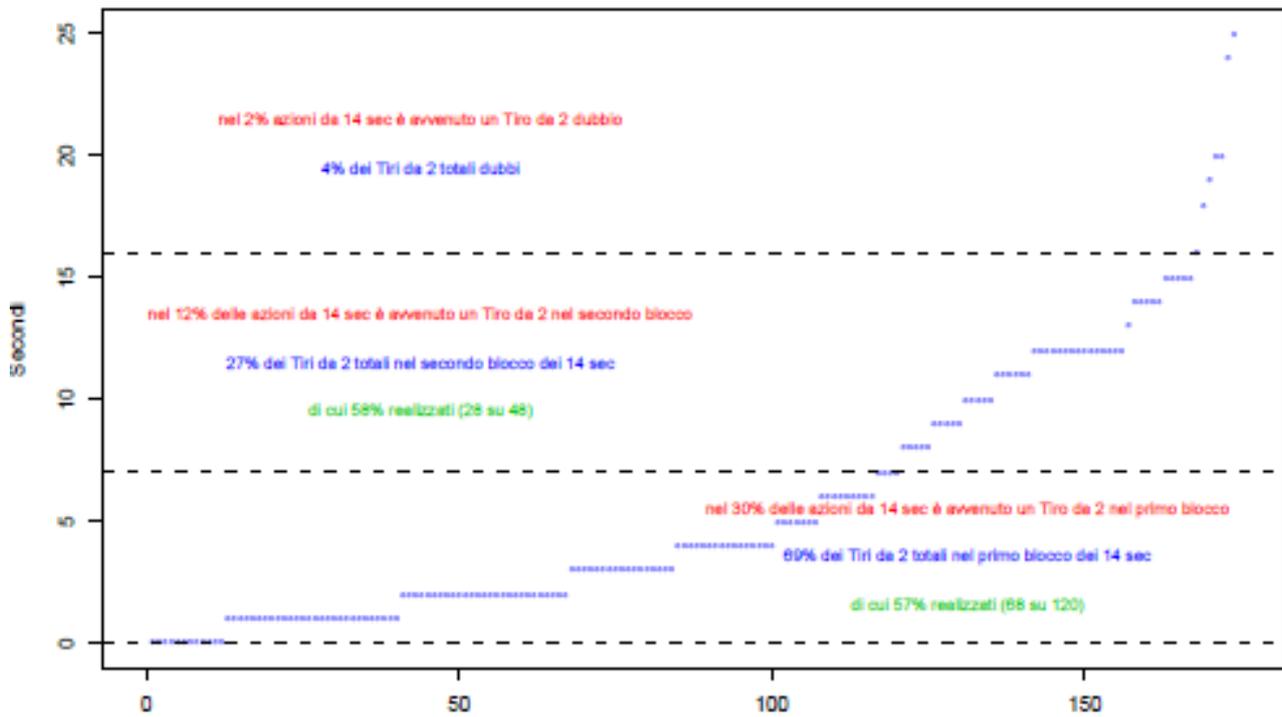


Figura 2.9 – Tiri da 2 punti nei 14 secondi

Tiri da 3 punti nei 14 sec

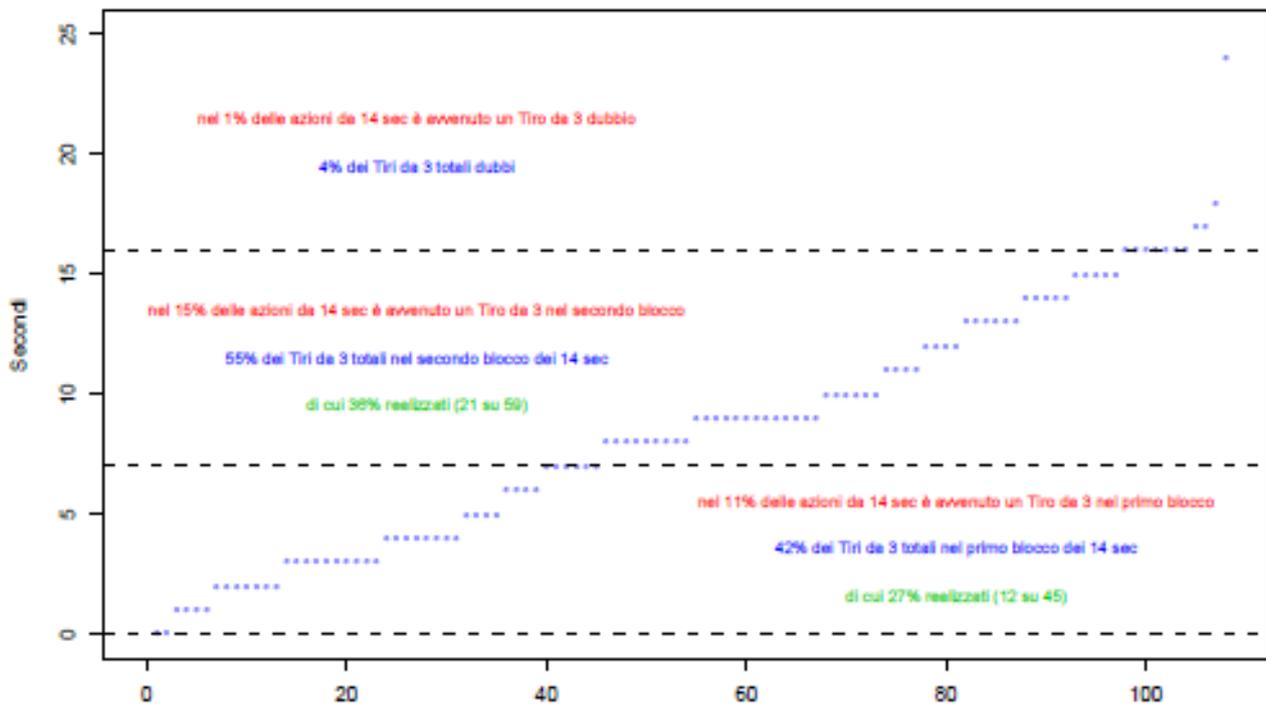


Figura 2.10 – Tiri da 3 punti nei 14 secondi

Infine, per avere un riassunto generale della performance di squadra, si forniva al coach una serie di report in cui si illustravano tutte le statistiche base di gioco (tiri da 2 punti, tiri da 3 punti, tiri liberi, rimbalzi offensivi e difensivi, palle perse ecc.) rilevate in base alla forbice ± 4 punti individuata precedentemente e rapportate ai minuti trascorsi in ciascuna fascia. Per la squadra in generale, vi erano 2 report: uno calcolato sull'intera partita e uno focalizzato sull'ultimo quarto di ogni partita. Stessa cosa per ciascun giocatore della squadra. In totale, coach Crespi disponeva di 26 report riguardanti la propria squadra costruiti come gli esempi riportati nelle figure 2.11, 2.12 (per l'intera squadra), 2.13 e 2.14 (per il singolo giocatore; l'esempio in questo caso fa riferimento a Lorenzo Saccaggi).

Margine – IP	Min	2pM	2pA	2p%	2pM40	3pM	3pA	3p%	3pM40	FTM	FTA	FT%	FT40	Pts40	PtsOpp40
A: sotto la forbice	243	97	220	44.09	15.96	58	188	30.85	9.54	89	117	76.07	14.64	75.19	69.27
B: interno alla forbice	441	203	423	47.99	18.43	95	307	30.94	8.62	93	132	70.45	8.44	71.18	61.28
C: sopra la forbice	422	165	336	49.11	15.63	80	253	31.62	7.58	118	166	71.08	11.18	65.17	68.39

Margine – IP	Min	FS	FS40	RO	RO40	TO	TO40
A: sotto la forbice	243	123	20.24	70	11.52	62	10.20
B: interno alla forbice	441	181	16.43	117	10.62	99	8.99
C: sopra la forbice	422	194	18.38	101	9.57	140	13.26

Legenda prima tabella	Margine – IP
Margine nelle tre categorie – Intera Partita	Margine – IP
Tempo di permanenza in minuti nelle tre categorie di margine	Min
Tiri da 2 riusciti	2pM
Tiri da 2 totali	2pA
Tiri da 2 % successo	2p%
Tiri da 2 rapportati al tempo (40 minuti)	2pM40
Tiri da 3 riusciti	3pM
Tiri da 3 totali	3pA
Tiri da 3 % successo	3p%
Tiri da 3 rapportati al tempo (40 minuti)	3pM40
Tiri liberi riusciti	FTM
Tiri liberi totali	FTA
Tiri liberi % successo	FT%
Tiri liberi rapportati al tempo (40 minuti)	FT40
Punti fatti rapportati al tempo (40 minuti)	Pts40
Punti subiti rapportati al tempo (40 minuti)	PtsOpp40

Legenda seconda tabella	Margine – IP
Margine nelle tre categorie – Intera Partita	Margine – IP
Tempo di permanenza in minuti nelle tre categorie di margine	Min
Falli subiti	FS
Falli subiti rapportati al tempo (40 minuti)	FS40
Rimbalzi offensivi	RO
Rimbalzi offensivi rapportati al tempo (40 minuti)	RO40
Palle perse	TO
Palle perse rapportate al tempo (40 minuti)	TO40

FORBICE: tra-4 e 4

Figura 2.11 – Report di Verona nelle fasce di margine (intera partita)

Margine - UQ	Min	2pM	2pA	2p%	2pM40	3pM	3pA	3p%	3pM40	FTM	FTA	FT%	FT40	Pts40	PtsOpp40
A: sotto la forbice	65	28	63	44.44	17.14	17	50	34.00	10.41	32	40	80.0	19.59	85.08	69.17
B: interno alla forbice	92	32	75	42.67	13.94	14	58	24.14	6.10	28	40	70.0	12.20	58.37	62.72
C: sopra la forbice	119	47	92	51.09	15.73	18	64	28.12	6.03	47	63	74.6	15.73	65.28	65.62

Margine - UQ	Min	FS	FS40	RO	RO40	TO	TO40
A: sotto la forbice	65	30	18.36	21	12.85	14	8.57
B: interno alla forbice	92	51	22.21	23	10.02	28	12.20
C: sopra la forbice	119	70	23.43	29	9.71	49	16.40

Legenda prima tabella	Margine - UQ
Margine nelle tre categorie - Ultimo Quarto	Margine - UQ
Tempo di permanenza in minuti nelle tre categorie di margine	Min
Tiri da 2 riusciti	2pM
Tiri da 2 totali	2pA
Tiri da 2 % successo	2p%
Tiri da 2 rapportati al tempo (40 minuti)	2pM40
Tiri da 3 riusciti	3pM
Tiri da 3 totali	3pA
Tiri da 3 % successo	3p%
Tiri da 3 rapportati al tempo (40 minuti)	3pM40
Tiri liberi riusciti	FTM
Tiri liberi totali	FTA
Tiri liberi % successo	FT%
Tiri liberi rapportati al tempo (40 minuti)	FT40
Punti fatti rapportati al tempo (40 minuti)	Pts40
Punti subiti rapportati al tempo (40 minuti)	PtsOpp40

Legenda seconda tabella	Margine - UQ
Margine nelle tre categorie - Ultimo Quarto	Margine - UQ
Tempo di permanenza in minuti nelle tre categorie di margine	Min
Falli subiti	FS
Falli subiti rapportati al tempo (40 minuti)	FS40
Rimbalzi offensivi	RO
Rimbalzi offensivi rapportati al tempo (40 minuti)	RO40
Palle perse	TO
Palle perse rapportate al tempo (40 minuti)	TO40

FORBICE: tra-4 e 4

Figura 2.12 – Report di Verona nelle fasce di margine (ultimo quarto)

Margine - IP	Min	2pM	2pA	2p%	2pM40	3pM	3pA	3p%	3pM40	FTM	FTA	FT%	FT40	Pts40
A: sotto la forbice	143	14	31	45.16	3.92	6	21	28.57	1.68	13	16	81.25	3.64	16.54
B: interno alla forbice	309	14	36	38.89	1.81	13	41	31.71	1.68	5	7	71.43	0.65	9.31
C: sopra la forbice	273	14	28	50.00	2.05	6	29	20.69	0.88	6	11	54.55	0.88	7.63

Margine - IP	Min	FS	FS40	RO	RO40	TO	TO40
A: sotto la forbice	143	15	4.20	3	0.84	11	3.08
B: interno alla forbice	309	29	3.75	2	0.26	14	1.81
C: sopra la forbice	273	19	2.79	3	0.44	10	1.47

Legenda prima tabella	Margine - IP
Margine nelle tre categorie - Intera Partita	Margine - IP
Tempo di permanenza in minuti nelle tre categorie di margine	Min
Tiri da 2 riusciti	2pM
Tiri da 2 totali	2pA
Tiri da 2 % successo	2p%
Tiri da 2 rapportati al tempo (40 minuti)	2pM40
Tiri da 3 riusciti	3pM
Tiri da 3 totali	3pA
Tiri da 3 % successo	3p%
Tiri da 3 rapportati al tempo (40 minuti)	3pM40
Tiri liberi riusciti	FTM
Tiri liberi totali	FTA
Tiri liberi % successo	FT%
Tiri liberi rapportati al tempo (40 minuti)	FT40
Punti fatti rapportati al tempo (40 minuti)	Pts40

Legenda seconda tabella	Margine - IP
Margine nelle tre categorie - Intera Partita	Margine - IP
Tempo di permanenza in minuti nelle tre categorie di margine	Min
Falli subiti	FS
Falli subiti rapportati al tempo (40 minuti)	FS40
Rimbalzi offensivi	RO
Rimbalzi offensivi rapportati al tempo (40 minuti)	RO40
Palle perse	TO
Palle perse rapportate al tempo (40 minuti)	TO40

FORBICE: tra-4 e 4

Minuti giocatore analizzati: 725

Figura 2.13 – Report di Lorenzo Saccaggi nelle fasce di margine (intera partita)

Margine - UQ	Min	2pM	2pA	2p%	2pM40	3pM	3pA	3p%	3pM40	FTM	FTA	FT%	FT40	Pts40
A: sotto la forbice	34	3	11	27.27	3.51	3	7	42.86	3.51	8	8	100	9.37	26.93
B: interno alla forbice	64	4	6	66.67	2.52	5	12	41.67	3.15	1	1	100	0.63	15.11
C: sopra la forbice	61	7	10	70.00	4.63	2	10	20.00	1.32	3	4	75	1.98	15.20

Margine - UQ	Min	FS	FS40	RO	RO40	TO	TO40
A: sotto la forbice	34	4	4.68	1	1.17	2	2.34
B: interno alla forbice	64	7	4.41	0	0.00	2	1.26
C: sopra la forbice	61	7	4.63	2	1.32	4	2.64

Legenda prima tabella	Margine - UQ	Legenda seconda tabella	Margine - UQ
Margine nelle tre categorie - Ultimo Quarto	Margine - UQ	Margine nelle tre categorie - Ultimo Quarto	Margine - UQ
Tempo di permanenza in minuti nelle tre categorie di margine	Min	Tempo di permanenza in minuti nelle tre categorie di margine	Min
Tiri da 2 riusciti	2pM	Falli subiti	FS
Tiri da 2 totali	2pA	Falli subiti riportati al tempo (40 minuti)	FS40
Tiri da 2 % successo	2p%	Rimbalzi offensivi	RO
Tiri da 2 riportati al tempo (40 minuti)	2pM40	Rimbalzi offensivi riportati al tempo (40 minuti)	RO40
Tiri da 3 riusciti	3pM	Palle perse	TO
Tiri da 3 totali	3pA	Palle perse riportate al tempo (40 minuti)	TO40
Tiri da 3 % successo	3p%		
Tiri da 3 riportati al tempo (40 minuti)	3pM40		
Tiri liberi riusciti	FTM		
Tiri liberi totali	FTA		
Tiri liberi % successo	FT%		
Tiri liberi riportati al tempo (40 minuti)	FT40		
Punti fatti riportati al tempo (40 minuti)	Pts40		

FORBICE: tra-4 e 4

Minuti giocatore analizzati: 158

Figura 2.14 – Report di Lorenzo Saccaggi nelle fasce di margine (ultimo quarto)

In conclusione, oltre a monitorare giornata per giornata l'andamento della squadra e dei singoli giocatori tramite i grafici e le tabelle appena esposti, tutte le analisi presentate venivano condotte anche sull'avversario della giornata successiva (squadra in generale e singoli giocatori nello specifico) in modo da tracciare un profilo generale della squadra, individuare i punti di forza del team avversario, individuare i giocatori più pericolosi ed elaborare, così, la migliore strategia di gioco possibile. Nelle figure dalla 2.15 alla 2.27 vengono riportati, a titolo esemplificativo, i grafici elaborati per uno degli avversari di Verona, ossia Centrale del Latte – Amica Natura Brescia (che terminò la regular season al secondo posto nel girone est e poi vinse i playoff, ottenendo la promozione in serie A). I report sul singolo giocatore (figure 2.26 e 2.27) riguardano Mirza Alibegovic, uno dei giocatori della formazione bresciana.

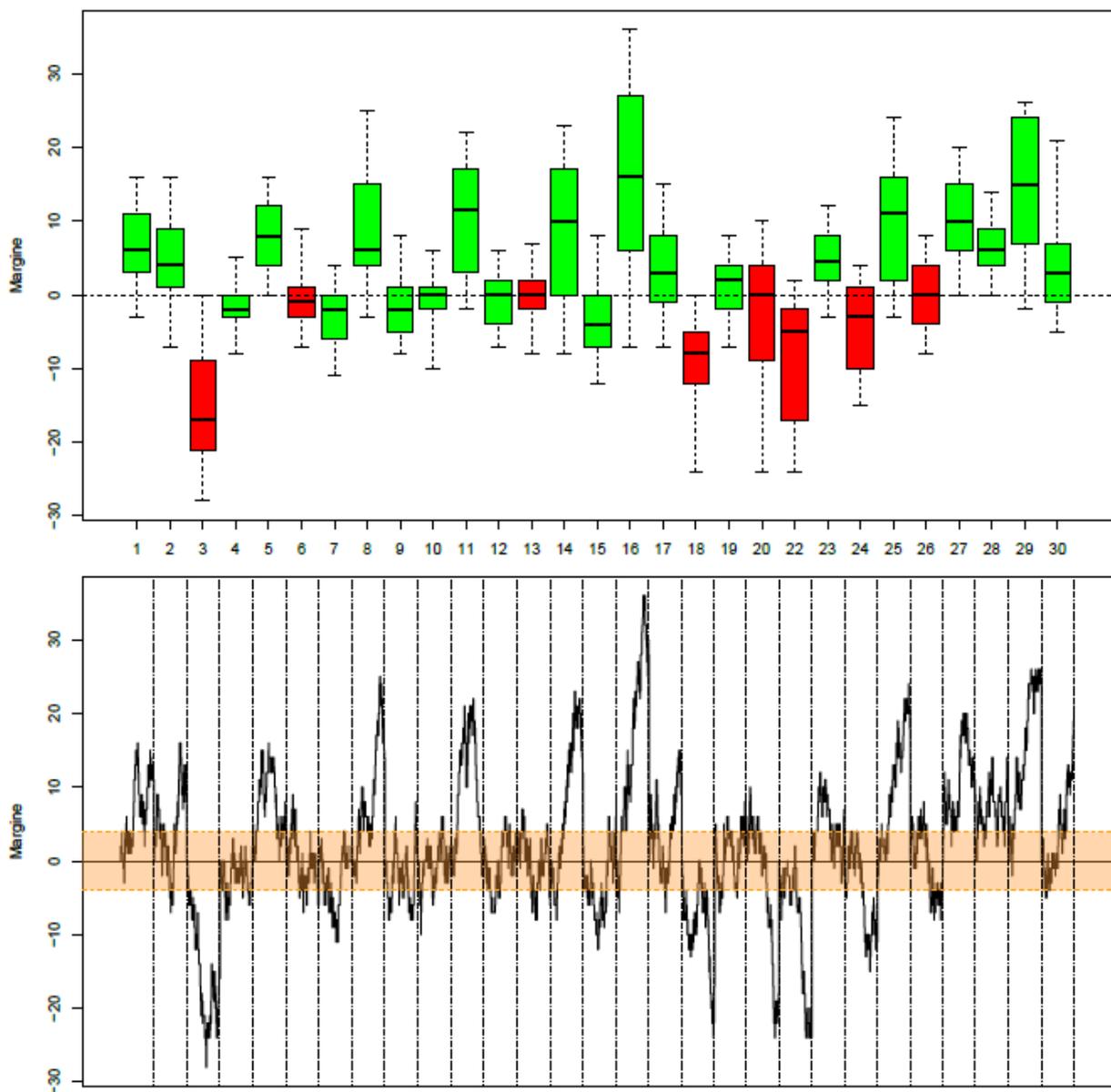


Figura 2.15 – Andamento del margine (Brescia)

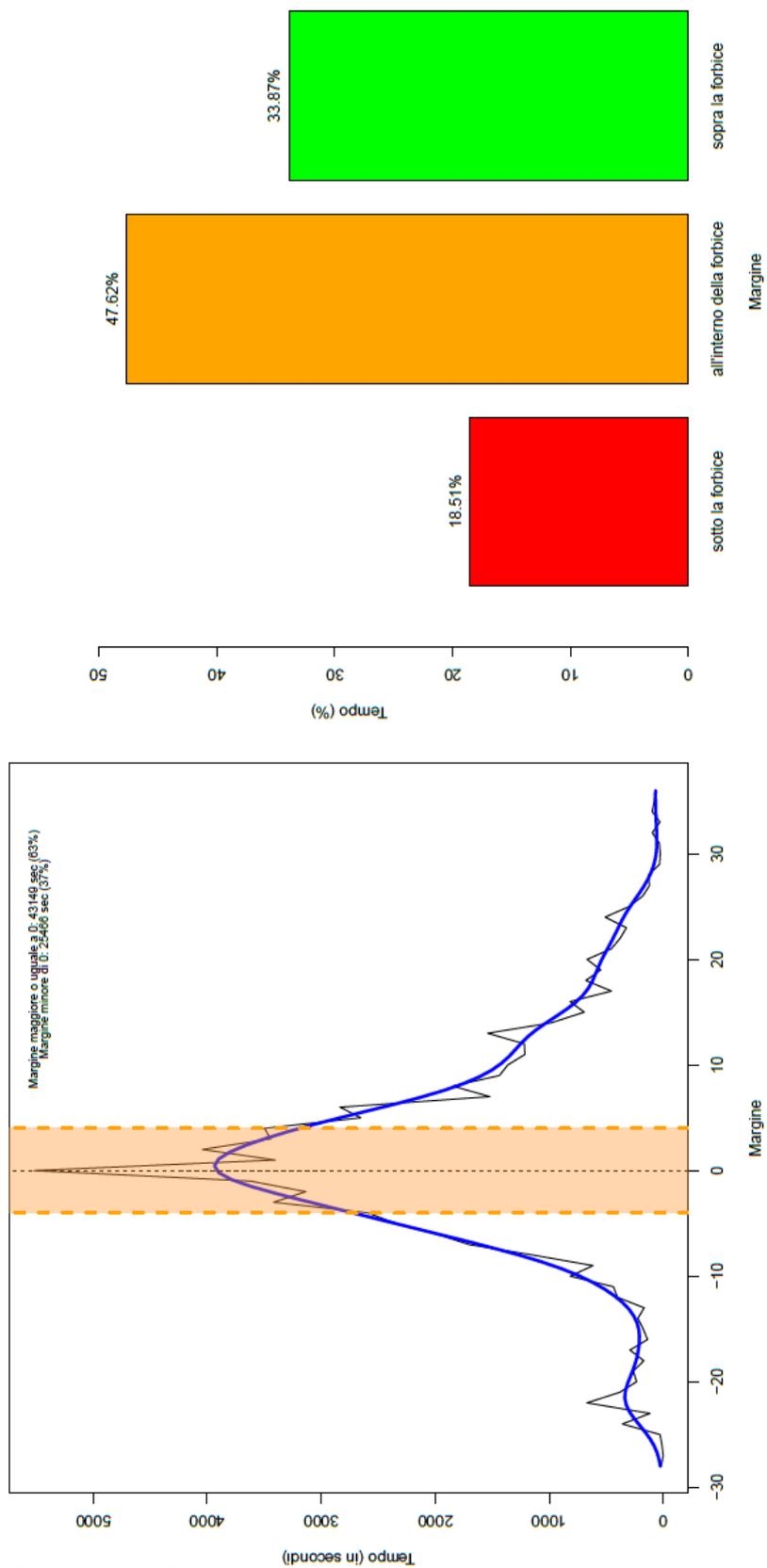


Figura 2.16 – Tempo di permanenza nelle fasce di margine (Brescia)

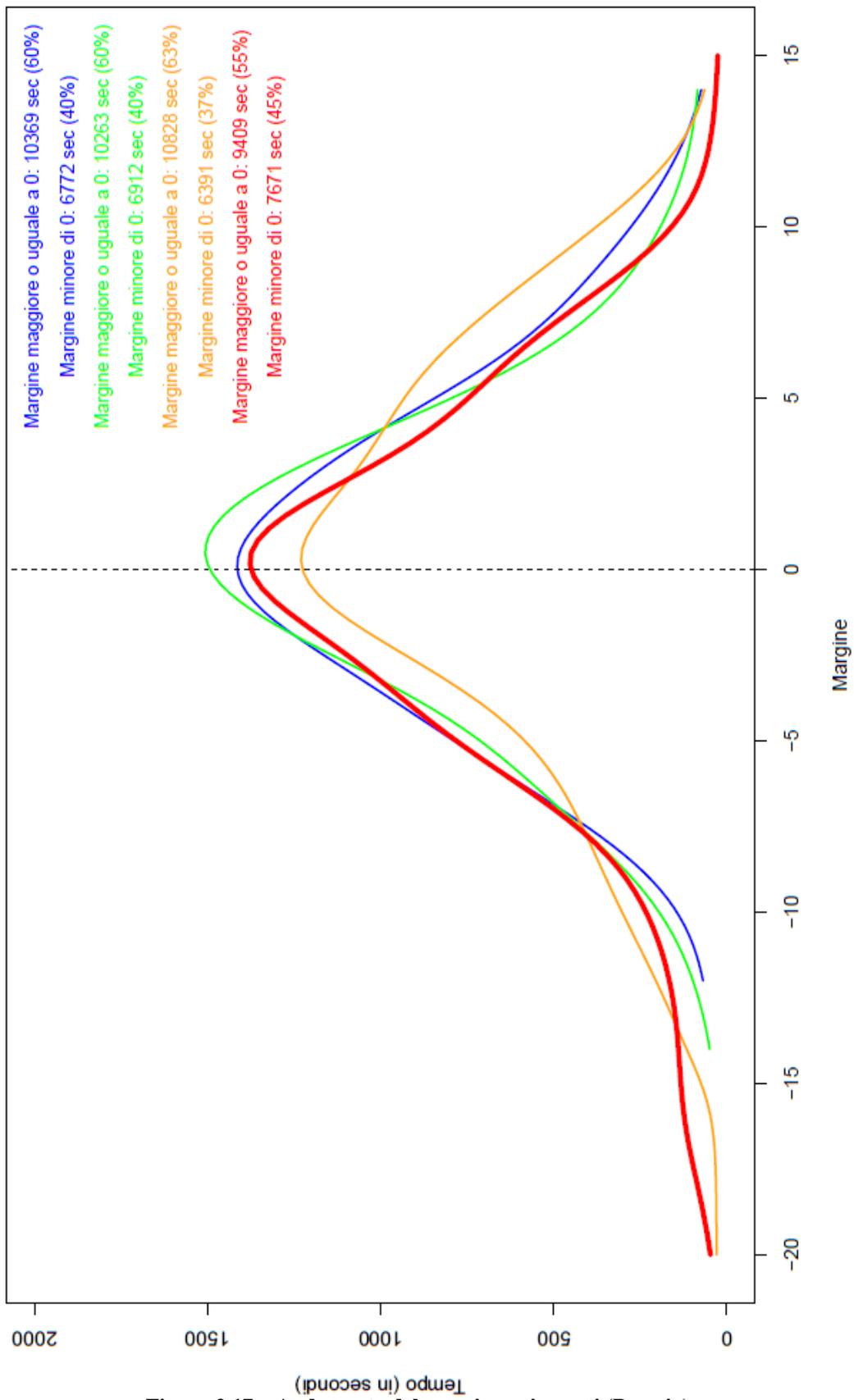
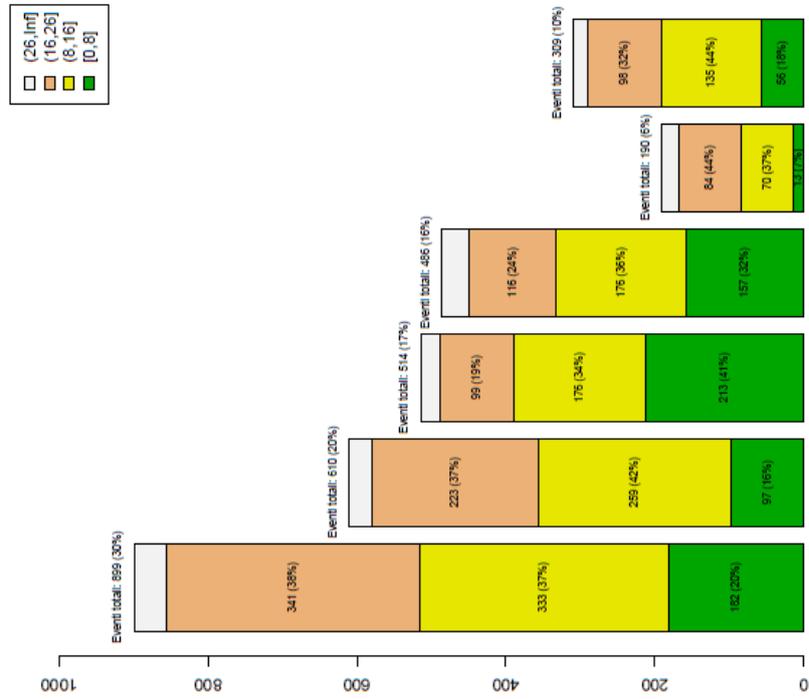


Figura 2.17 – Andamento del margine nei quarti (Brescia)

Eventi che accadono nelle azioni da 24 sec – distribuzione nei tre blocchi



Eventi che accadono nelle azioni da 24 sec – distribuzione nei tre blocchi

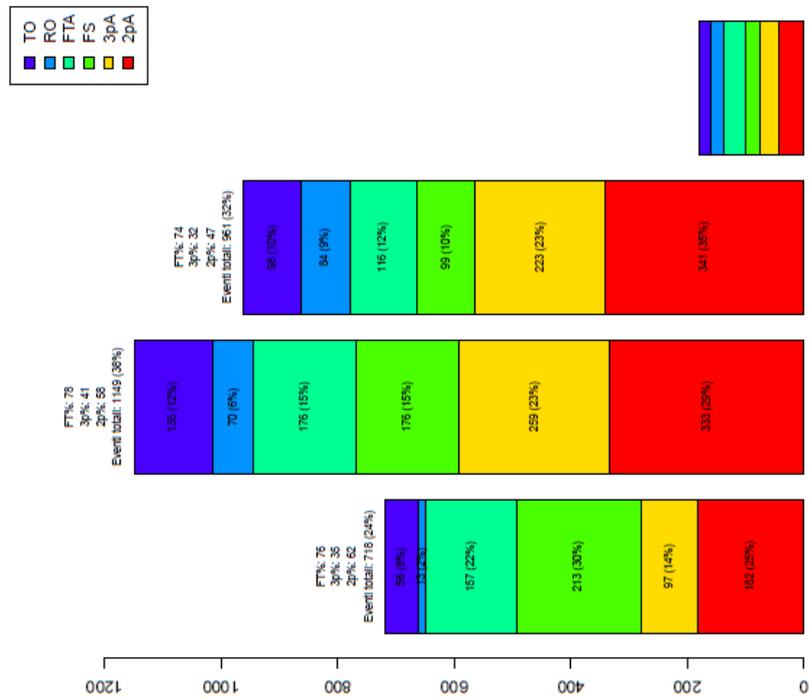


Figura 2.18 – Eventi interni alle azioni da 24 secondi (Brescia)

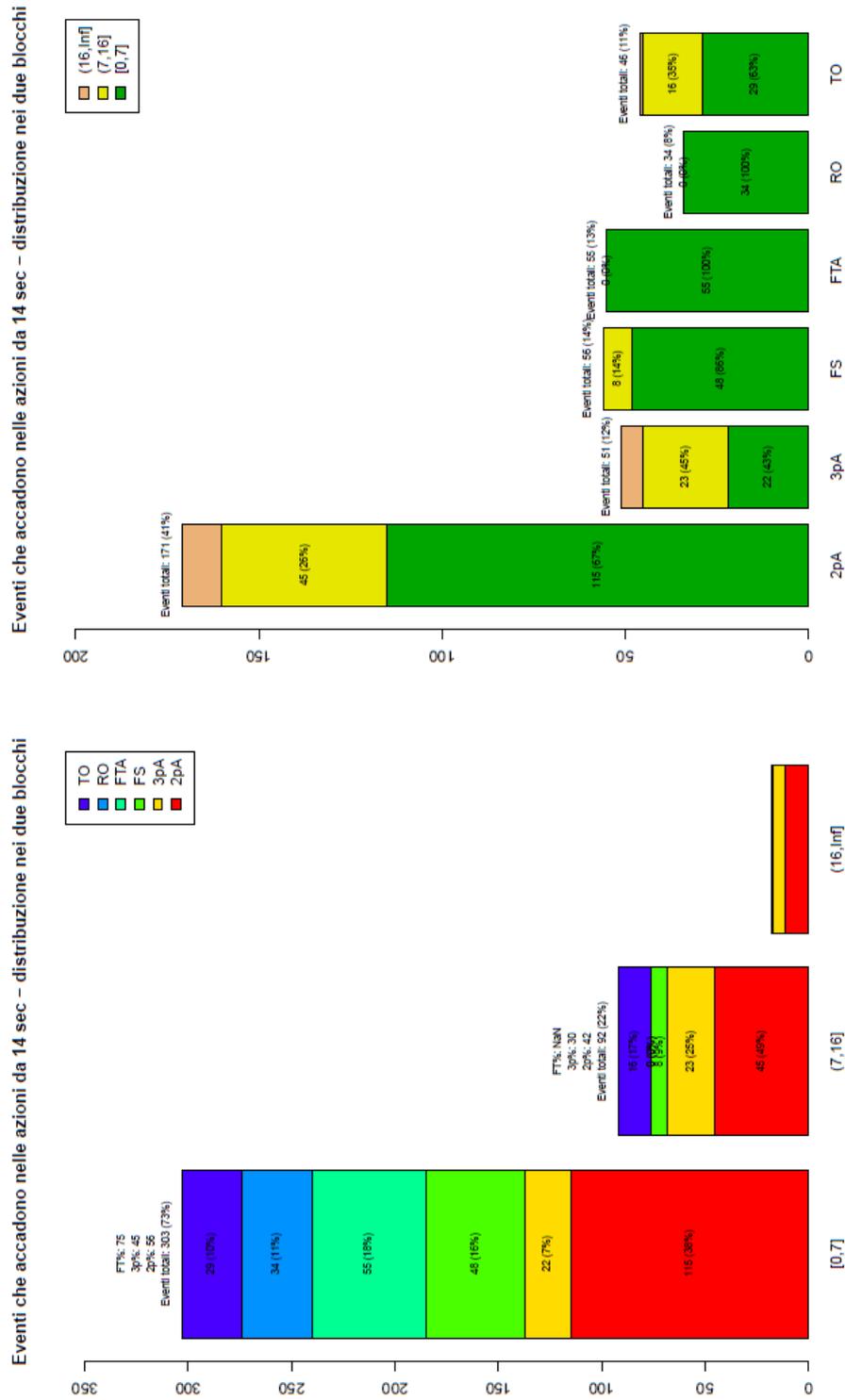


Figura 2.19 – Eventi interni alle azioni da 14 secondi (Brescia)

Tiri da 2 punti nei 24 sec

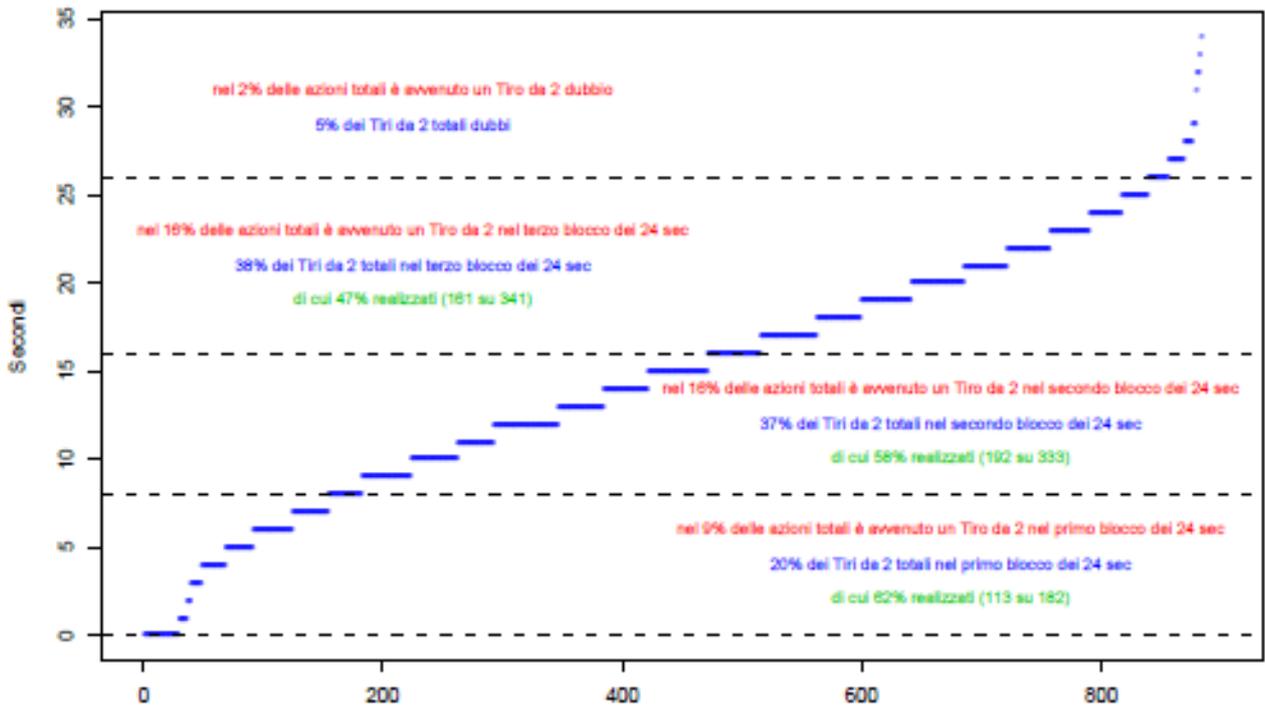


Figura 2.20 – Tiri da 2 punti nei 24 secondi (Brescia)

Tiri da 3 punti nei 24 sec

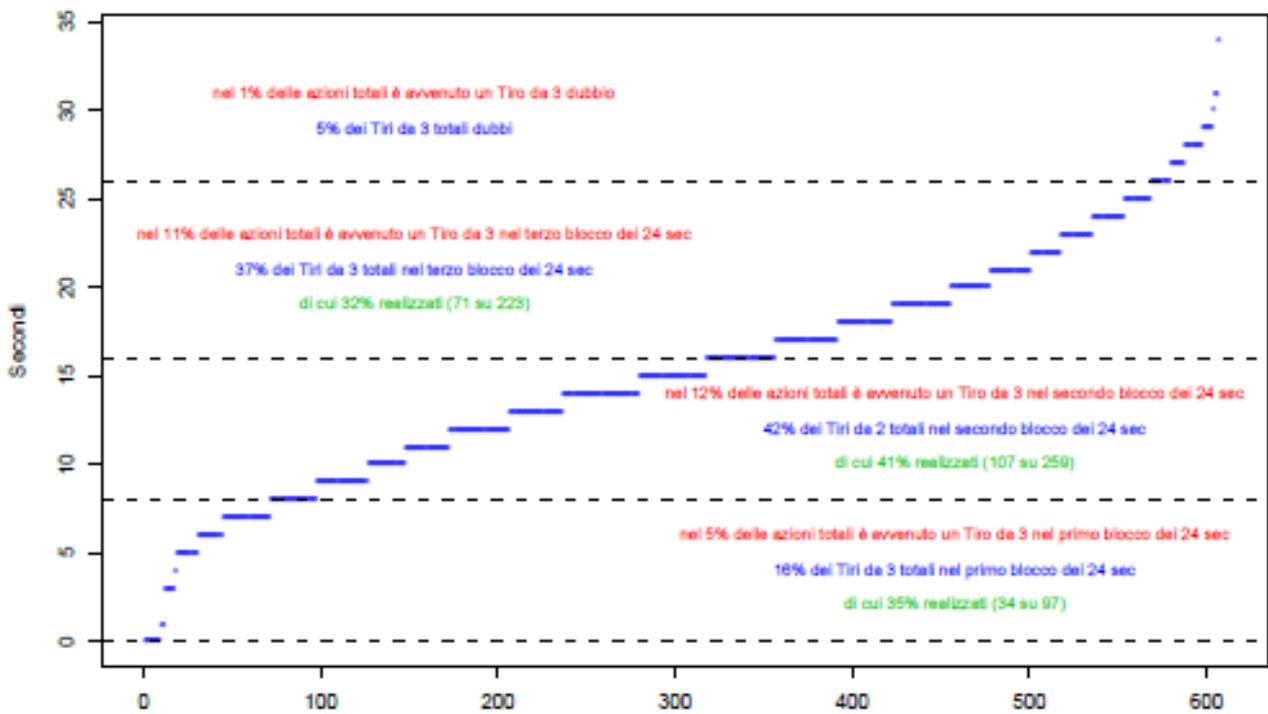


Figura 2.21 – Tiri da 3 punti nei 24 secondi (Brescia)

Tiri da 2 punti nei 14 sec

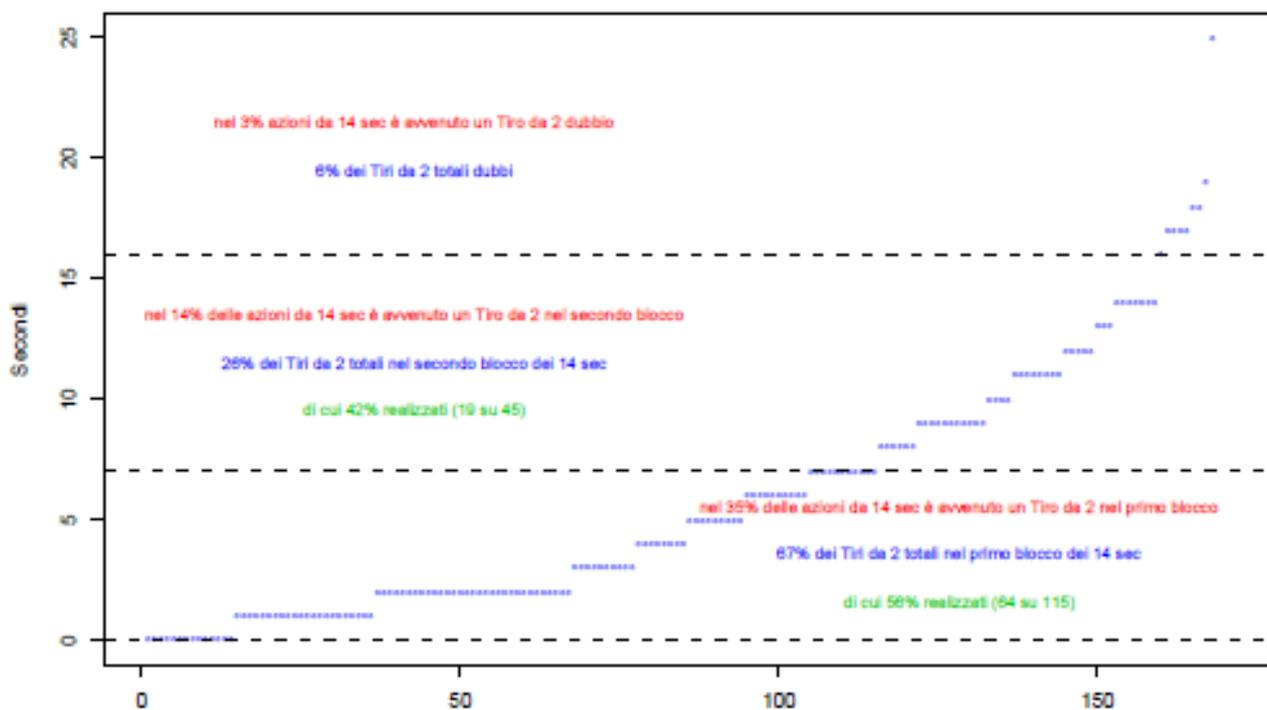


Figura 2.22 – Tiri da 2 punti nei 14 secondi (Brescia)

Tiri da 3 punti nei 14 sec

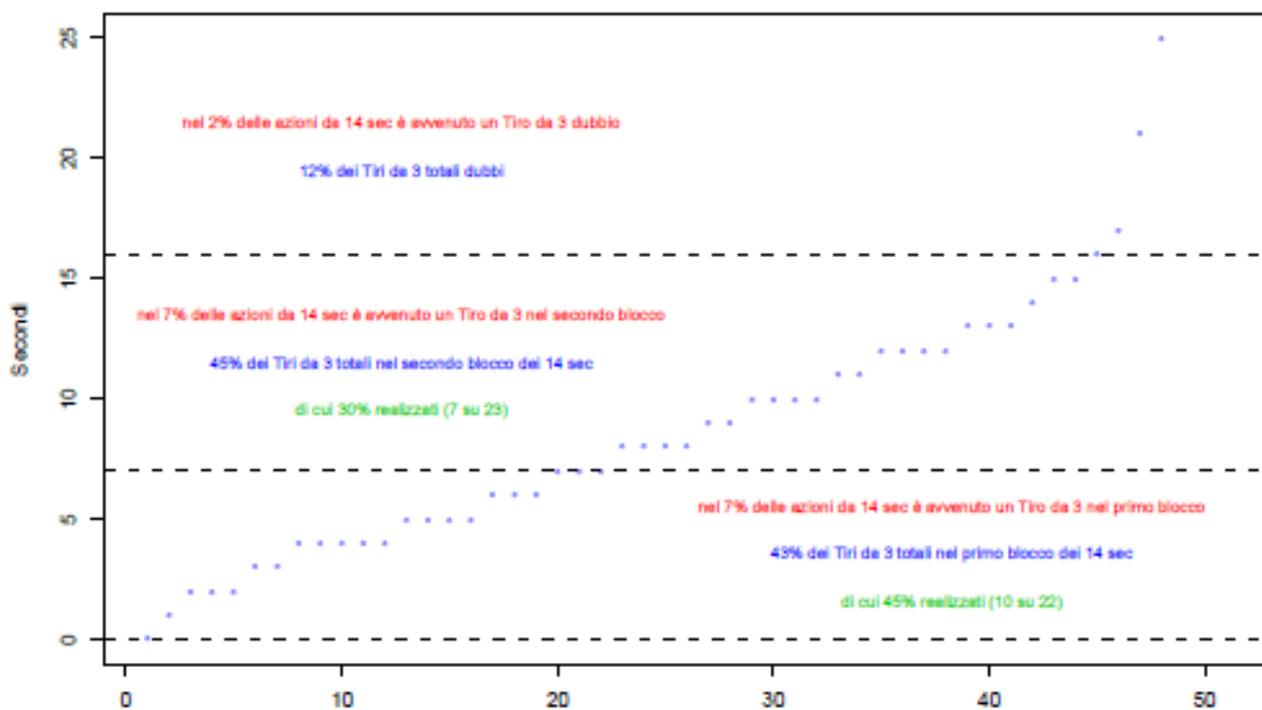


Figura 2.23 – Tiri da 3 punti nei 14 secondi (Brescia)

Margine - IP	Min	2pM	2pA	2p%	2pM40	3pM	3pA	3p%	3pM40	FTM	FTA	FT%	FT40	Pts40	PtsOpp40
A: sotto la forbice	212	103	220	46.82	19.46	56	137	40.88	10.58	106	139	76.26	20.03	90.69	68.40
B: interno alla forbice	545	299	525	56.95	21.96	104	319	32.60	7.64	170	220	77.27	12.49	79.33	72.21
C: sopra la forbice	387	170	325	52.31	17.56	78	205	38.05	8.05	138	182	75.82	14.25	73.53	83.54

Margine - IP	Min	FS	FS40	RO	RO40	TO	TO40
A: sotto la forbice	212	142	26.83	48	9.07	65	12.28
B: interno alla forbice	545	247	18.14	107	7.86	153	11.24
C: sopra la forbice	387	175	18.07	69	7.13	137	14.15

Legenda prima tabella	Margine - IP
Margine nelle tre categorie - Intera Partita	Margine - IP
Tempo di permanenza in minuti nelle tre categorie di margine	Min
Tiri da 2 riusciti	2pM
Tiri da 2 totali	2pA
Tiri da 2 % successo	2p%
Tiri da 2 rapportati al tempo (40 minuti)	2pM40
Tiri da 3 riusciti	3pM
Tiri da 3 totali	3pA
Tiri da 3 % successo	3p%
Tiri da 3 rapportati al tempo (40 minuti)	3pM40
Tiri liberi riusciti	FTM
Tiri liberi totali	FTA
Tiri liberi % successo	FT%
Tiri liberi rapportati al tempo (40 minuti)	FT40
Punti fatti rapportati al tempo (40 minuti)	Pts40
Punti subiti rapportati al tempo (40 minuti)	PtsOpp40

Legenda seconda tabella	Margine - IP
Margine nelle tre categorie - Intera Partita	Margine - IP
Tempo di permanenza in minuti nelle tre categorie di margine	Min
Falli subiti	FS
Falli subiti rapportati al tempo (40 minuti)	FS40
Rimbalzi offensivi	RO
Rimbalzi offensivi rapportati al tempo (40 minuti)	RO40
Palle perse	TO
Palle perse rapportate al tempo (40 minuti)	TO40

FORBICE: tra-4 e 4

Figura 2.24 – Report di Brescia nelle fasce di margine (intera partita)

Margine – UQ	Min	2pM	2pA	2p%	2pM40	3pM	3pA	3p%	3pM40	FTM	FTA	FT%	FT40	Pts40	PtsOpp40
A: sotto la forbice	59	29	58	50.00	19.53	12	30	40.00	8.08	44	54	81.48	29.64	92.96	92.96
B: interno alla forbice	98	50	95	52.63	20.34	15	55	27.27	6.10	38	51	74.51	15.46	74.43	63.85
C: sopra la forbice	127	57	111	51.35	17.96	23	58	39.66	7.25	61	84	72.62	19.22	76.89	88.24

Margine – UQ	Min	FS	FS40	RO	RO40	TO	TO40
A: sotto la forbice	59	46	30.99	13	8.76	16	10.78
B: interno alla forbice	98	56	22.78	26	10.57	29	11.79
C: sopra la forbice	127	67	21.11	19	5.99	47	14.81

Legenda prima tabella	Margine – UQ
Margine nelle tre categorie – Ultimo Quarto	Margine – UQ
Tempo di permanenza in minuti nelle tre categorie di margine	Min
Tiri da 2 riusciti	2pM
Tiri da 2 totali	2pA
Tiri da 2 % successo	2p%
Tiri da 2 rapportati al tempo (40 minuti)	2pM40
Tiri da 3 riusciti	3pM
Tiri da 3 totali	3pA
Tiri da 3 % successo	3p%
Tiri da 3 rapportati al tempo (40 minuti)	3pM40
Tiri liberi riusciti	FTM
Tiri liberi totali	FTA
Tiri liberi % successo	FT%
Tiri liberi rapportati al tempo (40 minuti)	FT40
Punti fatti rapportati al tempo (40 minuti)	Pts40
Punti subiti rapportati al tempo (40 minuti)	PtsOpp40

Legenda seconda tabella	Margine – UQ
Margine nelle tre categorie – Ultimo Quarto	Margine – UQ
Tempo di permanenza in minuti nelle tre categorie di margine	Min
Falli subiti	FS
Falli subiti rapportati al tempo (40 minuti)	FS40
Rimbalzi offensivi	RO
Rimbalzi offensivi rapportati al tempo (40 minuti)	RO40
Palle perse	TO
Palle perse rapportate al tempo (40 minuti)	TO40

FORBICE: tra-4 e 4

Figura 2.25 – Report di Brescia nelle fasce di margine (ultimo quarto)

Margine - IP	Min	2pM	2pA	2p%	2pM40	3pM	3pA	3p%	3pM40	FTM	FTA	FT%	FT40	Pts40
A: sotto la forbice	146	9	15	60.0	2.46	12	37	32.43	3.29	4	6	66.67	1.10	15.88
B: interno alla forbice	391	17	34	50.0	1.74	25	75	33.33	2.56	12	16	75.00	1.23	12.39
C: sopra la forbice	262	15	24	62.5	2.29	26	51	50.98	3.97	16	20	80.00	2.44	18.92

Margine - IP	Min	FS	RO	FS40	RO	RO40	TO	TO40
A: sotto la forbice	146	9	2.46	1	0.27	6	1.64	
B: interno alla forbice	391	20	2.05	6	0.61	15	1.54	
C: sopra la forbice	262	18	2.75	2	0.31	9	1.37	

Legenda prima tabella	Margine - IP
Margine nelle tre categorie - Intera Partita	Margine - IP
Tempo di permanenza in minuti nelle tre categorie di margine	Min
Tiri da 2 riusciti	2pM
Tiri da 2 totali	2pA
Tiri da 2 % successo	2p%
Tiri da 2 rapportati al tempo (40 minuti)	2pM40
Tiri da 3 riusciti	3pM
Tiri da 3 totali	3pA
Tiri da 3 % successo	3p%
Tiri da 3 rapportati al tempo (40 minuti)	3pM40
Tiri liberi riusciti	FTM
Tiri liberi totali	FTA
Tiri liberi % successo	FT%
Tiri liberi rapportati al tempo (40 minuti)	FT40
Punti fatti rapportati al tempo (40 minuti)	Pts40

Legenda seconda tabella	Margine - IP
Margine nelle tre categorie - Intera Partita	Margine - IP
Tempo di permanenza in minuti nelle tre categorie di margine	Min
Falli subiti	FS
Falli subiti rapportati al tempo (40 minuti)	FS40
Rimbalzi offensivi	RO
Rimbalzi offensivi rapportati al tempo (40 minuti)	RO40
Palle perse	TO
Palle perse rapportate al tempo (40 minuti)	TO40

FORBICE: tra-4 e 4

Minuti giocatore analizzati: 799

Figura 2.26 – Report di Mirza Alibegovic nelle fasce di margine (intera partita)

Margine - UQ	Min	2pM	2pA	2p%	2pM40	3pM	3pA	3p%	3pM40	FTM	FTA	FT%	FT40	Pts40
A: sotto la forbice	46	4	7	57.14	3.45	3	8	37.50	2.59	2	2	100.00	1.72	16.39
B: interno alla forbice	56	0	2	0.00	0.00	3	10	30.00	2.14	6	7	85.71	4.28	10.71
C: sopra la forbice	71	5	7	71.43	2.84	6	9	66.67	3.40	7	10	70.00	3.97	19.85

Margine - UQ	Min	FS	FS40	RO	RO40	TO	TO40
A: sotto la forbice	46	1	0.86	1	0.86	0	0.00
B: interno alla forbice	56	4	2.86	0	0.00	3	2.14
C: sopra la forbice	71	5	2.84	1	0.57	1	0.57

Legenda prima tabella	Legenda seconda tabella
Margine nelle tre categorie - Ultimo Quarto	Margine nelle tre categorie - Ultimo Quarto
Tempo di permanenza in minuti nelle tre categorie di margine	Tempo di permanenza in minuti nelle tre categorie di margine
Tiri da 2 riusciti	Falli subiti
Tiri da 2 totali	Falli subiti riportati al tempo (40 minuti)
Tiri da 2 % successo	Rimbalzi offensivi
Tiri da 2 riportati al tempo (40 minuti)	Rimbalzi offensivi riportati al tempo (40 minuti)
Tiri da 3 riusciti	Palle perse
Tiri da 3 totali	Palle perse riportate al tempo (40 minuti)
Tiri da 3 % successo	
Tiri da 3 riportati al tempo (40 minuti)	
Tiri liberi riusciti	
Tiri liberi totali	
Tiri liberi % successo	
Tiri liberi riportati al tempo (40 minuti)	
Punti fatti riportati al tempo (40 minuti)	

FORBICE: tra-4 e 4

Minuti giocatore analizzati: 173

Figura 2.27 – Report di Mirza Alibegovic nelle fasce di margine (ultimo quarto)

3 Gli alberi di decisione

In questo capitolo, si presentano alcuni concetti teorici necessari per comprendere la prima parte delle analisi svolte nei capitoli successivi. La teoria in questione riguarda gli alberi di decisione.

3.1 Alberi di classificazione e di regressione

L'albero decisionale (o Decision Tree) è uno schema di classificazione o di regressione ad albero, come suggerisce la parola stessa, e costituisce un metodo non parametrico per fare delle previsioni. È molto popolare in statistica grazie alla semplicità dei risultati che consente una facile comprensione ed interpretazione; inoltre, sono di particolare interesse qualora vengano applicati a insiemi di dati molto numerosi.

Un albero è costituito da un insieme finito di elementi detti nodi; dal nodo iniziale (o radice) partono i rami che individuano i nodi figli. Quelli terminali, per i quali non è ritenuta utile nessuna ulteriore suddivisione, sono chiamati foglie. Per ogni foglia è definita una precisa regola di classificazione delle osservazioni che può essere letta seguendo il percorso che collega il nodo iniziale alla foglia stessa.

Vi sono due tipologie di implementazione degli alberi di decisione:

- alberi di regressione, nei quali la variabile risposta è continua (pertanto si utilizzano per stimare i valori di variabili target che assumono solo valori numerici);
- alberi di classificazione, nei quali la variabile risposta è quantitativa discreta o qualitativa.

La procedura generale è iterativa di tipo top – down: su un insieme di individui nel nodo radice si valutano tutte le regole di suddivisione (splitting) per ottenere i nodi figli e, per ogni nodo, si decide se il processo deve arrestarsi o può proseguire secondo una regola di arresto (Fraire, Rizzi, 2011).

3.2 L'algoritmo CART

L'algoritmo CART (Classification And Regression Trees) è stato introdotto da Breiman *et al.* nel 1984: individua il modello ottimale costruendo un albero completo e, in seguito, sfolgendolo per raggiungere l'ottimalità basandosi sui risultati di una classificazione incrociata.

Nella costruzione dell'albero si individuano, quindi, due fasi:

1. crescita (o growing);
2. potatura (o pruning).

Nelle analisi che si basano su questo particolare tipo di algoritmo, intervengono quattro differenti tipi di componenti:

- la variabile che si vuole stimare (variabile target);
- le variabili note (predictors);
- l'insieme dei dati di apprendimento (learning data set) che contengono i valori delle variabili target e i predictors;
- un insieme di dati di verifica.

Le caratteristiche principali di questo algoritmo sono le seguenti:

- a) la procedura è non parametrica;
- b) le variabili non devono essere selezionate in anticipo;
- c) i risultati sono invarianti per trasformazioni monotone alle variabili (per esempio trasformazioni lineari, radici quadrate, logaritmi ecc. non modificano i risultati);
- d) è robusto rispetto agli outlier;
- e) può utilizzare combinazioni lineari di variabili (per esempio, operare sulle componenti principali);
- f) la procedura può operare anche in presenza di dati mancanti.

(Fraire, Rizzi, 2011)

3.2.1 L'indice di impurità: l'indice di Gini

L'albero viene costruito tramite una procedura ricorsiva di partizione binaria. Partendo dal nodo radice, i casi vengono classificati in 2 nodi figli; l'algoritmo continua a procedere in questo modo finché non giunge alla classificazione ottimale, individuata in base ad una regola di arresto (detta regola di splitting).

Nell'algoritmo CART l'indice di impurità, ossia una misura della purezza dei nodi (e, quindi, la regola di splitting), è rappresentato dall'indice di Gini:

$$G(i) = 1 - \sum [f(i, j)]^2$$

ove con i si intende il nodo ed $f(i, j)$ è la frequenza relativa della modalità a_j classificata nel nodo i . Tanto più piccolo sarà il valore di $G(i)$, tanto maggiore sarà l'omogeneità del nodo. Questo indice varia tra 0 e $(p - 1) / p$ ove p sono le unità statistiche appartenenti al nodo.

Come riportato nel paragrafo precedente, questo algoritmo consente all'albero di raggiungere la sua massima dimensione senza alcun criterio di arresto; il pruning è una procedura ricorsiva a ritroso che consente di ridurre la dimensione dell'albero eliminando alcuni rami.

Partendo dall'albero completo, si determinano alberi sempre più piccoli minimizzando la funzione di perdita:

$$R(T) = I(T) + \alpha N(T)$$

dove con T si indica un determinato albero, $I(T)$ è la funzione di impurità totale dell'albero (determinata sulla base dei $G(i)$), $N(T)$ è il numero delle foglie ed α è una costante che stabilisce la penalizzazione per la complessità desiderata (Fraire, Rizzi, 2011). Si tratta di una cosiddetta "funzione di costo – complessità": se l'impurità totale $I(T)$ diminuisce, la complessità $N(T)$ aumenta, rendendo così difficile ottenere un albero poco impuro e, al contempo, poco complesso. Occorre trovare un bilanciamento tra impurità e complessità, minimizzando la funzione $R(T)$.

3.2.2 La bontà del modello: l'indice AUC

La curva ROC (Receiver Operating Characteristics) è uno degli strumenti più utilizzati per valutare la qualità di un classificatore binario, ossia la capacità di fornire risultati corretti. Definendo sensibilità la frequenza relativa che un caso positivo sia correttamente classificato e specificità la probabilità di corretta classificazione per un caso negativo, e indicando con VP i veri positivi, FP i falsi positivi, VN i veri negativi, FN i falsi negativi, allora:

- sensibilità = $VP / VP + FN$
- specificità = $VN / VN + FP$.

Il grafico della curva ROC è ottenuto, quindi, riportando in ascissa la frequenza relativa dei falsi positivi (cioè $1 - \text{specificità}$) ed in ordinata la sensibilità al variare del punto di separazione tra le due classi della variabile target. Un esempio è rappresentato nella figura 3.1: la linea continua consiste nella curva ROC, la linea tratteggiata rappresenta la diagonale a 45° .

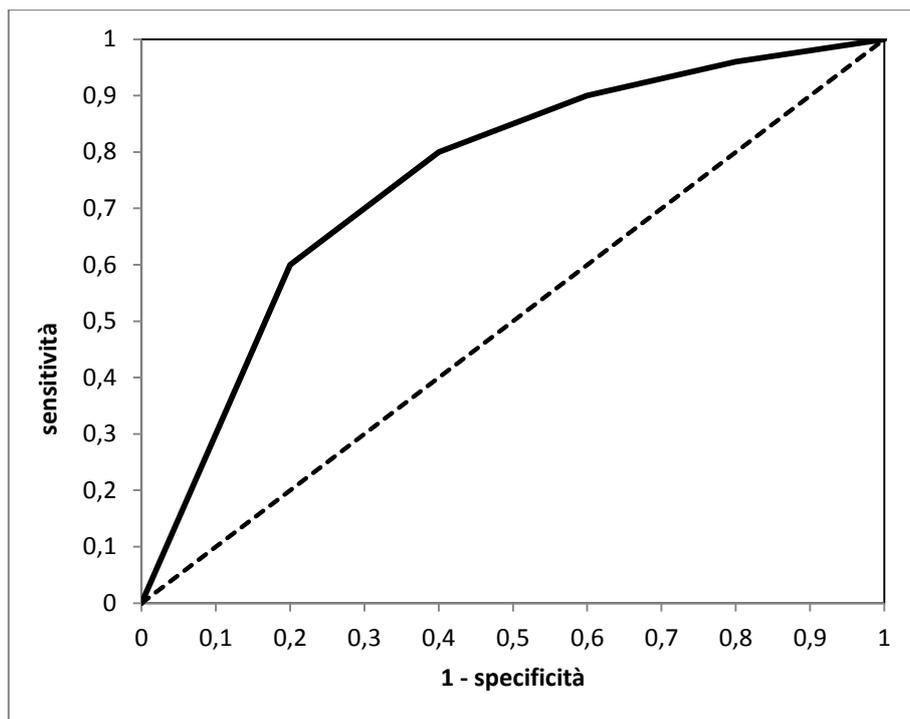


Figura 3.1 – Esempio di ROC Curve

Per un classificatore perfetto, la curva dovrebbe assumere in ordinata il valore 1 per tutti i valori delle ascisse e via via spostarsi verso la diagonale all'aumentare dell'errore di classificazione. Per un modello incapace di discriminare, la curva si appiattisce sulla diagonale. L'area sotto la curva (AUC, Area Under the ROC Curve) rappresenta, quindi, un indicatore della qualità del modello; raggiunge il suo massimo, uguale a 1, per un modello perfetto ed il minimo, uguale a 0,5, nel caso di classificatore randomico completamente inefficiente (Fraire, Rizzi, 2011).

4 La performance nel tiro in situazioni di stress

In questa sezione vengono presentati le situazioni di stress individuate, il modello elaborato che permette di determinare quali situazioni hanno, in media, un effetto significativo sulla probabilità di fare canestro e alcune nuove misurazioni della performance di tiro.

4.1 Le situazioni di stress

Madden *et al.* (1995) identificano 20 situazioni di gioco stressanti, tutte legate ad eventi offensivi, difensivi e neutrali che accadono durante le competizioni di basket. Le sei categorie di situazioni di stress, determinate grazie all'analisi delle componenti principali, sono interpretate dagli autori come *Being Outplayed*, *Errors in Personal Skills*, *Errors in General Play*, *Game Tension*, *Team Performance* e *Other Performance*.

In conformità con le definizioni generali di tali indicatori e considerando i dati a disposizione, si sono individuate, anche grazie all'aiuto di un coach esperto come Marco Crespi, 4 principali situazioni che possono generare pressione o stress sul giocatore mentre si appresta a tirare:

- a. quando il cronometro dei 24 secondi sta scadendo;
- b. quando il divario di punteggio tra le squadre è piccolo;
- c. quando la squadra, per qualunque ragione, non sta mantenendo una performance globale buona durante la partita, fino al momento considerato;
- d. quando il giocatore ha sbagliato il tiro precedente.

Potrebbe essere rilevante anche il tempo mancante alla fine del match, ma è logico ritenere che possa generare stress solo in interazione con altre variabili come, per esempio, il divario di punteggio. Per questo motivo, tale variabile rientra nell'analisi in relazione con altre variabili, come si vedrà nel paragrafo successivo.

Le variabili coinvolte per le 4 situazioni di stress identificate sono:

- a. *finea*: variabile numerica compresa tra 0 e 24 secondi, che indica il tempo sul cronometro dei 24 secondi nel quale ogni tiro viene effettuato;
- b. *forbice*: variabile numerica che indica il divario di punteggio rispetto alla squadra avversaria nel momento in cui ogni tiro viene effettuato;
- c. *missteamp*: variabile numerica compresa tra 0 e 1, che indica, per ogni partita, la frazione di tiri sbagliati dalla squadra fino al momento in cui ogni tiro viene effettuato;

- d. *misspl*: variabile categoriale che indica se il tiro precedente del giocatore è stato realizzato (“Done”) o sbagliato (“Missed”).

4.2 Il modello elaborato

Per elaborare un modello che stimasse l’impatto delle situazioni di gioco stressanti sulla probabilità di fare canestro e per tenere conto delle possibili interazioni tra le variabili, è stato utilizzato l’algoritmo CART (Breiman, Friedman, Stone, Olshen, 1984), i cui principi teorici sono esposti nel paragrafo 3.2.

Data la mole di informazioni contenuta nelle variabili quantitative, le covariate numeriche sono state convertite in covariate categoriali, come mostrato nella tabella 4.1, per alleggerire il processo computazionale e per facilitare l’interpretazione. Anche in questo caso, i consigli di coach Marco Crespi sono stati molto utili.

<i>finea</i>	<i>forbice</i>	<i>missteamp</i>
“just started”: primi 5 secondi	“sotto la forbice [-4,4]”	“good”: meglio del 75th percentile
“time end”: ultimi 2 secondi	“dentro la forbice [-4,4]”	“bad”: peggio del 25th percentile
“normal”: tutti gli altri casi	“sopra la forbice [-4,4]”	“medium”: tutti gli altri casi

Tabella 4.1 – Conversione delle covariate numeriche in categoriali

Inoltre sono state introdotte nell’analisi quattro ulteriori variabili: *T23F* (variabile categoriale che descrive il tipo di tiro, con categorie “2P”, “3P”, “FT” che stanno rispettivamente per tiro da 2 punti, da 3 punti e tiro libero), *Rank* (variabile numerica compresa tra 1 e 16 che indica la posizione in classifica della squadra), *Rank.Opp* (variabile numerica compresa tra 1 e 16 che indica la posizione in classifica della squadra avversaria) e *finemforb* (variabile categoriale che indica se la squadra si trova sotto, sopra o dentro la forbice negli ultimi 5 minuti del match; costituisce un’interazione tra la variabile *forbice* e il tempo mancante a fine partita).

Dopo alcune prove e attente osservazioni, ci si è accorti che le variabili *forbice*, *missteamp* e *finemforb* non risultavano avere un impatto significativo sulla probabilità di segnare un canestro. Occorre tenere presente che, per l’elaborazione dell’albero, sono stati presi in considerazione tutti i dati di tutte le squadre del campionato; è logico, quindi, pensare che tali variabili possano avere un effetto molto significativo per alcuni giocatori (e, quindi, per i team di cui fanno parte) e poco significativo per altri e che, di conseguenza, tali effetti si compensino. Dal momento che non sono risultate significative, comunque, sono state escluse dall’elaborazione dell’albero.

L'albero, quindi, è stato costruito usando le variabili *finea*, *misspl*, *T23F*, *Rank* e *Rank.Opp* come covariate e la variabile *Made* come variabile dipendente. La regola di splitting è rappresentata dall'indice di Gini. Dopo la fase di potatura, si è ottenuto l'albero raffigurato nella figura 4.1.

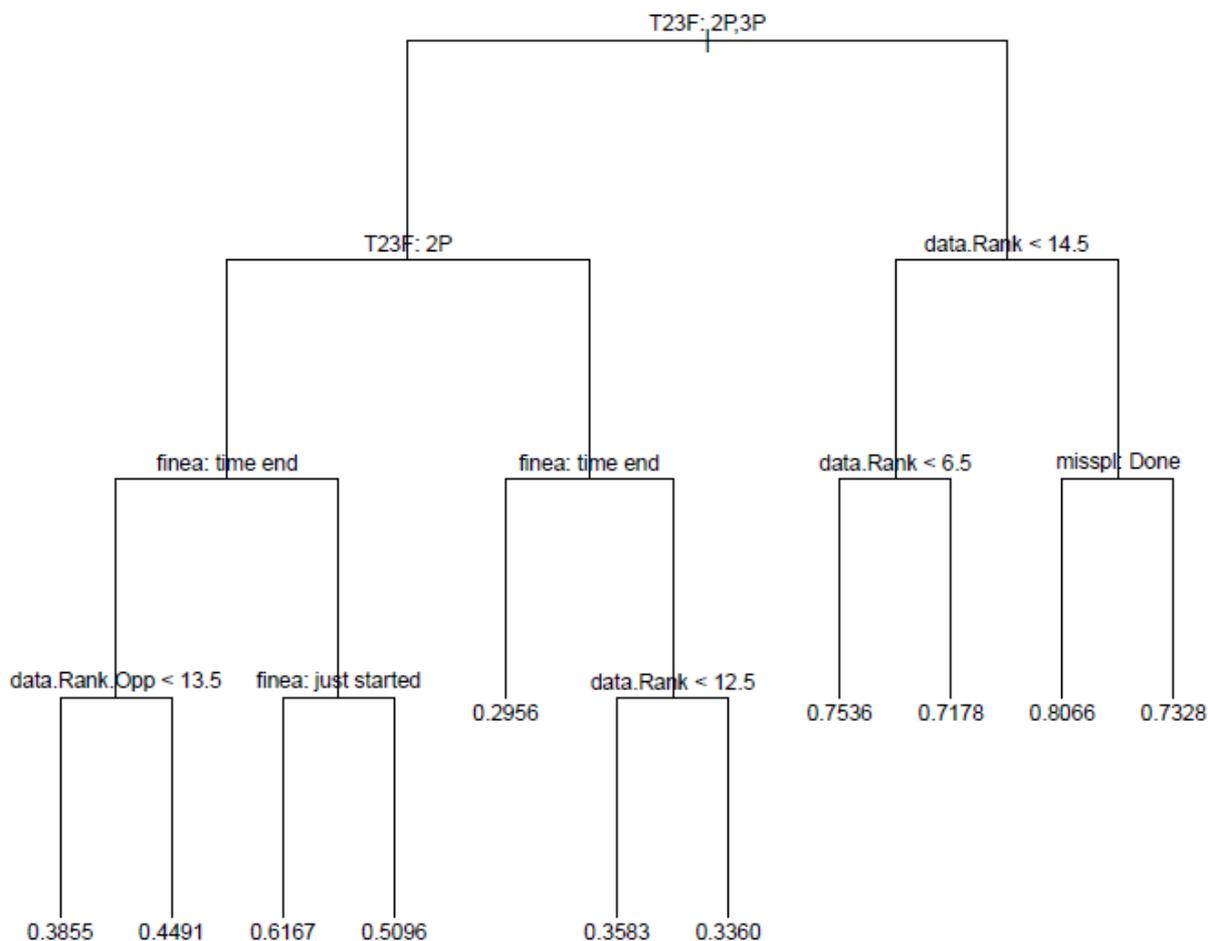


Figura 4.1 – Albero di classificazione (andata)

La bontà del modello, misurata tramite l'indice AUC (Area Under the ROC Curve), è pari a 0,6649: di per sé si tratta di un risultato medio – basso, ma in questo specifico caso si considera accettabile perché occorre tenere conto dell'alta imprevedibilità che caratterizza, per sua natura, il fenomeno analizzato. Più un fenomeno è imprevedibile (cioè caratterizzato da un'alta funzione di casualità), più è difficile prevederne l'esito e, di conseguenza, l'indice sarà più basso.

L'albero offre alcune interessanti interpretazioni riguardo all'impatto delle situazioni di stress sulla probabilità di segnare. Il primo fattore di discriminazione è, com'era logico aspettarsi, il tipo di tiro seguito dal cronometro dei 24 secondi per quanto riguarda i tiri dal campo (cioè tutti i tiri esclusi i tiri liberi). Nel dettaglio, l'albero rivela la seguente struttura di relazioni:

- Per i tiri liberi, un ruolo importante è giocato dalla posizione in classifica: per le squadre che occupano i primi 6 posti, la probabilità stimata di segnare un tiro libero è pari a 0,7536 mentre per le squadre dal 7° al 14° posto è di 0,7178. Per le squadre che occupano gli ultimi due posti in classifica, la probabilità stimata è pari a 0,8066 se il precedente tiro del giocatore è stato segnato, ma scende a 0,7328 se il precedente tiro è stato sbagliato.
- Per i tiri da 3 punti, la prima variabile rilevante è il cronometro dei 24 secondi. La probabilità stimata di segnare allo scadere dell'azione è di 0,2956. Se, invece, il tiro è effettuato prima degli ultimi 2 secondi, conta la posizione in classifica: per le squadre nei primi 12 posti, la probabilità stimata è di 0,3583 contro 0,3360 delle squadre di bassa classifica.
- Anche per i tiri da 2 punti la prima variabile rilevante è rappresentata dal cronometro dei 24 secondi. In caso di tiro effettuato negli ultimi 2 secondi, giocando contro un avversario di medio – alta classifica (fino al 13° posto) la probabilità stimata è 0,3855 contro 0,4491 qualora l'avversario occupi una posizione di bassa classifica. In caso di tiro effettuato nei primi 5 secondi dell'azione, la probabilità stimata è pari a 0,6167; negli altri casi è pari a 0,5096.

Come anticipato in precedenza, variabili come il divario di punteggio o il divario di punteggio negli ultimi minuti del match non sono risultati significativi ai fini delle stime della probabilità di segnare. Questo perché, come riportato nell'introduzione, alcuni giocatori vivono le situazioni di stress come competitive mentre, in altri, le stesse situazioni generano ansia e pressione. Come risultato totale, tale situazione sembra non influire sull'esito di un tiro. Per questo motivo, nei capitoli successivi si terranno conto anche di tali variabili e del loro possibile effetto sulle performance di tiro a livello di squadra.

4.3 Nuove misurazioni della performance di tiro

Il modello appena descritto può essere utile per sviluppare nuove misurazioni della performance di tiro, capaci di tenere conto delle circostanze nelle quali ciascun tiro è stato tentato. Il ragionamento alla base di tale affermazione consiste nel fatto che, secondo il modello CART, non tutti i tiri sono uguali. Per esempio, un tiro da 2 punti tentato negli ultimi 2 secondi dell'azione ha una probabilità di realizzazione compresa tra il 38% e il 45% (a seconda della posizione in classifica dell'avversario), mentre un tiro da 2 punti tentato nei primi 5 secondi ha una probabilità di realizzazione superiore al 60%. L'idea è quella di costruire una misurazione della performance che

dia il giusto valore ad ogni canestro, in corrispondenza con la rispettiva probabilità di realizzazione del tiro, in contrapposizione con le classiche misurazioni secondo le quali ogni tiro è identico agli altri.

Per ciascun tipologia di tiro T (2P: 2 punti; 3P: 3 punti; FT: tiro libero), sia J_T l'insieme di tutti i tiri tentati di tipo T . Inoltre, sia x_{ij} l'indicatore che assume valore 1 se il j -esimo tiro dell' i -esimo individuo (squadra, in questa analisi) è stato segnato e 0 se sbagliato, e sia π_{ij} la sua probabilità secondo il modello CART. Per ciascun tiro, la differenza $x_{ij} - \pi_{ij}$ può essere usata come misura della performance di tiro: infatti, è positiva se il tiro è stato realizzato (e più è bassa la probabilità, maggiore è il suo valore) e negativa se il tiro è stato sbagliato (e più alta è la probabilità, maggiore è il suo valore assoluto). Di conseguenza, un canestro è più rilevante se la probabilità di realizzazione del rispettivo tiro è bassa; per contro, un errore è più penalizzato se la probabilità di realizzazione del rispettivo tiro è alta.

Si può, quindi, definire la performance di tiro della squadra i per la tipologia di tiro T come:

$$P_i(T) = av_{j \in J_T}(x_{ij} - \pi_{ij})$$

dove $av_{j \in J_T}$ indica la media di tutti i tiri di tipo T tentati dalla squadra i .

Sfruttando questo nuovo indice, si possono ottenere altri indici che permettono di valutare la performance di tiro nelle seguenti tre situazioni di stress:

- a. negli ultimi 2 secondi di possesso palla;
- b. quando il divario di punteggio rispetto all'avversario è compreso nella forbice [-4;4];
- c. negli ultimi 5 minuti del match, quando il divario di punteggio rispetto all'avversario è compreso nella forbice [-4;4].

Definendo J_S l'insieme di tiri tentati nella situazione di stress S , vengono introdotti i seguenti indici (che possono essere calcolati per ogni tipologia di tiro T):

- differenziale di performance (rispetto alla performance generale) in S :

$$D_i^{(S)}(T) = av_{j \in J_T \cap J_S}(x_{ij} - \pi_{ij}) - av_{j \in J_T}(x_{ij} - \pi_{ij}) = P_i^{(S)}(T) - P_i(T)$$

dove $av_{j \in J_T \cap J_S}$ indica la media di tutti i tiri di tipo T tentati dalla squadra i nella situazione S . Questo indice stabilisce se, nella situazione S , la performance della squadra migliora o peggiora, calcolando l'incremento o il decremento assoluto dell'indice $P_i(T)$ quando i tiri sono effettuati nella situazione di stress S .

- propensione al tiro in S :

- per il caso a., si calcola la frazione (%) dei tiri tentati negli ultimi 2 secondi del possesso dalla squadra i :

$$R_i^{(S)}(T) = \left(\frac{\sum_{j \in J_T \cap J_S} x_{ij}}{\sum_{j \in J_T} x_{ij}} \right) \cdot 100$$

- per i casi b. e c., si calcola il rapporto tra la frazione di tiri del tipo T tentati in S e la frazione di minuti giocati in S dalla squadra i :

$$R_i^{(S)}(T) = \left[\frac{\frac{\sum_{j \in J_T \cap J_S} x_{ij}}{\sum_{j \in J_T} x_{ij}}}{\frac{T_i^{(S)}}{T_i}} - 1 \right] \cdot 100 = \left[\frac{\sum_{j \in J_T \cap J_S} x_{ij}}{\sum_{j \in J_T} x_{ij}} \cdot \frac{T_i}{T_i^{(S)}} - 1 \right] \cdot 100$$

dove T_i e $T_i^{(S)}$ indicano rispettivamente i minuti giocati dalla squadra i in totale e nella situazione S .

Questi indici stabiliscono la misura in cui la squadra prende il rischio di tirare durante la situazione di stress S .

Va sottolineato che gli indici proposti possono essere calcolati solo quando si dispone di un ragionevole ammontare di osservazioni sia in totale che nella situazione S , per ogni individuo e per ogni tipo di tiro. Per questa ragione, queste analisi sono state condotte sulle squadre e non sui singoli giocatori. L'analisi sui singoli giocatori, dopo aver selezionato solo quelli con un numero rilevante di tiri, è contenuta nel paper che verrà presentato alla conferenza internazionale sulla statistica di Siviglia (Zuccolotto, Manisera, Sandri, 2016).

5 Analisi sul girone di andata

5.1 Rappresentazioni grafiche dei nuovi indici

Dopo aver calcolato, per ogni squadra, gli indici esposti nel capitolo precedente, sono state costruite diverse rappresentazioni grafiche per analizzare la situazione al termine del girone di andata.

Il primo grafico costruito (figura 5.1) è rappresentato dallo scatterplot che riporta la performance nel tiro da 2 punti $P_i(2P)$ sull'asse delle ascisse e la performance nel tiro da 3 punti $P_i(3P)$ sull'asse delle ordinate; ogni squadra è rappresentata da un cerchio la cui dimensione indica il numero di tiri effettuati e il cui colore indica la performance nei tiri liberi $P_i(FT)$, in base alla legenda riportata a fianco.

Una prima considerazione che si può trarre dalla situazione generale raffigurata consiste nel fatto che sembra non esserci una relazione diretta tra le performance al tiro e la posizione in classifica. Si osservino, infatti, le classifiche dei due gironi al termine della prima parte del campionato:

GIRONE EST

1. Centrale del Latte – Amica Natura Brescia
2. Andrea Costa Imola
3. De' Longhi Treviso
4. Dinamica Generale Mantova
5. Tezenis Verona
6. Mec Energy Roseto
7. Eternedile Bologna
8. OraSi Ravenna
9. Remer Treviglio
10. Alma Agenzia per il Lavoro Trieste
11. Proger Chieti
12. Bondi Ferrara
13. Europromotion Legnano
14. Betulline Jesi
15. Basket Recanati
16. Bawer Matera

GIRONE OVEST

1. Moncada Agrigento
2. Givova Scafati
3. Novipiù Casale Monferrato
4. LightHouse Conad Trapani
5. Orsi Tortona
6. BCC Agropoli
7. Assigeco Casalpusterlengo
8. Benacquista Ass. Latina
9. Mens Sana Basket 1871 Siena
10. FMC Ferentino
11. Npc Rieti
12. Acea Roma
13. Bermè Reggio Calabria
14. Paffoni Omegna
15. Angelico Biella
16. La Briosa Barcellona

In generale, le squadre con le migliori performance sono Agropoli (nel tiro da 2 punti, nonostante una performance negativa nel tiro da 3 punti), Scafati (nel tiro da 3 punti) e Biella (nei tiri liberi, nonostante performance negative da 2 e 3 punti).

Tra le squadre più performanti (quadrante in alto a destra: performance da 2 e da 3 punti positive), figurano Scafati, Agrigento, Trapani e Tortona che occupano le prime posizioni nel loro girone; tuttavia, si trovano anche Ferrara, Casalpusterlengo, Ravenna e Latina che occupano le posizioni di medio – bassa classifica. Inoltre nel quadrante opposto (in basso a sinistra) con le performance negative, si trovano squadre come Verona, Treviso e Imola che sono tra le prime classificate nel loro girone. Curiosa la situazione di Brescia, prima nel girone est: buona performance da 2 punti e ai tiri liberi, un po' meno nel tiro dalla lunga distanza. All'incirca nella medesima situazione si trovano anche squadre come Chieti e Ferentino (che si trovano in zone medio – bassa classifica), Reggio Calabria e Roma (bassa classifica). Altrettanto curioso notare come Roseto, Recanati e Omegna si trovino quasi esattamente in corrispondenza dell'origine degli assi: significa che mantengono performance di tiro che si attestano sulla loro media, accompagnata da una performance leggermente negativa nei tiri liberi. Tuttavia, mentre la prima si trova al 6° posto nel suo girone, le seconde si trovano rispettivamente in penultima e terzultima posizione nei rispettivi gironi. Altre squadre, di medio – alta classifica, mostrano una situazione opposta: da una parte Bologna e Mantova, con performance leggermente negativa da 2 punti ma positiva da 3 punti e nei tiri liberi, dall'altra Agropoli, con performance positiva da 2 punti e negativa da 3 punti (nei tiri liberi, invece, si mantiene in linea con la media). Situazione simile a quest'ultima è quella di Barcellona, con la differenza di una performance migliore nei tiri liberi, che tuttavia occupa l'ultima posizione nel girone ovest. L'ultima della classe del girone est, invece, ossia Matera, manifesta performance negative in tutte le tipologie di tiro.

Si noti, quindi, come alcuni posizionamenti in classifica siano spiegati dalle performance di tiro (positive o negative), mentre alcune divergenze o analogie non sembrano rispecchiarsi in altrettante diversità o affinità nel grafico. Come si illustrerà successivamente nel prossimo capitolo, questo particolare aspetto del gioco che è stato analizzato può spiegare in parte alcune situazioni di classifica, così come alcune variazioni occorse in seguito al girone di ritorno; tuttavia, come è logico pensare, non spiega il gioco nella sua totalità poiché ne tralascia altri aspetti (ad esempio, le situazioni difensive) che sono altrettanto importanti e che potrebbero essere approfonditi, andando in tal modo ad integrare queste analisi, offrendo una visione più ampia.

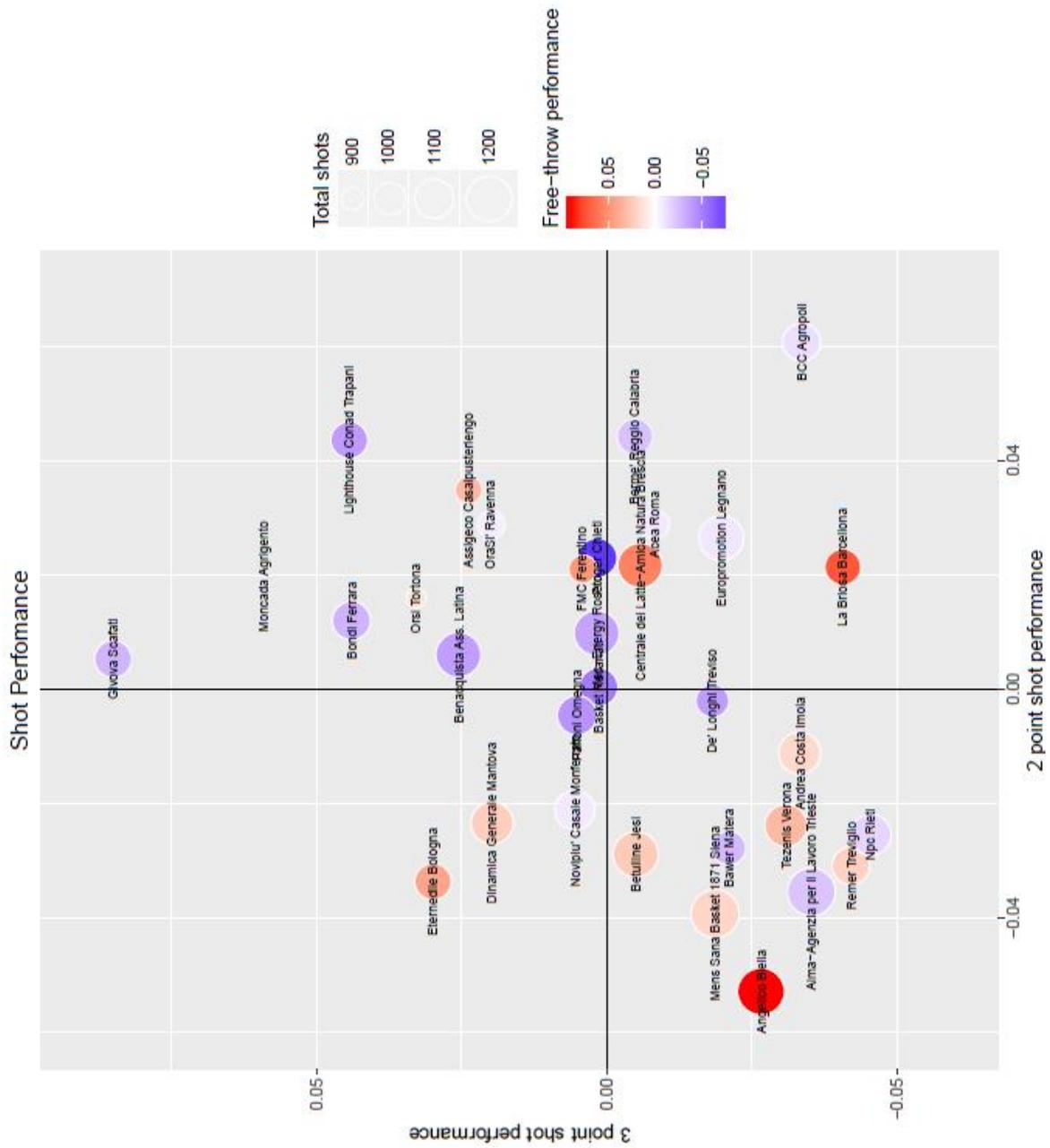


Figura 5.1 – Scatterplot: performance di tiro (andata)

Approfondendo l'analisi, sono stati costruiti anche i grafici riguardanti le situazioni di stress precedentemente identificate, sfruttando gli altri indici calcolati (propensione al tiro e differenziale di performance).

Per quanto riguarda il tiro negli ultimi secondi dell'azione, nel grafico è riportata sull'asse delle ascisse la percentuale di realizzazione generale da 2 o 3 punti, mentre sull'asse delle ordinate è riportato il differenziale di performance negli ultimi 2 secondi. La grandezza del cerchio indica il numero di tiri effettuati in questa situazione e il colore ne rappresenta la percentuale di sul totale di tiri effettuati. Sono stati tralasciati i tiri liberi poiché sono tiri effettuati a cronometro fermo e, pertanto, non possono essere influenzati dai secondi di possesso palla. I risultati sono riportati in figura 5.2 e 5.3.

Come si può notare, la grande maggioranza delle squadre si trova sotto l'asse x: significa che sono molto poche le squadre che mostrano un miglioramento di performance nei tiri effettuati negli ultimi 2 secondi di possesso. Occorre tenere presente che difficilmente si arriva a tirare volutamente nelle immediate vicinanze dello scadere dei 24 secondi; tendenzialmente, si tira prima di arrivare "con l'acqua alla gola"! Infatti, la durata media delle azioni si attesta tra 16 – 17 secondi. Di conseguenza, è intuitivo pensare che la maggior parte dei tiri effettuati in questa situazione possano essere i cosiddetti "tiri della disperazione", ossia un tiro effettuato in condizioni non ottimali (solitamente con forte pressione difensiva) pur di non far scadere il cronometro senza aver concluso a canestro. Quindi, non deve trarre in inganno il colore rosso dei cerchi raffiguranti le squadre poiché la percentuale di tiri in questa situazione rimane, comunque, molto bassa (tra il 10% e il 18% circa). Per quanto riguarda i tiri da 2 punti, si nota un leggero miglioramento di performance da parte di Recanati e Agropoli. Per quanto riguarda il tiro da 3 punti, un miglioramento più marcato è visibile per Ferentino, Tortona e Roma, meno marcato quello di Imola, Siena e Treviso.

In generale, questi grafici possono dare indicazioni sulla tendenza delle squadre ad effettuare azioni offensive particolarmente lunghe e a come si comportano, in termini di realizzazione; contro squadre particolarmente efficienti in questa situazione, è necessario mantenere alta la concentrazione e la pressione difensiva fino all'ultimo secondo, ancora più del solito. Viceversa, contro squadre che tendono ad effettuare azioni più brevi e che diminuiscono la performance con il passare dei secondi (in gergo, le squadre che adottano la tattica del "run and gun"), potrebbe essere utile "allungare" la difesa, cioè mettere forte pressione fin dalla rimessa, anche tentando raddoppi sul portatore di palla, in modo da far perdere più secondi possibile all'avversario e costringerlo a prendere un tiro affrettato negli ultimi secondi di possesso.

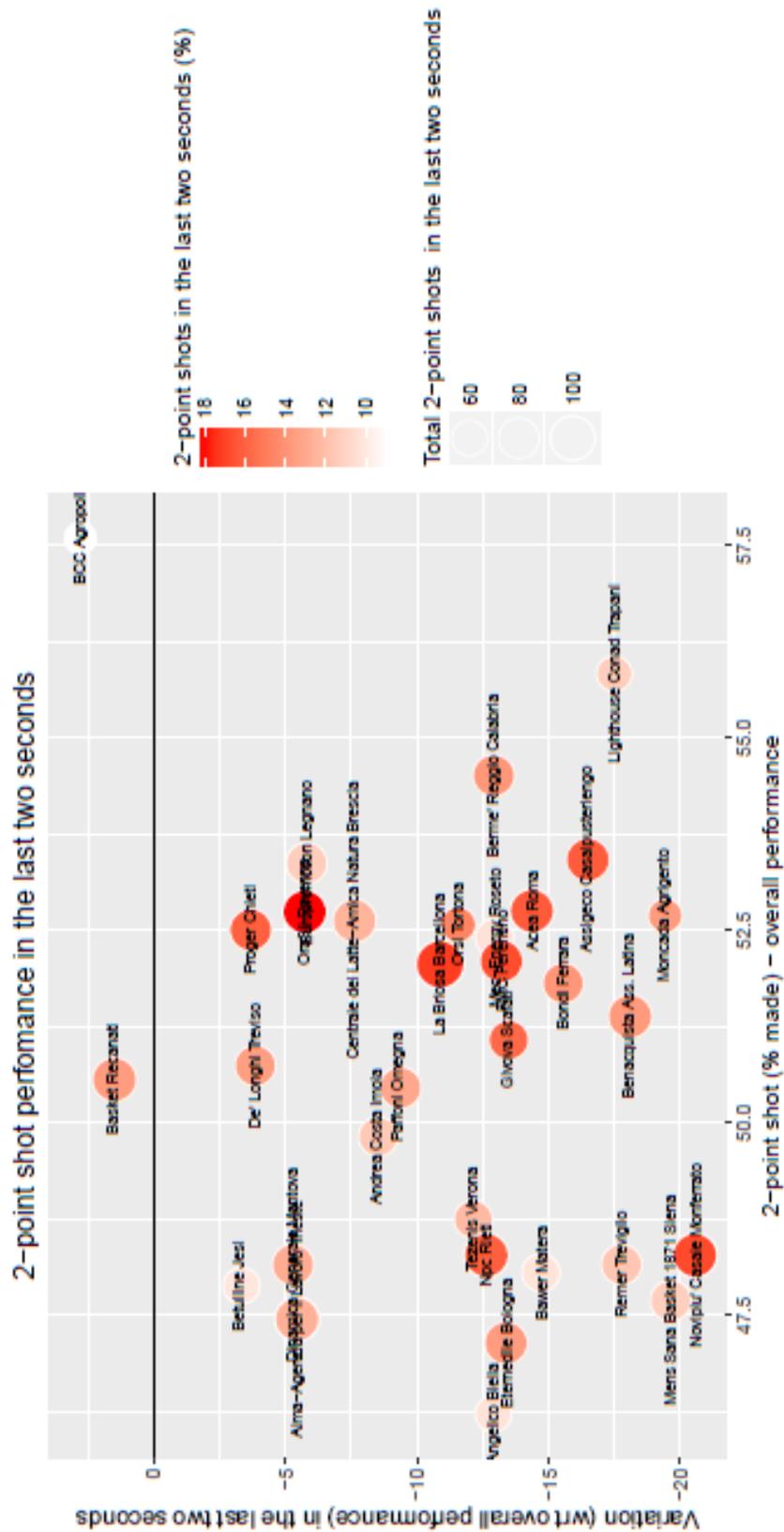


Figura 5.2 – Scatterplot: tiro da 2 pt negli ultimi 2 secondi (andata)

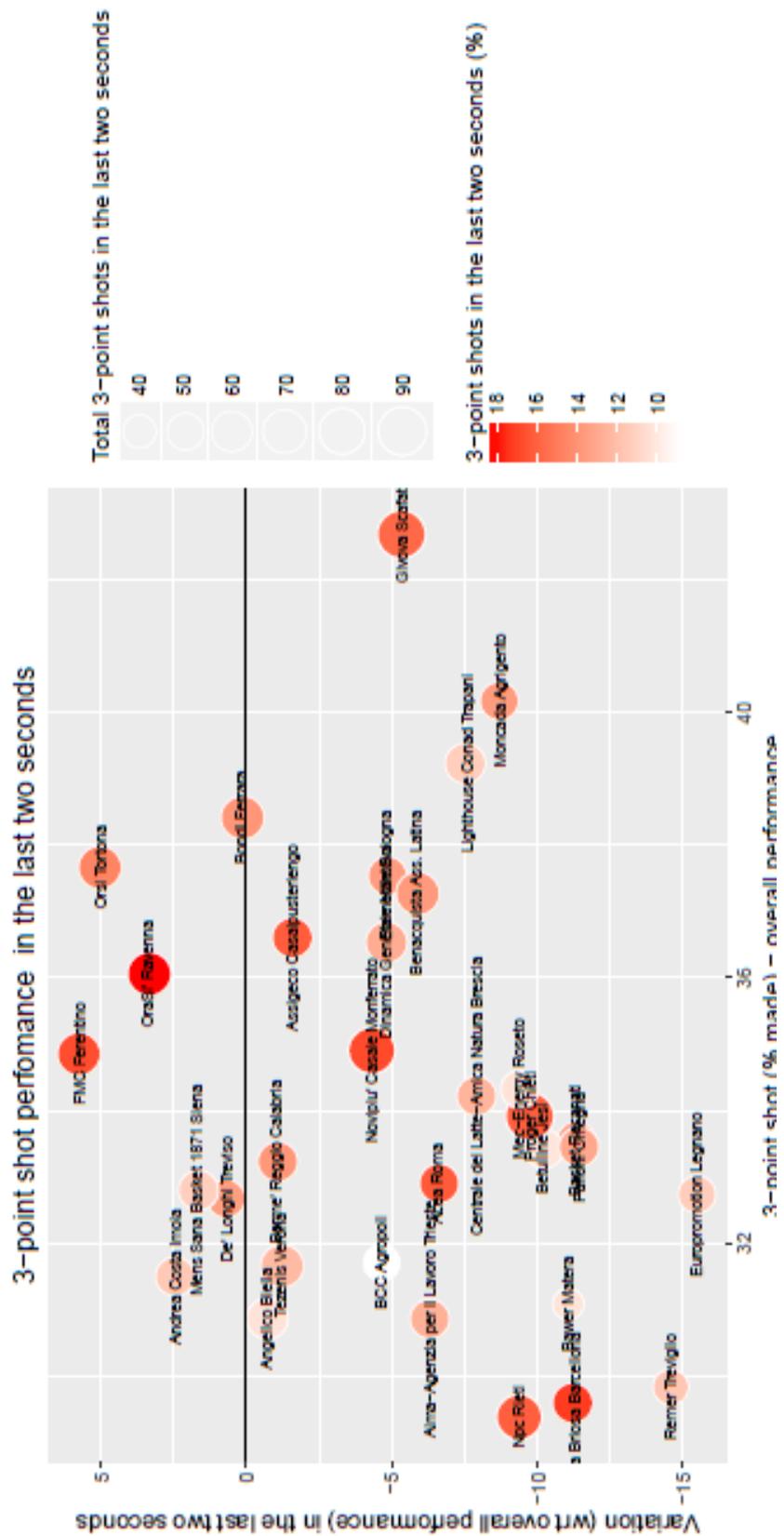


Figura 5.3 – Scatterplot: tiro da 3 pt negli ultimi 2 secondi (andata)

I seguenti grafici riportati in figure 5.4, 5.5 e 5.6 rappresentano, invece, le performance di tiro quando le squadre si trovano all'interno della forbice [-4; 4]: sull'asse delle ascisse è riportata la percentuale di realizzazione generale, sull'asse delle ordinate il differenziale di performance. In questa situazione, il numero di tiri è indicato dalla grandezza del cerchio e la propensione al tiro dal colore. Per quanto riguarda il tiro da 2 punti, significativi sono i miglioramenti nella performance da parte di Brescia (nonostante la bassa propensione al tiro in tale situazione) e Treviso. Significativa, anche se per motivi opposti, pure la situazione di Verona: nonostante l'alta propensione al tiro, la performance della squadra tende a peggiorare. Confermate, inoltre, le buone prestazioni di Ferentino, Agropoli e Scafati. Le squadre che fanno registrare, per contro, peggioramenti significativi sono Recanati in primis, seguita da Biella, Reggio Calabria e Trapani (nonostante queste ultime mostrino delle buone percentuali di realizzazione generale). Per quanto riguarda il tiro da 3 punti, un miglioramento importante nella performance è mostrato da Scafati, Mantova e Trapani, anche se quest'ultima mostra una più bassa propensione al tiro. Da sottolineare come la situazione di Verona (ma anche di Ferentino e Trieste) sia la stessa mostrata nel tiro da 2 punti: alta propensione al tiro, ma peggioramento (in questo caso, più lieve) della performance. Inoltre, non deve trarre in inganno la posizione di Matera e Barcellona: nonostante ci sia un effettivo miglioramento della performance (pur con basse percentuali di realizzazione), il colore blu indica una propensione al tiro molto bassa, cioè indica come, in realtà, i tiri effettuati dalle squadre in questa situazione siano pochi (analoga situazione delle stesse squadre nel tiro da 2 punti). Brescia, invece, mostra un peggioramento della performance e si trova nella stessa situazione di Biella, Roma, Chieti, Recanati, Treviso e Reggio Calabria, ossia tra le situazioni peggiori. Per quanto riguarda i tiri liberi, la situazione è abbastanza omogenea. Si notano il miglioramento della performance da parte di Roma e i peggioramenti da parte di Chieti, Reggio Calabria e Latina (da sottolineare, però, la dimensione dei cerchi di Roma e Reggio che indicano un basso numero di tiri liberi effettuati). Tra le squadre con maggiore propensione al tiro figurano Verona, Trieste e Ferentino. Tuttavia, il maggior numero di tiri liberi guadagnati non è adeguatamente sfruttato poiché la performance rimane ancorata intorno alla media e non mostra un miglioramento sensibile. In generale, il fatto che le squadre siano ben disperse sopra e sotto la linea dello zero conferma quanto esposto in precedenza riguardo le reazioni personali dei giocatori in questa situazione (e, conseguentemente, delle squadre di cui fanno parte): alcuni migliorano, mentre altri tendono a peggiorare. Il risultato è che tali effetti si bilanciano e, per questo motivo, l'algoritmo CART non ha rilevato effetti di questa particolare variabile.

Questi grafici, riguardanti questa particolare situazione, possono essere utili per variare la strategia di gioco in base alle performance generali della squadra avversaria: ad esempio, contro una squadra che, in situazione di testa a testa, tende ad avere un calo nella prestazione dalla lunga distanza può essere efficiente schierare la difesa a zona (che, portando i giocatori ad occuparsi di zone del campo e non dei singoli avversari, permette di occupare l'area e lasciare più sguarnita la zona intorno all'arco dei 3 punti, concedendo qualche tiro in più agli avversari).

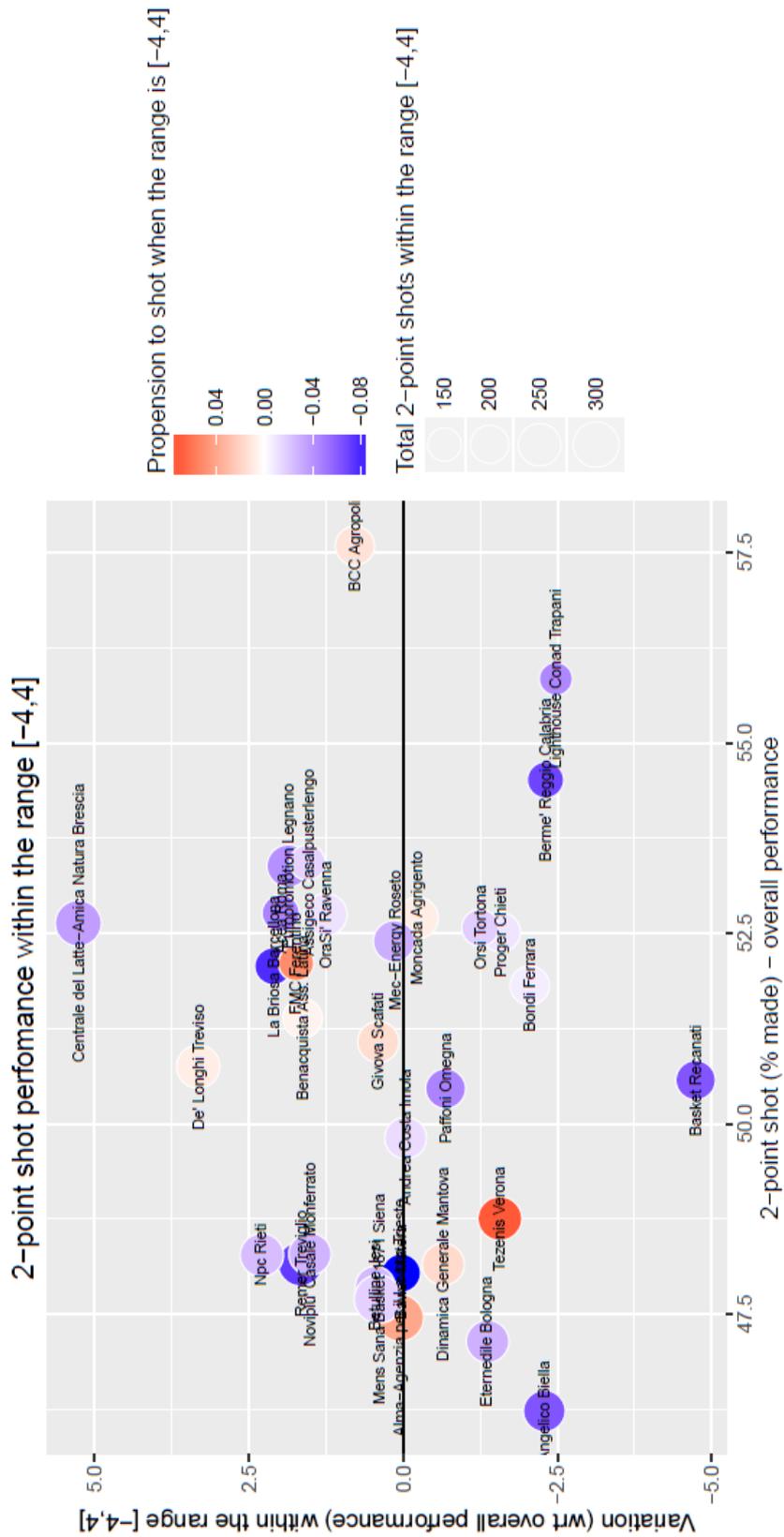


Figura 5.4 – Scatterplot: tiro da 2 pt all’interno della forbice (andata)

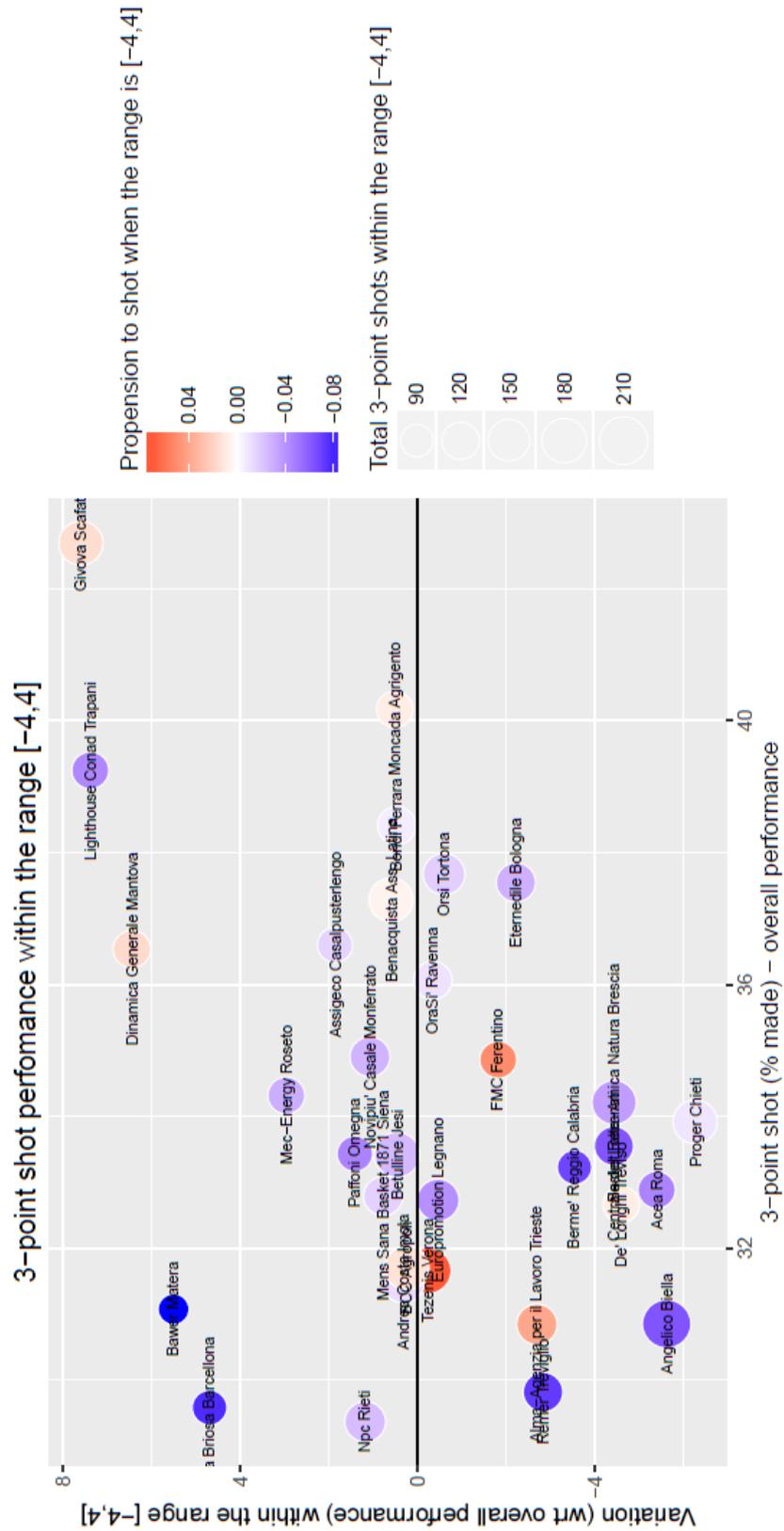


Figura 5.5 – Scatterplot: tiro da 3 pt all'interno della forbice (andata)

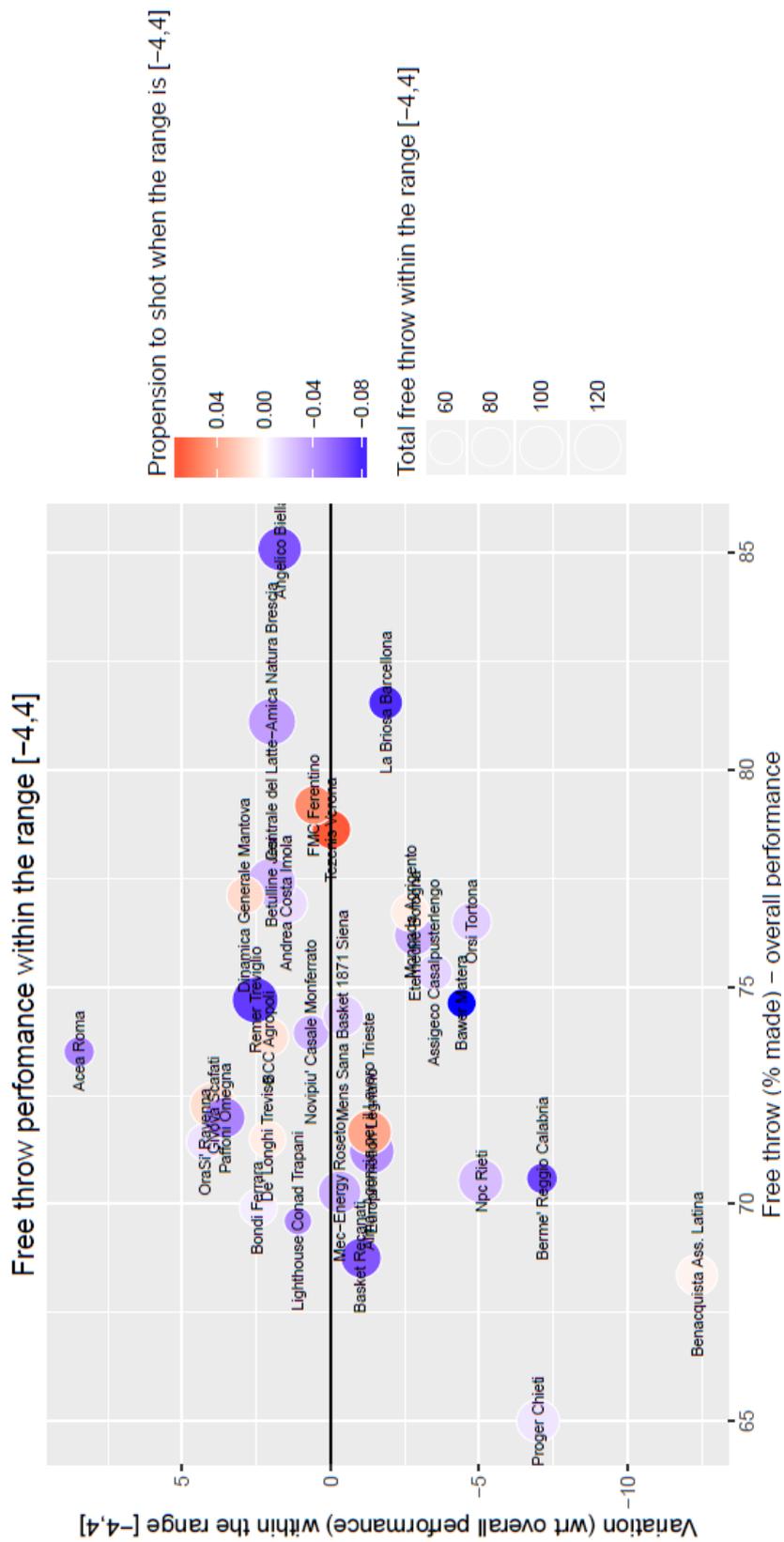


Figura 5.6 – Scatterplot: tiro libero all'interno della forbice (andata)

L'ultima situazione analizzata riguarda le performance di tiro nel caso in cui il punteggio sia all'interno della forbice [-4; 4] negli ultimi 5 minuti di partita. I seguenti grafici nelle figure 5.7, 5.8 e 5.9 raffigurano tale situazione e valgono le stesse chiavi di lettura fornite per i precedenti grafici. Da tenere presente, in questo caso, l'esiguo numero di osservazioni (come si nota dalla legenda).

Per quanto riguarda i tiri da 2 punti, i miglioramenti più consistenti sono fatti registrare da Latina, Agrigento e Legnano (nonostante la poca propensione al tiro). I miglioramenti di Reggio Calabria e Matera non devono ingannare poiché sono caratterizzati da un ristretto numero di tiri e da poca propensione. Altre squadre come Siena e Treviglio migliorano, ma partono da una bassa percentuale di realizzazione generale. Tra le squadre che tirano di più in questa situazione e che hanno alte propensioni (Trieste, Brescia e Imola), invece, non vi sono particolari variazioni di performance, che rimane pressappoco ferma intorno alla media.

Per quanto riguarda il tiro da 3 punti, spicca il consistente miglioramento di Scafati (che oltretutto fa registrare la migliore percentuale di realizzazione). Migliora anche Jesi, che tuttavia paga la minore vena realizzativa. Anche Casalpusterlengo e Agrigento migliorano, ma con una propensione al tiro più bassa. Ancora una volta, le squadre che tirano maggiormente e che hanno maggiore propensione, non sembrano manifestare variazioni significative di performance o manifestano un peggioramento (Brescia, Imola, Ferentino, Bologna, Tortona, Omegna, Treviglio ecc.).

Per quanto riguarda i tiri liberi, si notano miglioramenti significativi da parte di Biella (con la percentuale di realizzazione più alta), Imola, Mantova, Scafati, Ravenna e Agrigento (seppure con minore propensione e minor numero di tiri). Per contro, si osserva il netto peggioramento di Chieti, che è anche la squadra con la percentuale di realizzazione più bassa. In lieve calo anche Brescia, mentre il peggioramento di Reggio Calabria è caratterizzato da bassa propensione e pochi tiri liberi effettuati.

Interessante notare, infine, la situazione di Verona: alta propensione al tiro, ma peggioramento della performance in tutte le tipologie di tiro (particolarmente evidente nel tiro da 2 punti, solo lievemente nel tiro da 3 e nei tiri liberi). Come evidenziato nelle analisi preliminari (paragrafo 2.2), quando la squadra si trova in svantaggio per buona parte del match, raramente riesce poi a vincere. Inoltre, dopo attente ed approfondite discussioni con il coach Marco Crespi e tutto lo staff tecnico, era emerso che la squadra sembrava perdere concentrazione nei momenti chiave della partita. Come si suole dire in gergo: quando conta, la squadra non segna!

Grazie agli indici sviluppati, si è riusciti a dare un risvolto pratico a quelle sensazioni: la scarsa precisione nei momenti chiave è stata quantificata e rappresentata graficamente, nelle varie situazioni.

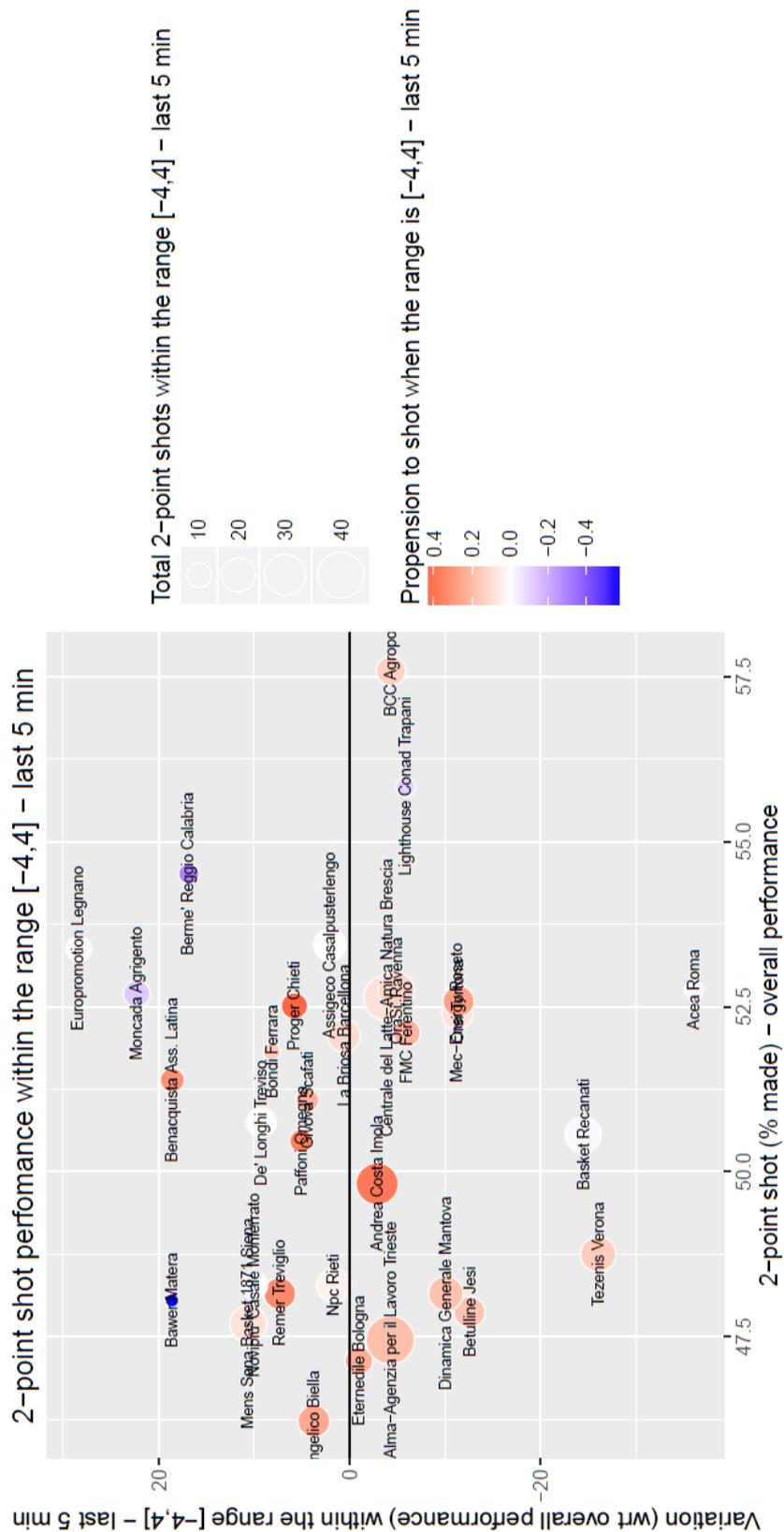


Figura 5.7 – Scatterplot: tiro da 2 pt all'interno della forbice-ultimi 5 min (andata)

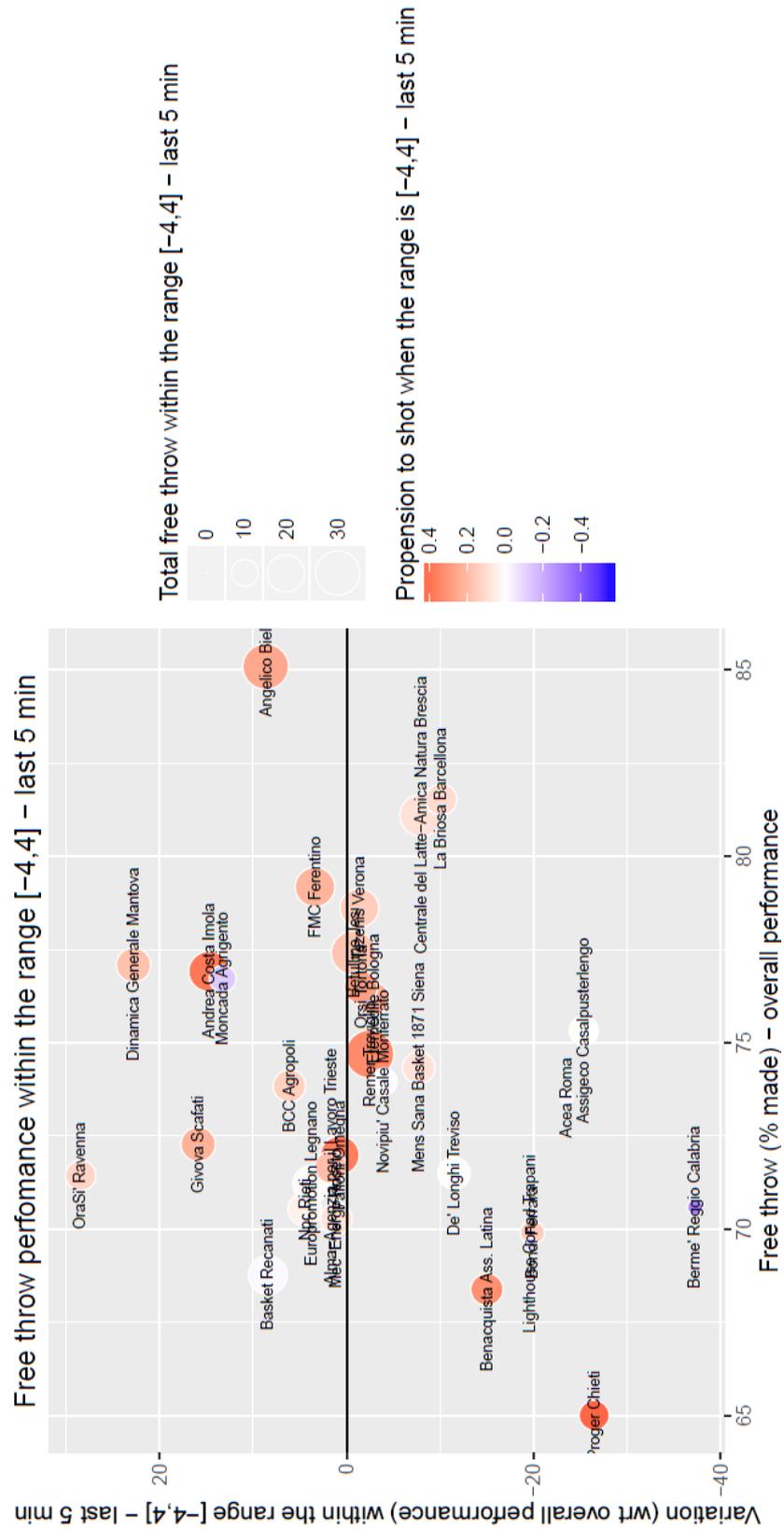


Figura 5.9 – Scatterplot: tiro libero all'interno della forbice-ultimi 5 min (andata)

6 Analisi sul girone di ritorno

6.1 Variazioni dell'albero di classificazione

Al termine delle analisi sul girone di andata, è stato costruito l'albero di classificazione sul girone di ritorno, riportato di seguito in figura 6.1.

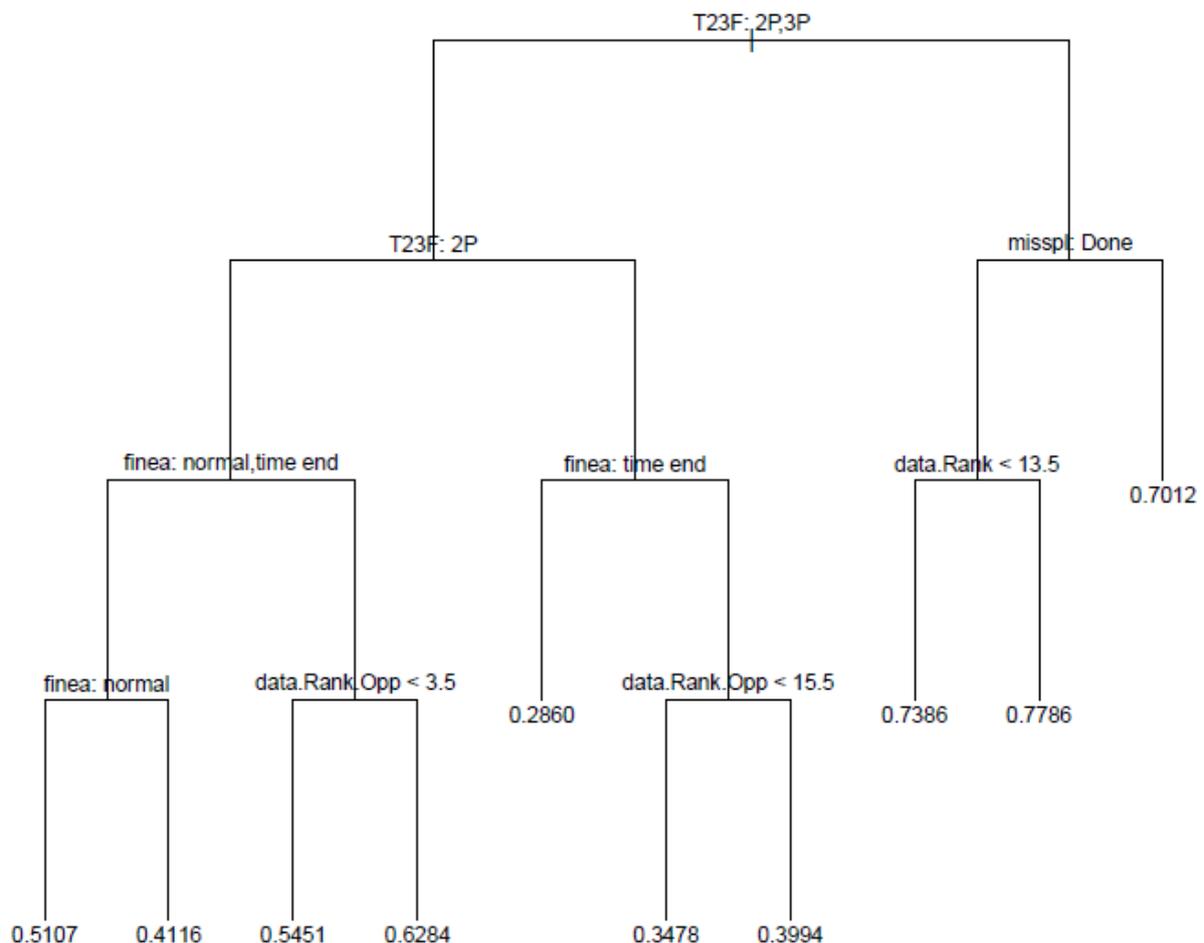


Figura 6.1 – Albero di classificazione (ritorno)

Lo si confronti con l'albero di classificazione relativo al girone di andata, riportato in figura 4.1; si possono notare alcune lievi differenze, ma tutto sommato la struttura generale è molto simile. In particolare:

- la bontà del modello, misurata tramite l'indice AUC (Area Under the ROC Curve), è pari a 0,6614 ossia sullo stesso livello della bontà dell'albero precedente (0,6649);

- il primo fattore di discriminazione rimane il tipo di tiro, com'è logico che sia, seguito dal cronometro dei 24 secondi per i tiri dal campo;
- per i tiri liberi, il fattore che più incide è l'esito del tiro precedente: se è stato sbagliato, la probabilità stimata di segnare è del 70,12%. Se, viceversa, è stato segnato, allora entra in gioco la posizione in classifica: fino al 13° posto la probabilità stimata è del 73,86%, per le ultime 3 squadre la probabilità stimata è del 77,86%. Leggermente diverso l'albero sul girone di andata, nel quale la posizione in classifica rappresentava il primo fattore di discriminazione e l'esito del tiro precedente era rilevante solo per le ultime 2 squadre in classifica;
- per i tiri da 3 punti, un ruolo importante è svolto ancora dai secondi di possesso palla. La probabilità stimata di fare canestro negli ultimi secondi dell'azione è pari a 28,6% (nel girone di andata era pari a 29,56%), mentre negli altri casi entra in gioco la posizione in classifica dell'avversario (nel girone di andata, invece, la posizione in classifica della propria squadra). Se si affronta l'ultima squadra in classifica, la probabilità stimata è quasi 40%, altrimenti è pari a 34,78%;
- per i tiri da 2 punti, quando ci si trova a fine azione, la probabilità stimata è 41,16%. Se ci si trova, invece, nei primi secondi dell'azione, allora entra in gioco la posizione in classifica degli avversari: contro le prime 3 classificate, la probabilità stimata è 54,51%, mentre sale al 62,84% contro le altre squadre. In caso di tiro effettuato nei secondi centrali di possesso palla, la probabilità stimata è pari a 51,07%. Leggermente diversa la situazione nel girone di andata, dove la posizione in classifica degli avversari influenzava il tiro negli ultimi secondi dell'azione.

Rispetto alla situazione illustrata al termine del girone di andata (in figura 4.1), è logico pensare che ci siano delle differenze: cambiamo le performance delle squadre, senza contare che alcune di esse hanno cambiato diversi giocatori in seguito al mercato invernale, cosa che verosimilmente può aver portato a variazioni di schemi e strategie in alcuni team. Inoltre, da un punto di vista statistico, gli alberi di classificazione sono molto sensibili ai cambiamenti nei dati: si parla, a tal proposito, di "instabilità" degli alberi. E, in questo caso, si tratta di dati completamente diversi: le partite del girone di andata sono totalmente diverse dalle partite del girone di ritorno (fatta eccezione, ovviamente, per le squadre coinvolte). Il fatto, quindi, di aver ottenuto due alberi molto simili tra loro è un risultato importante che conferisce più veridicità al modello elaborato, più robustezza alle stime elaborate e, in generale, più valore all'analisi svolta.

6.2 Variazioni degli indici e movimenti di classifica

In questa sezione vengono riportate le analisi condotte sul solo girone di ritorno con il fine di individuare le squadre che hanno manifestato miglioramenti o peggioramenti significativi nelle performance di tiro rispetto al girone di andata ed osservare se ad essi corrispondano effettivi cambiamenti significativi in classifica.

Si osservi lo scatterplot costruito con le performance di tiro in figura 6.2 e lo si confronti con la figura 5.1, riportata insieme per comodità di lettura. Si confrontino, inoltre, le classifiche dei due gironi al termine della stagione regolare:

GIRONE EST

1. De' Longhi Treviso
2. Centrale del Latte – Amica Natura Brescia
3. Dinamica Generale Mantova
4. Andrea Costa Imola
5. Mec Energy Roseto
6. Alma Agenzia per il Lavoro Trieste
7. Eternedile Bologna
8. Tezenis Verona
9. OraSi Ravenna
10. Remer Treviglio
11. Bondi Ferrara
12. Proger Chieti
13. Europromotion Legnano
14. Betulline Jesi
15. Basket Recanati
16. Bawer Matera

GIRONE OVEST

1. Givova Scafati
2. BCC Agropoli
3. Orsi Tortona
4. FMC Ferentino
5. Mens Sana Basket 1871 Siena
6. Moncada Agrigento
7. LightHouse Conad Trapani
8. Novipiù Casale Monferrato
9. Bermè Reggio Calabria
10. Angelico Biella
11. Assigeco Casalpusterlengo
12. Benacquista Ass. Latina
13. Npc Rieti
14. Acea Roma
15. Paffoni Omegna
16. La Briosa Barcellona

Dal confronto di grafici e classifiche, emergono alcune osservazioni interessanti. Innanzitutto, Verona mostra un peggioramento nella performance ai tiri liberi, a sostanziale parità delle altre performance, e, in effetti, perde qualche posizione in classifica. Brescia mantiene pressoché inalterata la performance da 2 punti, migliora nel tiro da 3 punti e peggiora nei tiri liberi: perde una sola posizione in classifica. Viceversa, Scafati migliora nel tiro da 2 e peggiora in quello da 3

(mantenendo inalterata la performance ai tiri liberi) e guadagna una posizione. Spicca il netto miglioramento in tutte le performance da parte di Ferentino, che compie un vero e proprio balzo in classifica dal 10° al 4° posto. Da segnalare anche il miglioramento nei tiri dal campo da parte di Biella che, pur peggiorando leggermente la performance (comunque positiva) nei tiri liberi, sale dal 15° al 10° posto. Da notare, infine, i miglioramenti in alcune performance e in classifica di altre squadre. Ad esempio, Siena migliora sensibilmente la performance da 3 punti, mantenendo pressoché invariata la performance da 2 punti e ai tiri liberi: in classifica guadagna 4 posizioni. Stessa cosa per Treviso, che guadagna 2 posizioni chiudendo al primo posto nel girone est. Agropoli migliora da 3 punti e ai liberi e cala leggermente da 2 punti, attestandosi intorno alla media (cioè verso l'origine degli assi): guadagna ben 4 posizioni e chiude al secondo posto il girone ovest. Anche Bologna mantiene performance medie, migliorando da 2 punti e peggiorando lievemente da 3 punti; tuttavia, conferma la sua posizione in classifica. Roseto migliora le sue performance nei tiri da 2 e 3 punti, cosa che porta la squadra a scalare un gradino nella classifica del girone est. Imola mantiene pressoché invariate le performance dal campo e peggiora nei tiri liberi, scendendo di 2 posti in classifica. Curioso notare come Imola e Matera siano rimaste sempre nel quadrante peggiore (addirittura Matera ha una performance migliore nei tiri liberi), ma la prima si è qualificata per i play-off mentre la seconda è retrocessa al termine del campionato.

Come si può osservare, ci sono casi in cui ad un miglioramento nelle performance di tiro corrisponde un miglioramento della posizione in classifica e casi in cui non sembra esserci questa corrispondenza. Si può, quindi, confermare quanto già evidenziato in precedenza, ossia come effettivamente la performance nel tiro spieghi buona parte del gioco e giustifichi parte dei movimenti di classifica; tuttavia, non spiega tutto il gioco. Come si è visto, alcune variazioni di classifica non corrispondono ad effettive variazioni nelle performance: significa che ci sono altri aspetti del gioco (ad esempio, la difesa), non analizzati in questo lavoro, che incidono sulla performance generale, intesa in senso più ampio, della squadra e che, se analizzati a fondo, possono integrare queste analisi in modo da dare una visione più completa ed ancora più approfondita.

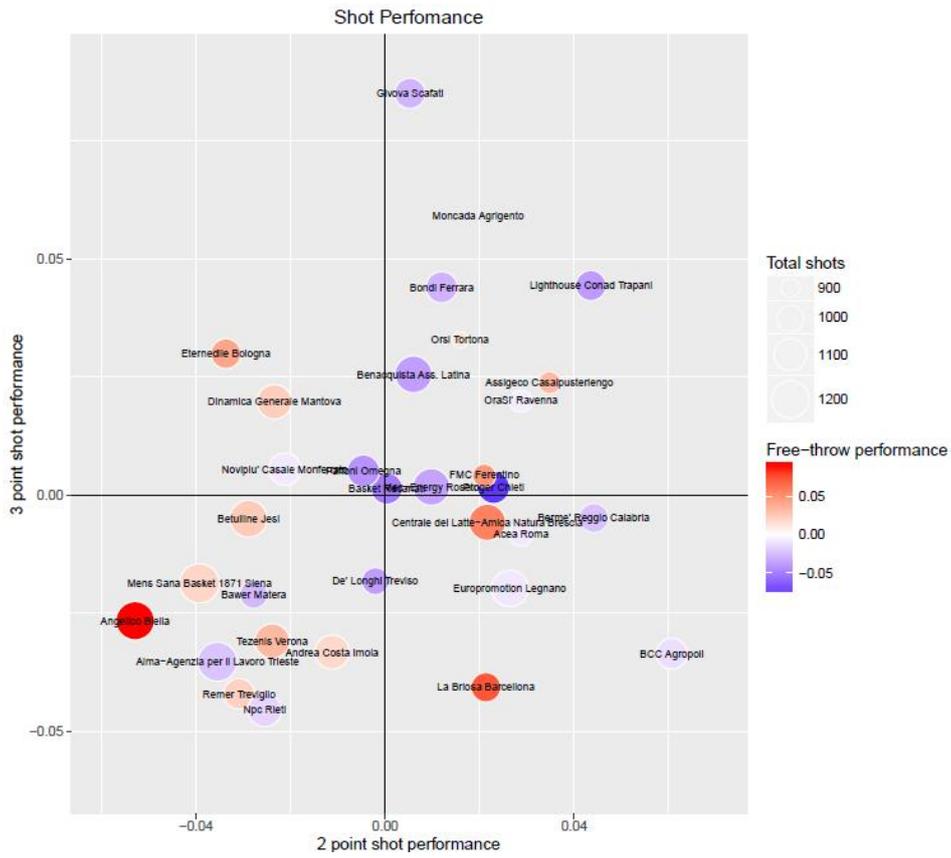


Figura 5.1 – Scatterplot: performance di tiro (andata)

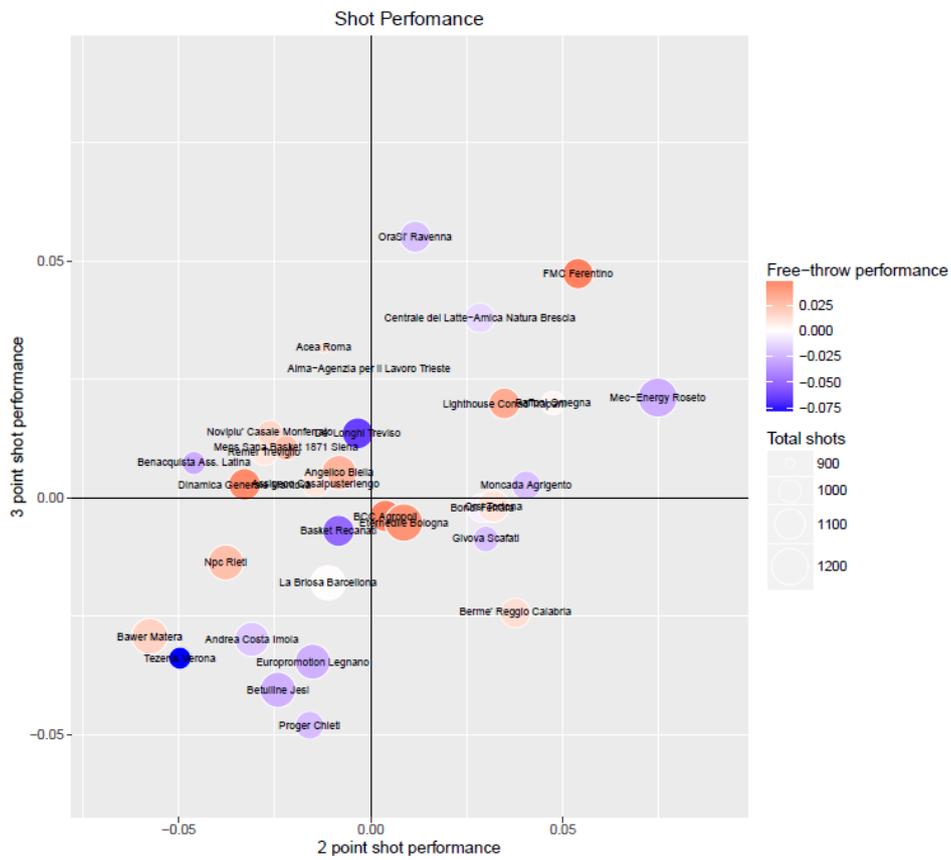


Figura 6.2 – Scatterplot: performance di tiro (ritorno)

Proseguendo con l'analisi, si passa ora alle situazioni di stress. I seguenti grafici in figura 6.3 e 6.4 rappresentano la performance di tiro negli ultimi 2 secondi di possesso nel girone di ritorno. Anche in questo caso, come nel girone di andata (figura 5.2 e 5.3), sono poche le squadre che mostrano la tendenza a migliorare la performance di tiro in questa situazione.

Le squadre che hanno manifestato miglioramenti nel girone di andata (Recanati e Agropoli nei tiri da 2 punti; Ferentino, Tortona, Ravenna, Imola, Siena e Treviso nei tiri da 3 punti) non si sono riconfermate nel girone di ritorno, nel quale sono emerse, sempre in maniera contenuta, Jesi nel tiro da 2 e Treviglio, Recanati, Reggio Calabria e Matera nel tiro da 3.

Nel tiro da 2 punti, Verona migliora leggermente la performance in questa situazione, comunque negativa, ma peggiora la percentuale di realizzazione totale. Situazione opposta per Brescia, che migliora leggermente in entrambe le statistiche; così come Roseto, che migliora sensibilmente la percentuale di realizzazione, ma cala nella precisione in questa situazione (il pallino è diventato più bianco).

Nel tiro da 3 punti, Ferentino e Ravenna (ma anche Brescia) hanno avuto un calo nel differenziale di performance in questa situazione, ma hanno aumentato la percentuale di realizzazione totale. Verona, invece, fa registrare un leggero calo di entrambi gli indicatori.

In generale, valgono le stesse considerazioni espresse con riferimento alle figure 5.2 e 5.3 nel capitolo precedente. Sembra, infatti, non esserci una relazione tra l'andamento in classifica e la performance di tiro negli ultimi 2 secondi, poiché si tratta maggiormente di tiri effettuati in condizioni non ottimali, di forte pressione difensiva e, molto spesso, lontano da canestro. Potrebbe, comunque, essere interessante disporre di questo genere di informazioni per eventuali accorgimenti difensivi esposti in precedenza, con riferimento alle figure del capitolo precedente.

I seguenti grafici (figure 6.5, 6.6 e 6.7) mostrano le varie performance quando il divario di punteggio è all'interno della forbice [-4; 4] e possono essere confrontati con i rispettivi grafici del girone di andata (5.4, 5.5 e 5.6).

Per quanto riguarda il tiro da 2 punti, Ferrara e Trapani migliorano sia la performance in questa situazione che la percentuale di realizzazione generale (Ferrara mostra anche una propensione al tiro maggiore rispetto al girone di andata). Ferentino migliora leggermente la percentuale di realizzazione, così come Brescia che aumenta anche la propensione al tiro. Agropoli rimane in una buona posizione, nonostante il calo delle percentuali al tiro. Siena e Treviglio aumentano la propensione al tiro e la percentuale di realizzazione, ma calano nel differenziale di performance in questa situazione. Non deve trarre in inganno il miglioramento di performance, in caso di punteggio in bilico, da parte di Verona: la percentuale di realizzazione generale, infatti, è inferiore a quella mantenuta nel girone di andata, così come la propensione al tiro.

Per quanto riguarda il tiro da 3 punti, le migliori squadre del girone di andata non si confermano: Scafati, Trapani e Mantova mostrano un calo di percentuale di realizzazione e del differenziale di performance (anche se per Scafati rimane positivo). Spiccano i netti miglioramenti sia nella percentuale di realizzazione, che nel differenziale di performance da parte di Ferentino, Siena e Biella (per queste ultime, aumento anche della propensione al tiro). Migliora leggermente Brescia, soprattutto nella propensione e nelle percentuali al tiro, anche se il differenziale rimane negativo. Legnano migliora sensibilmente il differenziale di performance, che raggiunge il livello di Siena e Biella, tuttavia è accompagnato da un calo di percentuale di realizzazione che si attesta tra i più bassi in assoluto.

Per quanto riguarda i tiri liberi, si nota innanzitutto un generale calo di percentuale di realizzazione rispetto al girone di andata. Biella e Roma si confermano tra le squadre migliori in questa situazione. Trieste e Treviglio aumentano la propensione al tiro e anche il differenziale, mentre Ferentino non riesce a mantenere la performance tenuta nel girone di andata. Stessa cosa per Brescia, Siena, Imola e Ferrara: pur avendo alte propensioni, non riescono a sfruttare a loro vantaggio questa situazione come si intuisce dal peggioramento della performance. Si può, inoltre, notare il netto peggioramento di Verona: oltre ad un calo di percentuale di realizzazione generale, risulta evidente il peggioramento in questa precisa situazione di stress.

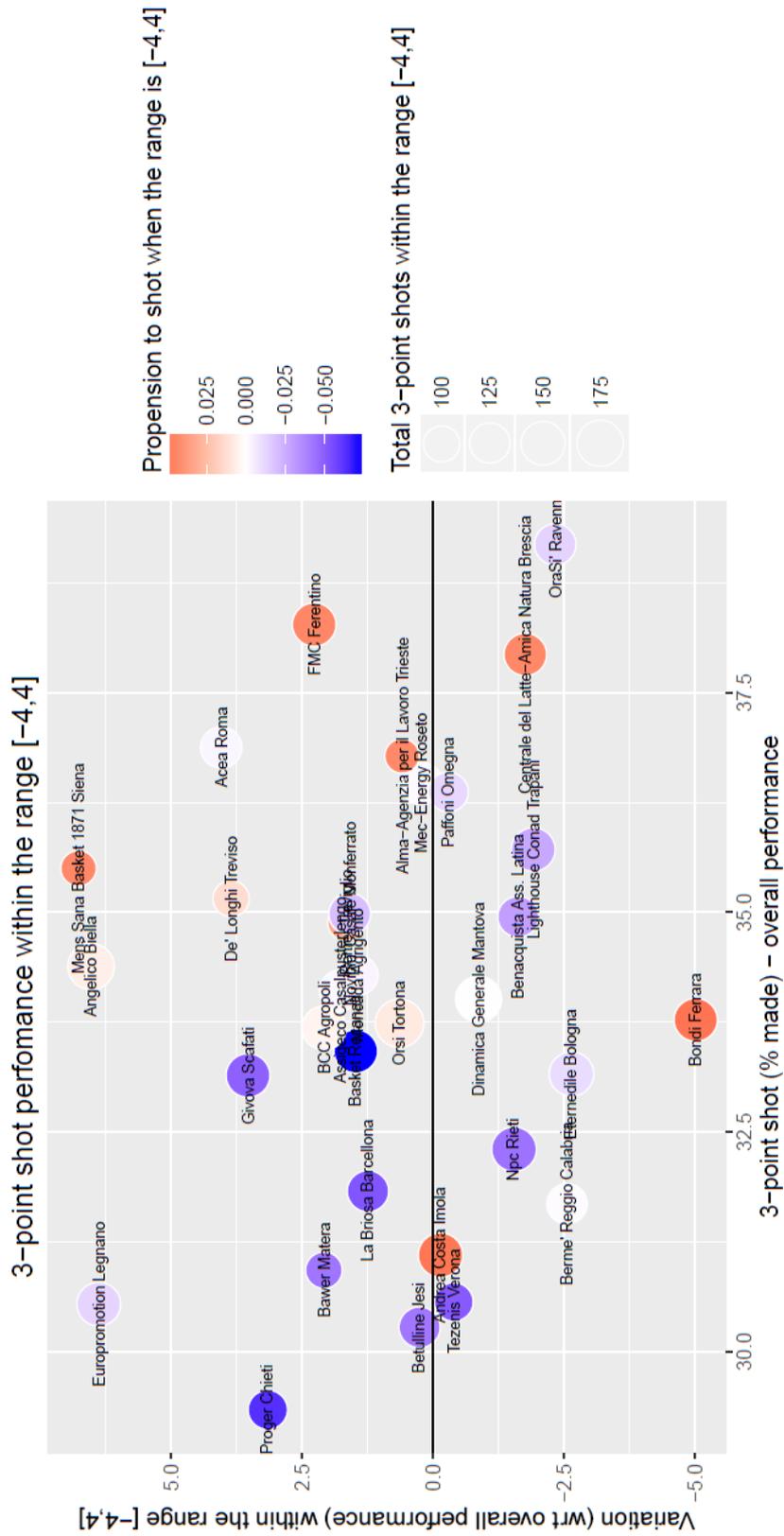


Figura 6.6 – Scatterplot: tiro da 3 pt all'interno della forbice (ritorno)

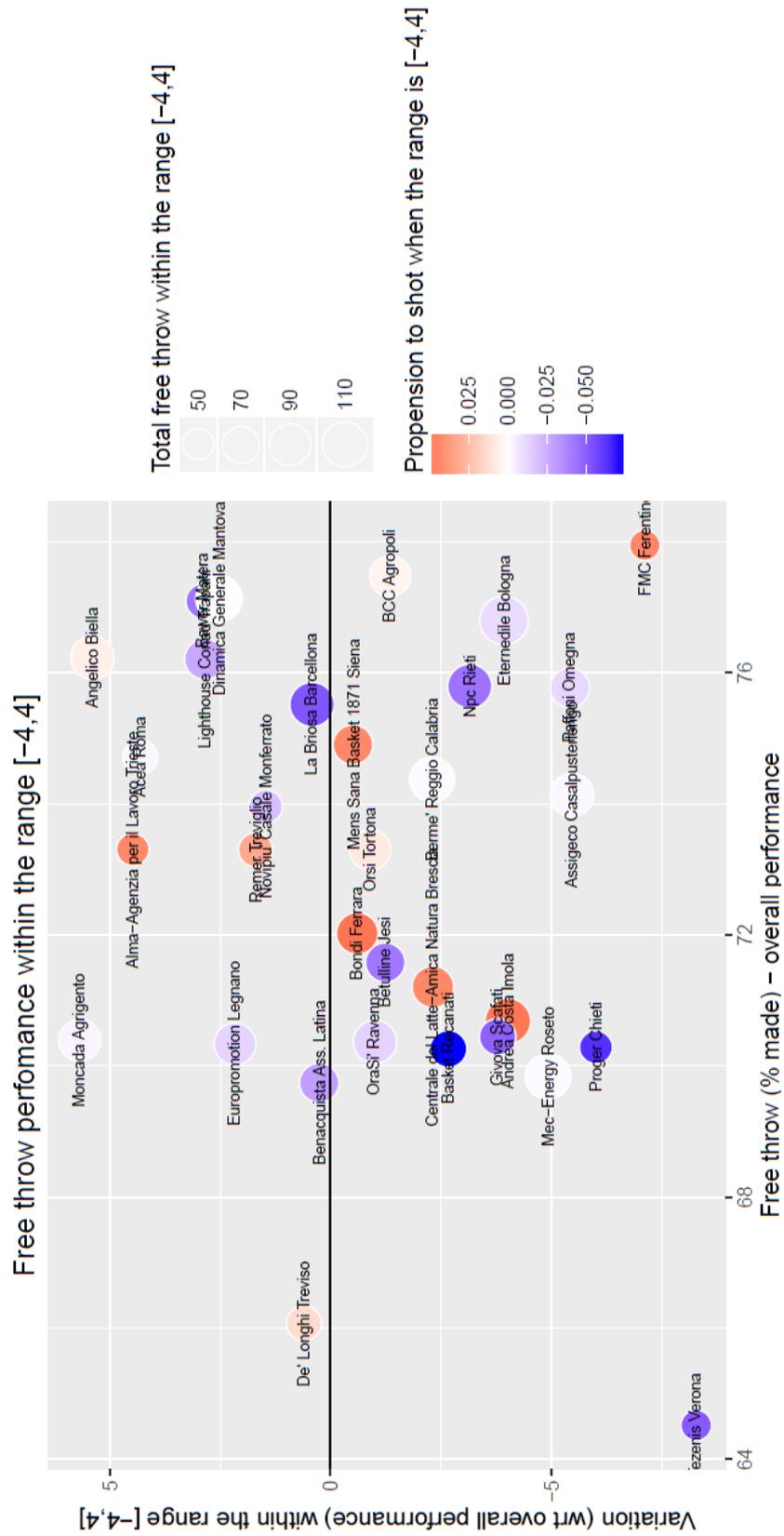


Figura 6.7 – Scatterplot: tiro libero all'interno della forbice (ritorno)

Approfondendo, ora, la situazione della forbice [-4; 4] negli ultimi 5 minuti (tenendo conto dell'esiguo numero di osservazioni), si osservino i grafici in figura 6.8, 6.9 e 6.10 e li si mettano a confronto con le figure 5.7, 5.8 e 5.9.

Si nota, per prima cosa, come l'apparente miglioramento di Verona nella performance da 2 punti mostrata nella figura 6.3 venga ridimensionato: nel momento chiave della partita, la squadra mostra un peggioramento di performance. Il miglioramento della performance nel tiro da 3 punti sembra non compensare tale peggioramento, dal momento che la percentuale di realizzazione generale è inferiore rispetto al girone di andata (come pure la propensione al tiro). Ancora più marcato, inoltre, risulta il peggioramento nei tiri liberi.

Curioso, invece, come anche il peggioramento di performance generale nei tiri liberi di Brescia (figura 6.1) sia ridimensionato: nel momento chiave della partita, la squadra riesce a mantenere una performance positiva.

Nei tiri da 2 punti, Agropoli migliora differenziale e propensione, ma peggiora la percentuale di realizzazione; Casale Monferrato migliora il differenziale di performance, a sostanziale parità degli altri indicatori, mentre Trieste migliora nel differenziale anche se cala la percentuale di realizzazione. Brescia e Tortona migliorano la loro performance, mentre Treviglio peggiora drasticamente. Da notare come l'alta propensione al tiro di Reggio Calabria (maggiore rispetto all'andata) non si traduca necessariamente in maggiore efficienza: la percentuale di realizzazione è rimasta praticamente uguale, ma il differenziale è calato e ora è leggermente negativo.

Nei tiri da 3 punti, evidente il peggioramento di Scafati e la diminuzione della percentuale di realizzazione e del differenziale di performance da parte di Jesi. Per contro, evidenti i miglioramenti di Biella, Siena, Ravenna e Imola. Parziali miglioramenti anche da parte di Tortona e Chieti, anche se per loro diminuisce la percentuale di realizzazione. Controverse le situazioni di Recanati e Roma: la prima manifesta un miglioramento nel differenziale di performance e un calo nella propensione al tiro, mentre la seconda mostra un drastico calo del differenziale che non consente di sfruttare l'aumento della percentuale di realizzazione.

Nei tiri liberi, come già evidenziato in precedenza, si nota un generale calo delle percentuali di realizzazione. Detto questo, Biella si conferma una delle squadre con le performance migliori. Sensibili miglioramenti da parte di Siena, Trieste e Matera (anche se per queste ultime due si notino l'esiguo numero di tiri e la bassa propensione). Ambigua la situazione di Treviso: aumento della propensione, aumento del differenziale di performance, ma calo della percentuale di realizzazione. Netto, infine, il miglioramento di Reggio Calabria: più tiri, maggiore propensione e aumento di differenziale e di percentuale di realizzazione.

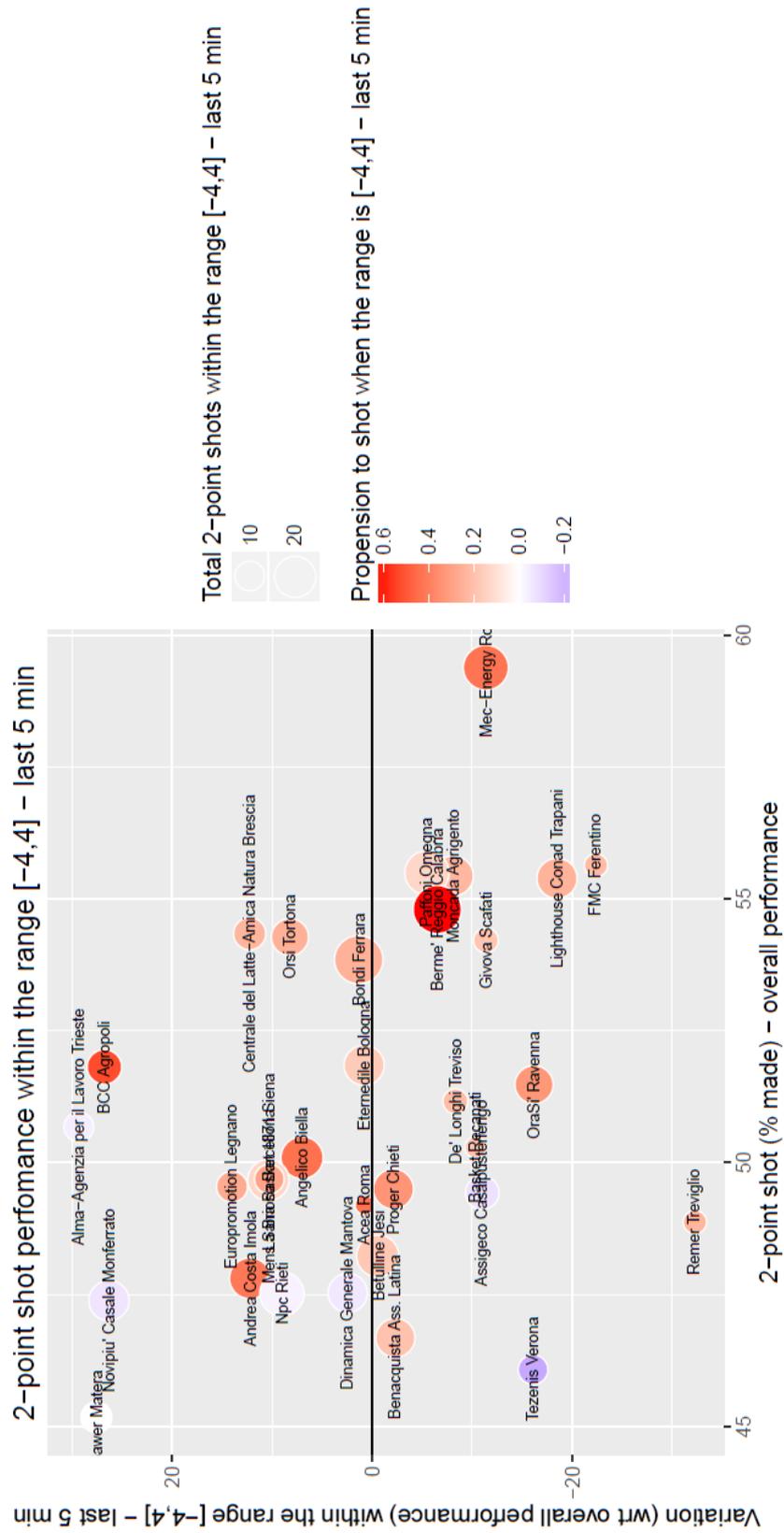


Figura 6.8 – Scatterplot: tiro da 2 pt all'interno della forbice-ultimi 5 min (ritorno)

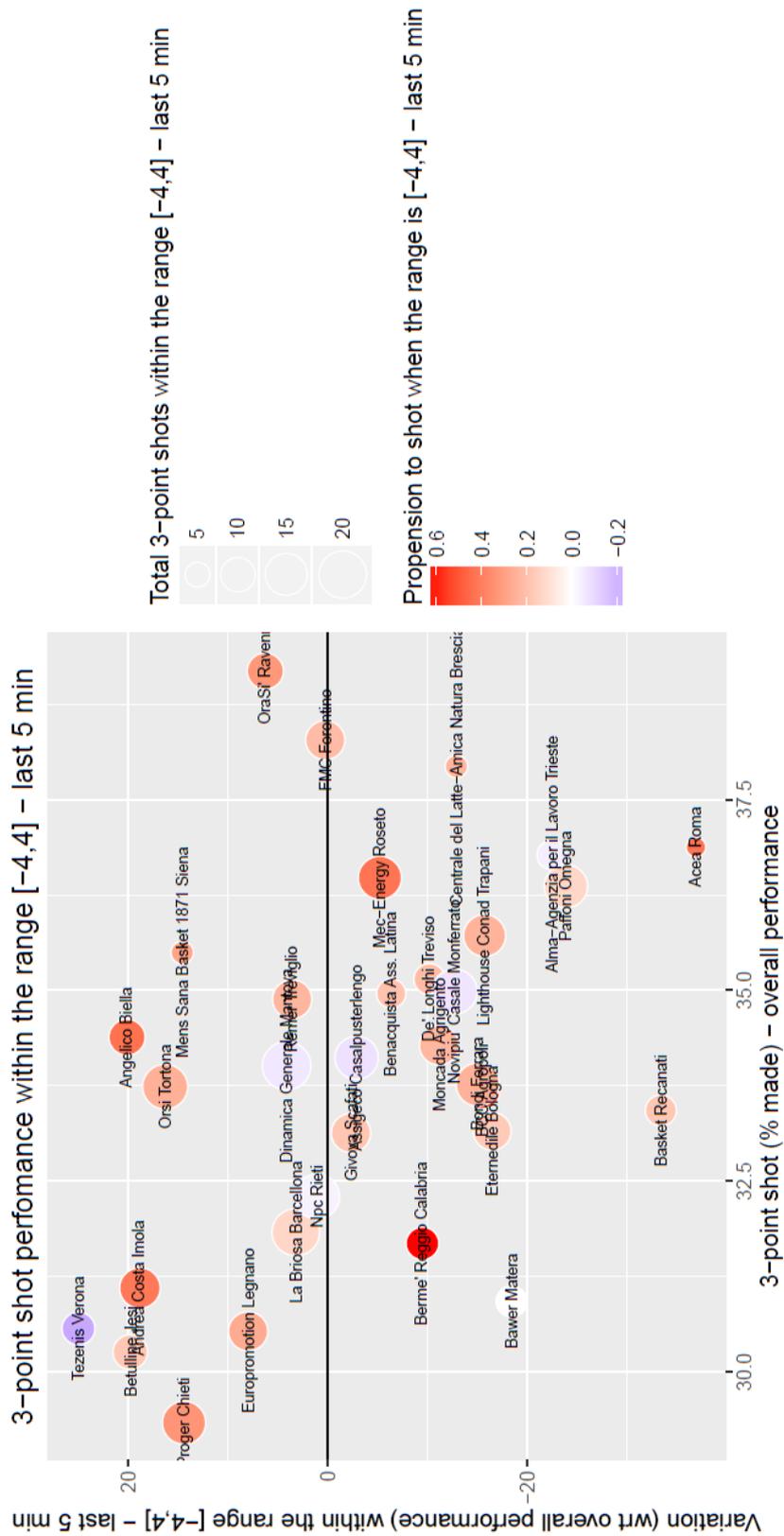


Figura 6.9 – Scatterplot: tiro da 3 pt all'interno della forbice-ultimi 5 min (ritorno)

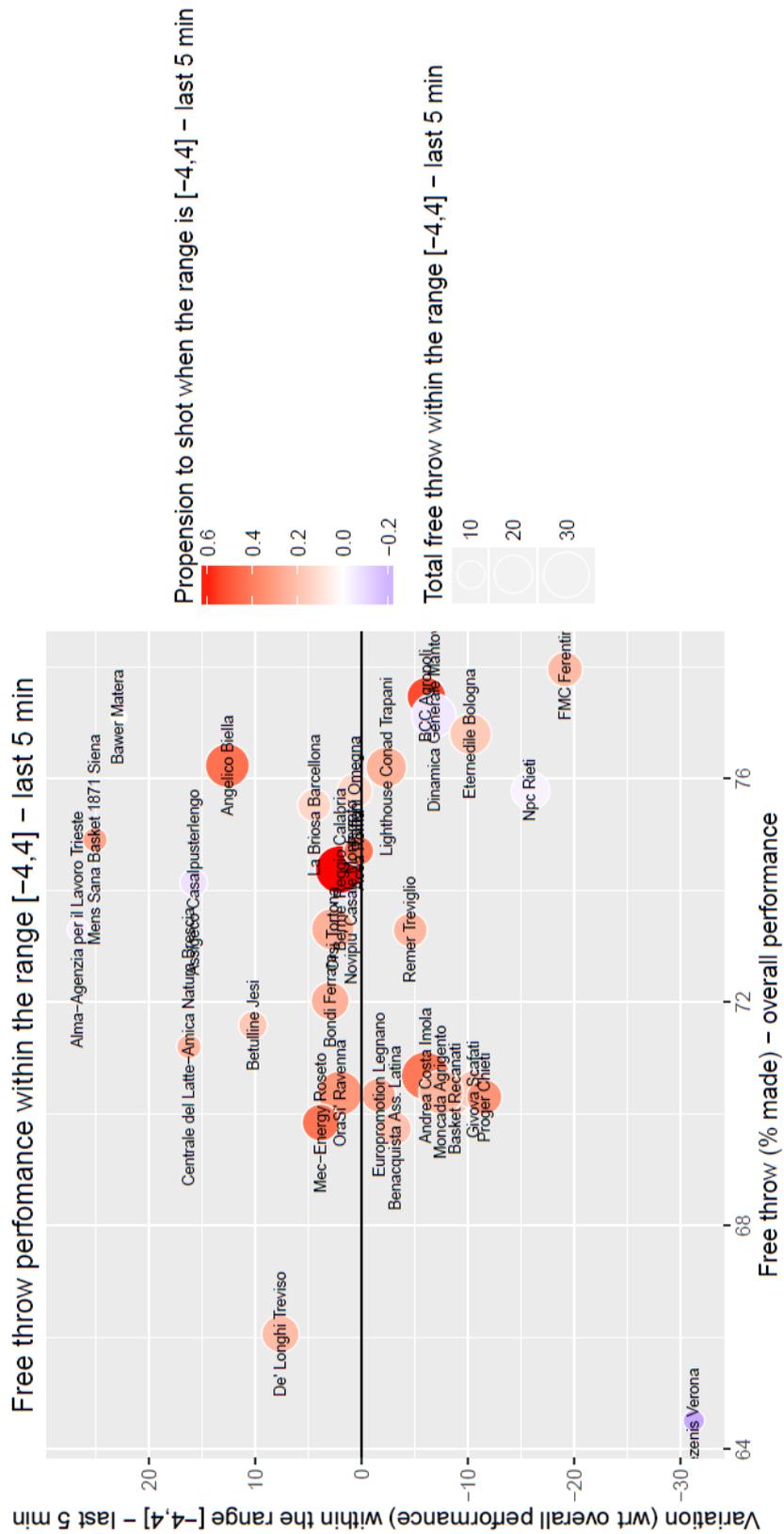


Figura 6.10 – Scatterplot: tiro libero all'interno della forbice-ultimi 5 min (ritorno)

7 Conclusioni

Il punto di partenza per tutto il lavoro realizzato è costituito dalle analisi svolte per la Scaligera Basket Verona, sotto l'attenta guida ed i preziosi consigli del coach Marco Crespi.

Analizzando i primi play-by-play relativi alle partite giocate da Verona, si è intravista la possibilità di condurre una serie di analisi più approfondite e su più larga scala. Si è deciso, così, di ampliare la ricerca a tutti i dati disponibili dell'intero campionato di serie A2, ottenendo dei risultati interessanti.

Dal momento che le analisi svolte per Verona riguardavano principalmente la distribuzione dei tiri nell'arco di un'azione da 24 secondi e l'andamento del divario di punteggio tra le squadre, ci si è concentrati su questi aspetti: l'efficienza offensiva, intesa come performance al tiro. Dopo aver individuato alcune situazioni inerenti al gioco che possono generare stress o pressione, tramite l'algoritmo CART è stato stimato il loro possibile impatto sulla probabilità di fare canestro. L'albero di classificazione risultante da tale algoritmo comprende i dati riguardanti le partite del girone di andata; in questo modo si può effettuare una prima analisi della situazione a metà campionato e, successivamente, si può condurre un'analisi dinamica considerando il solo girone di ritorno per poi confrontare i risultati ottenuti. L'albero così ottenuto ha permesso di elaborare alcuni nuovi indici di tiro che possono integrare le normali statistiche di base di largo uso nel basket. Tali nuovi indicatori sono:

- performance di tiro, capace di dare a ciascun canestro (o errore) il suo peso specifico in relazione alla situazione in cui viene effettuato il tiro;
- differenziale di performance, capace di cogliere il miglioramento o peggioramento della performance di tiro in determinate situazioni;
- propensione al tiro, che mette in relazione il numero di tiri effettuati in una determinata situazione ed il numero di minuti giocati in quella specifica situazione.

Queste nuove misurazioni permettono di stabilire la reale percentuale di realizzazione in diverse situazioni di gioco, di quantificare e rappresentare graficamente l'effettivo miglioramento o peggioramento rispetto alla performance generale. È stata condotta un'analisi approfondita delle varie situazioni sul girone di andata per poter individuare le squadre che hanno tenuto performance migliori e osservare se fossero le stesse che occupano i primi posti in classifica. Si è osservato che non sempre si manifesta questa corrispondenza. In seguito, è stato preso in considerazione il solo girone di ritorno e, effettuando le stesse analisi, si è notato come i nuovi indicatori spieghino e giustifichino, almeno in parte, alcuni movimenti di classifica.

Occorre, tuttavia, tenere presente che è necessario un numero elevato di osservazioni per poter ottenere stime robuste e, di conseguenza, indici significativi. Per questo motivo, il lavoro è stato condotto a livello di squadra.

Infine è stato elaborato, tramite il medesimo algoritmo CART utilizzato in precedenza, l'albero di classificazione sul girone di ritorno. Confrontandolo con quello ottenuto al termine del girone di andata, si è osservato come ci siano alcune differenze lievi ma, sostanzialmente, presentino una struttura molto simile. Si tratta di un risultato particolarmente importante: cambiando completamente i dati, si sono ottenuti due alberi molto simili tra loro. Ciò conferisce maggior credibilità al modello, alle stime e agli indici elaborati, andando ad accrescere il valore del lavoro svolto.

Ipotizzando di integrare le analisi preliminari, svolte nella prima parte, con un'analisi di questo tipo (sfruttando i big data e le tecniche di machine learning), basata sui nuovi indici di tiro, e con altre possibili analisi su altri aspetti del gioco (ad esempio, le situazioni difensive), si potrebbe tracciare un profilo preciso e approfondito di ciascuna squadra di interesse. Oltre ad essere uno scenario appetibile per gli esperti di statistica, potrebbe costituire un motivo di interesse pratico per le squadre, che potrebbero analizzare meglio il proprio profilo per individuare possibili miglioramenti e che potrebbero analizzare gli avversari per preparare al meglio le strategie di gioco. Inoltre, ipotizzando di avere sufficienti dati a disposizione per ogni giocatore, tali analisi potrebbero essere condotte su ciascun giocatore di interesse. E disporre di un profilo specifico, approfondito e preciso per ogni giocatore sarebbe una grande risorsa per qualsiasi squadra. Oltre ad essere utilizzabili in fase di preparazione delle strategie di gioco per una partita, infatti, potrebbero anche essere utilizzati per effettuare confronti tra i giocatori stessi durante la fase del mercato: analizzare il profilo di un cestista, al di là delle semplici informazioni di base, confrontare giocatori diversi e valutarne il rendimento, quantificarne il contributo alla performance di squadra per poter individuare quelli che meglio si adattano all'idea di gioco e alle preferenze del coach.

Si è visto, quindi, come la statistica applicata al gioco (in questo caso, ad una parte di esso) riesca a spiegarne molte situazioni, ad offrire un supporto reale alle loro interpretazioni e a rendere oggettivamente quantificabili eventuali sensazioni percepite durante lo svolgimento del gioco stesso. Gli aspetti presentati in questo lavoro, uniti all'enorme quantità di dati disponibili per questo sport e ai molteplici strumenti statistici da adottare potrebbero aprire nuovi orizzonti nell'analisi dei dati nel basket.

Infatti è molto importante disporre di dati approfonditi; ma ancora più importante è il sapere sfruttare i dati a disposizione in coerenza con gli obiettivi prefissati. Ci sono molti aspetti, nel

gioco, che vengono tralasciati o sottovalutati o considerati solo in via superficiale; riuscire a capire che la statistica possa approfondire tali aspetti, offrire nuove chiavi di interpretazione del gioco e aiutare nella sua comprensione potrebbe portare a sensibili miglioramenti nella generale concezione della pallacanestro, spesso offuscata da pregiudizi e preconcetti. Elaborare ed adottare strategie ottimali basate su dati, indici e misure oggettivamente visibili e verificabili consente di giungere, nel medio – lungo periodo, al successo.

Indice tabelle

Tabella 2.1 – Principali caratteristiche del dataset	p. 10
Tabella 2.2 – Partite escluse dall'analisi	p. 13 – 14
Tabella 4.1 – Conversione delle covariate numeriche in categoriali	p. 46

Indice figure

Figura 2.1 – Esempio di play-by-play	p. 8
Figura 2.2 – Andamento del margine	p. 16
Figura 2.3 – Tempo di permanenza nelle fasce di margine	p. 17
Figura 2.4 – Andamento del margine nei quarti	p. 18
Figura 2.5 – Eventi interni alle azioni da 24 secondi	p. 19
Figura 2.6 – Eventi interni alle azioni da 14 secondi	p. 20
Figura 2.7 – Tiri da 2 punti nei 24 secondi	p. 23
Figura 2.8 – Tiri da 3 punti nei 24 secondi	p. 23
Figura 2.9 – Tiri da 2 punti nei 14 secondi	p. 24
Figura 2.10 – Tiri da 3 punti nei 14 secondi	p. 24
Figura 2.11 – Report di Verona nelle fasce di margine (intera partita)	p. 26
Figura 2.12 – Report di Verona nelle fasce di margine (ultimo quarto)	p. 27
Figura 2.13 – Report di Lorenzo Saccaggi nelle fasce di margine (intera partita)	p. 28
Figura 2.14 – Report di Lorenzo Saccaggi nelle fasce di margine (ultimo quarto)	p. 29
Figura 2.15 – Andamento del margine (Brescia)	p. 30
Figura 2.16 – Tempo di permanenza nelle fasce di margine (Brescia)	p. 31
Figura 2.17 – Andamento del margine nei quarti (Brescia)	p. 32
Figura 2.18 – Eventi interni alle azioni da 24 secondi (Brescia)	p. 33
Figura 2.19 – Eventi interni alle azioni da 14 secondi (Brescia)	p. 34
Figura 2.20 – Tiri da 2 punti nei 24 secondi (Brescia)	p. 35
Figura 2.21 – Tiri da 3 punti nei 24 secondi (Brescia)	p. 35
Figura 2.22 – Tiri da 2 punti nei 14 secondi (Brescia)	p. 36
Figura 2.23 – Tiri da 3 punti nei 14 secondi (Brescia)	p. 36
Figura 2.24 – Report di Brescia nelle fasce di margine (intera partita)	p. 37
Figura 2.25 – Report di Brescia nelle fasce di margine (ultimo quarto)	p. 38

Figura 2.26 – Report di Mirza Alibegovic nelle fasce di margine (intera partita)	p. 39
Figura 2.27 – Report di Mirza Alibegovic nelle fasce di margine (ultimo quarto)	p. 40
Figura 3.1 – Esempio di ROC Curve	p. 44
Figura 4.1 – Albero di classificazione (andata)	p. 47
Figura 5.1 – Scatterplot: performance di tiro (andata)	p. 53 – 70
Figura 5.2 – Scatterplot: tiro da 2 pt negli ultimi 2 secondi (andata)	p. 55
Figura 5.3 – Scatterplot: tiro da 3 pt negli ultimi 2 secondi (andata)	p. 56
Figura 5.4 – Scatterplot: tiro da 2 pt all'interno della forbice (andata)	p. 59
Figura 5.5 – Scatterplot: tiro da 3 pt all'interno della forbice (andata)	p. 60
Figura 5.6 – Scatterplot: tiro libero all'interno della forbice (andata)	p. 61
Figura 5.7 – Scatterplot: tiro da 2 pt all'interno della forbice-ultimi 5 min (andata)	p. 63
Figura 5.8 – Scatterplot: tiro da 3 pt all'interno della forbice-ultimi 5 min (andata)	p. 64
Figura 5.9 – Scatterplot: tiro libero all'interno della forbice-ultimi 5 min (andata)	p. 65
Figura 6.1 – Albero di classificazione (ritorno)	p. 66
Figura 6.2 – Scatterplot: performance di tiro (ritorno)	p. 70
Figura 6.3 – Scatterplot: tiro da 2 pt negli ultimi 2 secondi (ritorno)	p. 72
Figura 6.4 – Scatterplot: tiro da 3 pt negli ultimi 2 secondi (ritorno)	p. 73
Figura 6.5 – Scatterplot: tiro da 2 pt all'interno della forbice (ritorno)	p. 75
Figura 6.6 – Scatterplot: tiro da 3 pt all'interno della forbice (ritorno)	p. 76
Figura 6.7 – Scatterplot: tiro libero all'interno della forbice (ritorno)	p. 77
Figura 6.8 – Scatterplot: tiro da 2 pt all'interno della forbice-ultimi 5 min (ritorno)	p. 79
Figura 6.9 – Scatterplot: tiro da 3 pt all'interno della forbice-ultimi 5 min (ritorno)	p. 80
Figura 6.10 – Scatterplot: tiro libero all'interno della forbice-ultimi 5 min (ritorno)	p. 81

Ringraziamenti

Per la realizzazione di questo lavoro, si ringrazia:

la prof.ssa Paola Zuccolotto, per l'opportunità di stage proposta, la disponibilità, la collaborazione e l'aiuto durante tutto lo svolgimento del lavoro;

la prof.ssa Marica Manisera, per la disponibilità, la collaborazione e l'aiuto;

il coach Marco Crespi, per la disponibilità mostrata, per la collaborazione, per l'esperienza ed i preziosi consigli;

il coach Michele Dall'Ora, per la collaborazione, per l'aiuto ed il supporto nelle analisi;

il dott. Andrea Sordelli, per la gestione amministrativa dello stage;

la società Scaligera Basket Verona, per l'opportunità concessa;

la società Lombardo s.r.l. e tutte le persone che ne fanno parte, per il supporto mostrato nei mesi trascorsi con loro;

i genitori, gli amici, i compagni del corso di laurea in marketing all'università di Brescia e tutte le persone che hanno mostrato il proprio supporto.

Progetto svolto in collaborazione con il Big&Open Data Innovation Laboratory (BODaI-Lab), Università degli Studi di Brescia, nell'ambito del progetto BDSports.

Riferimenti bibliografici

- Albert J., J. Bennet, and J. J. Cochran (2005): *Anthology of statistics in sports*, volume 16, SIAM.
- Albert J., and R. H. Koning (2007): *Statistical thinking in sports*, CRC Press.
- Annis D. H., et al. (2006): “Optimal end-game strategy in basketball”, *Journal of Quantitative Analysis in Sports*, 2, 1.
- Breiman L., J. Friedman, C. J. Stone, and R. A. Olshen (1984): *Classification and regression trees*, CRC Press.
- Brown M., J. Sokol, et al. (2010): “An improved lrmc method for NCAA basketball prediction”, *Journal of Quantitative Analysis in Sports*, 6, 1 – 23.
- Carpita M., M. Sandri, A. Simonetto, and P. Zuccolotto (2014): “Football mining with r”, in Y. Z. and Y. C., eds., *Data Mining Applications with R*, Elsevier, Capitolo 14.
- Cooper W. W., J. L. Ruiz, and I. Sirvent (2009): “Selecting non-zero weights to evaluate effectiveness of basketball players with dea”, *European journal of operational research*, 195, 563 – 574.
- Deshpande S. K., and S. T. Jensen (2016): “Estimating an NBA players impact on his team chances of winning”, *Journal of Quantitative Analysis in Sports*, 12, 51 – 72.
- Fearnhead P., and B. M. Taylor (2011): “On estimating the ability of NBA players”, *Journal of Quantitative Analysis in Sports*, 7.
- Fraire M., and A. Rizzi (2011): *Analisi dei dati per il Data Mining*, Carocci Editore.
- Gupta A. A. (2015): “A new approach to bracket prediction in the NCAA mens basketball tournament based on a dual-proportion likelihood”, *Journal of Quantitative Analysis in Sports*, 11, 53 – 67.
- Han J., J. Pei and M. Kamber (2011): *Data mining: concepts and techniques*, Elsevier.
- Hand D. J., H. Mannila, and P. Smyth (2011): *Principles of data mining (adaptive computation and machine learning)*, MIT Press.
- Kubatko J., D. Oliver, K. Pelton, D. T. Rosenbaum, et al. (2007): “A starting point for analyzing basketball statistics”, *Journal of Quantitative Analysis in Sports*, 3, 1 – 22.
- Loeffelholz B., E. Bednar, K. W. Bauer, et al. (2009): “Predicting NBA games using neural networks”, *Journal of Quantitative Analysis in Sports*, 5, 1 – 15.
- Lopez M. J., and G. J. Matthews (2015): “Building an NCAA mens basketball predictive model and quantifying its success”, *Journal of Quantitative Analysis in Sports*, 11, 5 – 12.

- Madden C. C., R. J. Kirkby, D. McDonald, J. J. Summers, D. F. Brown, and N. J. King (1995): “Stressfull situations in competitive basketball”, *Australian Psychologist*, 30, 119 – 124.
- Manner H. (2016): “Modeling and forecasting the outcomes of NBA basketball games”, *Journal of Quantitative Analysis in Sports*, 12, 31 – 41.
- Ozmen M. U. (2012): “Foreign player quota, experience and efficiency of basketball players”, *Journal of Quantitative Analysis in Sports*, 8.
- Page G. L., B. J. Barney, and A. T. McGuire (2013): “Effect of position, usage rate, and per game minutes played on NBA player production curves”, *Journal of Quantitative Analysis in Sports*, 9, 337 – 345.
- Page G. L., G. W. Fellingham, and C. S. Reese (2007): “Using box-scores to determine a position’s contribution to winning basketball games”, *Journal of Quantitative Analysis in Sports*, 3.
- Piette J., S. Anand, and K. Zhang (2010): “Scoring and shooting abilities of NBA players”, *Journal of Quantitative Analysis in Sports*, 6.
- Ruiz F. J., and F. Perez-Cruz (2015): “A generative model for predicting outcomes in college basketball”, *Journal of Quantitative Analysis in Sports*, 11, 39 – 52.
- Shorridge A., K. Goldsberry, and M. Adams (2014): “Creating space to shoot: quantifying spatial relative field goal efficiency in basketball”, *Journal of Quantitative Analysis in Sports*, 10, 303 – 313.
- Skinner B. (2010): “The price of anarchy in basketball”, *Journal of Quantitative Analysis in Sports*, 6.
- Taylor J. (1987): “Predicting athletic performance with self-confidence and somatic and cognitive anxiety as a function of motor and physiological requirements in six sports”, *Journal of personality*, 55, 139 – 153.
- West B. T. et al. (2008): “A simple and flexible rating method for predicting success in the NCAA basketball tournament: Updated results from 2007”, *Journal of Quantitative Analysis in Sports*, 4, 8.
- Yuan L.-H., A. Liu, A. Yeh, A. Kaufman, A. Reece, P. Bull, A. Franks, S. Wang, D. Illushin, and L. Bornn (2015): “A mixture-of-modelers approach to forecasting NCAA tournaments outcomes”, *Journal of Quantitative Analysis in Sports*, 11, 13 – 27.
- Zuccolotto P., Manisera M., Sandri M. (2016): *Big data analytics to model scoring probabilit  in basketball: the effect of shootinh under high-pressure conditions.*