



UNIVERSITÀ
DEGLI STUDI
DI BRESCIA

DIPARTIMENTO DI ECONOMIA

Corso di Laurea
in ECONOMIA E GESTIONE AZIENDALE

Relazione Finale

INDICI STATISTICI PER LA MISURAZIONE DELLE PERFORMANCE NEL GIOCO DEL CALCIO: EXPECTED GOALS ED EXPECTED POINTS

Relatore: Chiar.ma Prof.ssa Paola Zuccolotto

Laureando:

Maraschi Andrea

Matricola n. 719873

Anno Accademico 2019/2020

INDICE

INTRODUZIONE

CAPITOLO 1:

I CONCETTI CHIAVE DELLA STATISTICA APPLICATA ALLO SPORT

1.1.Introduzione sull'importanza della statistica	1
1.2.Binomio statistica-sport	1
1.3.Evoluzione della statistica applicata al calcio	4

CAPITOLO 2:

EXPECTED GOAL (xG) E INDICI DERIVATI

(expected goal against e expected goal difference)

2.1. Nozioni sull'indice xG	8
2.2.I parametri per definire xG.....	9
2.3.Campi di utilizzo dell'indice	13
2.4. I limiti dell'indice	16
2.5.Expected goal against.....	19
2.6. Expected goal difference	20

CAPITOLO 3:

EXPECTED POINTS

3.1.Nozione Expected points	24
3.2.Criteri per l'assegnazione del punteggio	24
3.3.Scostamenti dell'indice dalla realtà	28
3.4. Campi di utilizzo	29

CAPITOLO 4:

ANALISI STAGIONE CALCISTICA 2017/2018

4.1. Premessa iniziale	36
4.2. Il possesso palla	37
4.3. Valutazione capacità offensiva ed xG	39

4.4. Valutazione forza difensiva	45
4.5. Rielaborazione classifica finale attraverso l'indice xPTS	50

CONCLUSIONE

SITOGRAFIA

RINGRAZIAMENTI

INTRODUZIONE

Nel presente elaborato l'autore si propone di analizzare in maniera dettagliata gli indici Expected goals ed Expected points, che rappresentano la più avanzata frontiera della statistica applicata al gioco del calcio.

Nel primo capitolo introduttivo si evidenzieranno le difficoltà di compiere un'analisi statistica nel calcio, facendo particolare riferimento alla diversità strutturale tra calcio e basket. Inoltre, vi sarà una breve storia sull'evoluzione della statistica nel mondo del pallone.

Il secondo capitolo, si pone l'obiettivo di definire ed illustrare in maniera approfondita l'indice Expected goals (e indici derivati), sottolineando i campi di utilizzo, i vantaggi e svantaggi dell'indicatore che ha rivoluzionato la statistica applicata del gioco del calcio.

Nel terzo capitolo viene descritto l'indice Expected points, focalizzando principalmente l'attenzione sulla distribuzione delle probabilità di Bernoulli che sta alla base del metodo di attribuzione del punteggio di tale indicatore.

Il quarto capitolo, invece, è dedicato all'analisi della stagione calcistica 2017/2018 tramite gli indici illustrati nei capitoli precedenti.

In particolare, partendo dalla classifica reale, si andrà ad esaminare i dati più significativi nelle varie zone di campo, per poi rielaborare una graduatoria finale basata sulle prestazioni attraverso l'indice Expected points.

CAPITOLO 1:

CONCETTI CHIAVE DELLA STATISTICA APPLICATA ALLO SPORT

1.1 Introduzione sull'importanza della statistica

La statistica dispone della facoltà di soccorrere l'inevitabile incompletezza del nostro sapere, infatti basta conoscere solamente alcune parti di realtà (campione statistico) per generare indicazioni sull'inconosciuto.

“Servirsi della statistica significa riconoscere che una persona nella vita non avrà il tempo di leggere tutti i libri o guardare tutte le partite, ma selezionando opportunamente un determinato numero di elementi di ciascuna delle categorie interessate esiste la possibilità di ricavare una buona approssimazione dell'intero complesso di oggetti, persone e idee che ci circondano”.¹ Il teorema del limite centrale è la legge che esprime in termini matematici il concetto che è alla base di gran parte delle applicazioni della statistica moderna. Questo teorema che a un certo punto tra il settecento e ottocento inizia ad essere usato, consiste di fare inferenza su parametri ignoti di una popolazione basandosi sulla sola osservazione di un campione.²

1.2 Binomio statistica-sport

Nel panorama sportivo, il basket sembra quasi essere stato concepito appositamente per diventare terra di conquista dell'analisi statistica. Una delle condizioni indispensabili per mettere in moto la capacità statistica è, difatti, la riproducibilità dell'evento che si vuole studiare. In una partita di basket ci sono in media circa 200 tiri a canestro ed una percentuale tra il 50 e il 60 di essi viene realizzata³, quindi si può affermare che avvengono moltissimi eventi misurabili; inoltre, considerando la sola NBA, in ogni stagione regolare ognuna delle 30 squadre gioca 82 partite⁴, il totale di tutto questo non è l'infinito richiesto dal limite centrale, ma per compiere uno studio statistico è assai più che sufficiente. Grazie all'aiuto della tecnologia, negli ultimi anni sono state studiate statistiche più sofisticate, che rispondono a domande

¹ <https://www.rivistaundici.com/2019/04/03/calcio-statistiche/>

² https://it.wikipedia.org/wiki/Teoremi_centrali_del_limite

³ <https://it.global.nba.com/statistics/>

⁴ <https://it.wikipedia.org>

sempre più dettagliate. Infatti si è passato dal chiederci “qual è la probabilità di fare canestro?” a “qual è la probabilità fare canestra in una precisa zona di campo, con una certa situazione difensiva, in un determinato momento della partita?

Dal punto di vista statistico/matematico, dare una risposta a quest’ultimo interrogativo significa inserire variabili aggiuntive alla propria valutazione quantitativa, significa creare una base di dati enorme che per ogni tiro tentato in una partita non fornisca solamente l’esito finale (canestro/non canestro), ma tutta una sequela di dati addizionali. Al di là della disponibilità concreta di dati di questo tipo (nei campi in NBA sono installate avanzatissime telecamere che tracciano i movimenti di tutti i cestisti istante dopo istante), il punto chiave è che il basket si presta perfettamente ad essere descritto da tali dati. La grande quantità di eventi che hanno luogo durante una partita fa sì che non solo i tiri liberi, che rappresentano il gesto sportivo che più si avvicina alle condizioni sperimentali ideali, ma che anche i tiri da due e tre punti possano essere analizzati in maniera dettagliata.

Ad esempio, al giorno d’oggi possiamo sapere che Stephen Curry (il miglior tiratore da 3 punti dell’NBA in attività)⁵ con il proprio difendente ad una distanza compresa tra 50 centimetri ed un metro realizza 44 tiri su 100, ovvero ogni volta che sta per rilasciare il tiro in queste condizioni è legittimo pensare che segnerà con una probabilità del 44%, e compagni e avversari lo devono tenere presente.

Esempio di statistica applicata al basket.

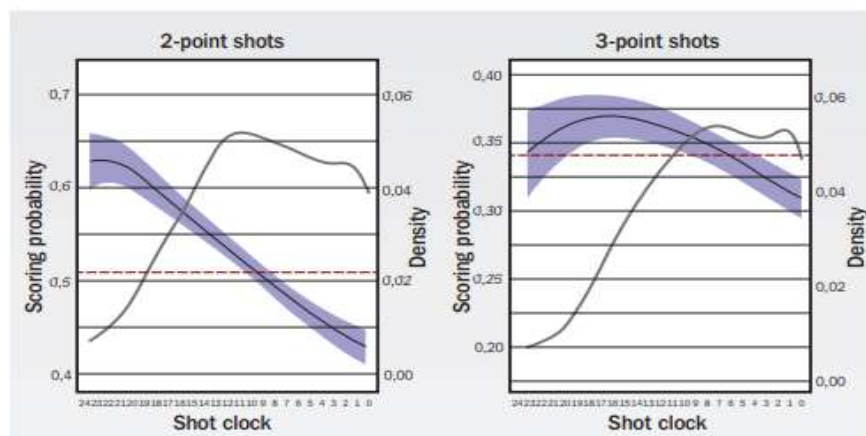
Come varia la percentuale realizzativa in funzione del cronometro dei 24 secondi?

Attraverso l’analisi dei dati del campionato A2ITA durante la stagione regolare 2015/2016 si può affermare che esiste una relazione che lega in modo stabile e deciso la probabilità di fare canestro al cronometro dei 24 secondi (shot clock).⁶ La figura 1 mostra i risultati ottenuti separatamente per i tiri da 2 punti (sinistra) e da 3 punti (destra). L’area colorata in ciascun grafico va

⁵ https://it.wikipedia.org/wiki/Migliori_realizzatori_da_3_punti_NBA

⁶ *sds-scuola dello sport* / xxxvii / 117, pag. 41 di Paola Zuccolotto e Monica Manisera

dal 5° al 95° percentile delle stime bootstrap della probabilità di fare canestro per ogni valore di shot clock (che va da 24 a 0, sull'asse orizzontale). La linea orizzontale rossa tratteggiata rappresenta la probabilità di successo (rapporto tra il numero di tiri realizzati e il numero di tiri tentati) calcolata su tutti i tiri di ogni tipologia. La linea grigia rappresenta la densità stimata dei tiri (per ogni valore di shot clock), la cui misura è riportata sull'asse verticale di destra. Con riferimento ai dati A2ITA, lo shot clock ha un effetto significativo sulla probabilità di successo, che si riduce sensibilmente man mano che la sirena di fine azione si avvicina. Nel caso dei tiri da 2 punti, i tiri tentati nei primi secondi di possesso hanno la più alta probabilità di successo (essi avvengono solitamente in contropiede). Invece, i tiri da 3 punti tentati nei primissimi secondi hanno meno probabilità di successo (normalmente non si tratta di contropiede, in quanto è raro che si ricorra a un tiro da 3 punti in tale situazione, ma di tiri tentati al termine del quarto, quando lo shot clock sarebbe ancora elevato ma il tempo complessivo di gioco sta per scadere). In ogni caso, la gran parte dei tiri sono tentati negli ultimi 12 secondi, come indicato dalla linea grigia della densità di tiro. I grafici della figura 1 includono nell'analisi tutte le squadre del rispettivo campionato; naturalmente, ogni squadra può discostarsi o meno dall'andamento medio, e può pertanto essere interessante osservare l'andamento delle squadre singole rispetto a quello medio generale.



Figura

1.1: Andamento della probabilità di successo (“scoring probability”) dei tiri da 2 punti (sinistra) e 3 punti (destra) in funzione del cronometro dei 24 secondi (“shot clock”), dataset A2ITA (grafici sopra, in blu) e RIO16 (sotto, in verde) – area colorata: stima bootstrap kernel; linea grigia: distribuzione dei tiri tentati (Zuccolotto et al. 2017).

Perché nel calcio un'analisi statistica così accurata sembra difficilmente proponibile?

Contrariamente alla pallacanestro, gli avvenimenti che determinano una partita sono infrequenti, in certi casi del tutto assenti, infatti fare goal è definito come “evento raro” molto più difficile che fare canestro, nel calcio gli avversari sono molto più numerosi, il campo del gioco è più grande e la porta è difesa da un portiere. Come abbiamo visto, alcune delle premesse fondamentali della statistica hanno bisogno di grandi numeri, di ripetibilità, di situazioni che assomiglino più possibile a esperimenti controllati (ad esempio, il lancio di una moneta o l'esecuzione di un tiro libero), mentre i goal differiscono l'uno dall'altro sia nella fase preparatoria che in quella conclusiva.⁷ Si può quindi affermare che il gioco del basket sia concepito come un insieme di gesti riproducibili, mentre per quanto riguarda il calcio lo si potrebbe considerare come un “regno della variabilità.

Nonostante le basi non siano del tutto favorevoli staff professionistici investono tempo e denaro per acquisire nei loro database decine di migliaia di informazioni sui goal e occasioni da goal create nelle ultime stagioni, sperando nel fatto che all'interno della loro “banca dati” ci sia qualche giocata simile, azioni già viste in passato che potranno essere utilizzate come *benchmark* per prevedere l'esito di azioni che avverranno in futuro.

1.3 Evoluzione della statistica applicata al calcio

Fin dalla nascita del computer alcuni avanguardisti del mondo del calcio hanno provato a giudicare i calciatori utilizzando la statistica. Tra essi vale la pena ricordare Arsène Wenger, ex storico *coach* dell'Arsenal, laureato in economia e con la “fissa” dei numeri.

Già alla fine degli anni '80, quando allenava il Monaco usava Top Score, un software sviluppato da un amico per registrare i dati più elementari (goal fatti/goal subiti, percentuale possesso palla ecc..).⁸

⁷ <https://www.rivistaundici.com/2019/04/03/calcio-statistiche/>

⁸ <https://storiedicalcio.altervista.org/blog/calcio-e-numeri.html>

La sterzata definitiva, però, si è verificata nel 1996, dopo che Opta Index iniziò ad elaborare i dati della Premier League (massimo campionato di calcio Inglese). Per la prima volta, tutte le società riuscirono a venire a conoscenza dei chilometri corsi da un certo giocatore ed il numero di passaggi effettuati.⁹

L'analisi delle performance divenne talmente rilevante che, durante la stagione 2001/2002, il Manchester United, guidato da Alex Ferguson, vendette a sorpresa il giocatore Jaap Stam alla società S.S. Lazio¹⁰, questa cessione fu dettata dai risultati delle statistiche, infatti, analizzando i dati si era ritenuto che l'allora difensore 29enne, fosse nella fase calante della sua carriera, ma alla luce dei fatti sbagliando; questo episodio curioso è passato alla storia come il primo trasferimento calcistico dettato fortemente dalla statistica.

Nel 2003 la rivoluzione della statistica applicata al calcio ricevette una nuova corrente di pensiero dagli Stati Uniti d'America, quando Michael Lewis pubblicò il suo libro sul baseball intitolato "*Moneyball*", che alcuni addetti ai lavori del calcio inglese analizzarono rimanendone stupiti.¹¹ Il libro, successivamente rifatto in un film, racconta come il general manager degli Oakland Athletics, Billy Beane, utilizzasse le statistiche per valutare i suoi giocatori, con l'aiuto di tali dati, in poco tempo la piccola squadra in poco tempo ottenne risultati ben oltre le aspettative, finché anche i club più grandi cominciarono ad assumere statistici (i Boston Red Sox vinsero due campionati adottando i metodi descritti in Moneyball).¹²

In sintesi la teoria descritta da Lewis studia ed analizza i dati dei giocatori, non basandosi esclusivamente sulla media battuta o sugli homeruns battuti, quanto su ogni dato che emerge nel corso di una partita (basi per balls, abilità difensive...).

All'inizio diversamente dal baseball, nel mondo calcistico, le teorie descritte nel libro "*Moneyball*" non furono adottate dalle maggiori squadre europee, in quanto il mondo del pallone continuava a diffidare dei numeri e si riteneva

⁹ <https://storiedicalcio.altervista.org/blog/calcio-e-numeri.html>

¹⁰ https://it.wikipedia.org/wiki/Jaap_Stam

¹¹ <https://storiedicalcio.altervista.org/blog/calcio-e-numeri.html>

¹² https://it.wikipedia.org/wiki/L%27arte_di_vincere

difficoltosa la convivenza tra un ottimo statistico ed un allenatore dai metodi tradizionali.

Solamente il Bolton e l'Arsenal, sempre guidata da Wenger, rimasero entusiaste da questo nuovo pensiero.

Un esempio si può ricondurre al 2005, dove l'allenatore francese decise di sostituire l'allora campione Patrick Vieira con un giovane sconosciuto di nome Flamini, poiché attraverso lo studio dei dati statistici emerse che questo calciatore era in grado di correre fino a 14 chilometri a partita.

In poco tempo da soggetto sconosciuto riuscì a distinguersi nella nuova squadra e a diventare un titolare fisso.

In questi anni le statistiche più utilizzate nel panorama sportivo furono il calcolo dei tackle (contrasti) e i chilometri percorsi da ogni singolo calciatore, ma ben presto divenne chiaro che queste informazioni di dominio pubblico erano insufficienti a fornire indicazioni particolarmente approfondite riguardo le performance delle squadre.

Infatti, non vi è prova del fatto che ci sia una relazione tra la distanza totale percorsa e il numero delle vittorie, ed anche per quanto riguarda il numero di tackle sembra un parametro non così influente.

Proprio in merito a quest'ultimo dato, un esempio lampante è Paolo Maldini, ex capitano del Milan e della nazionale italiana, il quale aveva un numero bassissimo di contrasti per partita, ma secondo gli esperti calcistici è stato uno dei migliori difensori della storia di questo sport.

Negli ultimi anni gli analisti sono riusciti ad individuare i dati che interessano davvero, ad esempio, invece che i chilometri percorsi, ora, i match analyst preferiscono esaminare i tratti percorsi ad alta velocità.

Daniele Tognaccini componente dello staff di "MilanLab" (centro di ricerca scientifica interdisciplinare di proprietà dell'A.C. Milan)¹³ già dalla stagione 2007/2008 notò un legame tra numero di scatti e una vittoria (si intende scatto la situazione in cui un calciatore raggiunge i 25 km/h).¹⁴

¹³ <https://www.acmilan.com/it/>

¹⁴ <https://www.scienzemotorie.com/>

Grazie alle nuove tecnologie, la rivoluzione statistica si può definire solo all'inizio, infatti, dalla seconda metà dell'ultimo decennio dei nuovi indicatori di performance hanno fatto la loro apparizione nel mondo sportivo, si sta parlando degli Expected goals ed Expected points¹⁵¹⁶¹⁷¹⁸, indici qualitativi in grado di analizzare ex post i dati di calciatori e club attraverso l'uso di algoritmi.

Questi indicatori verranno trattati nello specifico nel capitolo secondo e terzo capitolo, in quanto tema principale di questa relazione finale.

¹⁵ https://en.wikipedia.org/wiki/Expected_goals

¹⁶ <https://understat.com/>

¹⁷ <https://www.ultimouomo.com/expected-goals/>

¹⁸ <https://www.optasports.com/>

CAPITOLO 2:

EXPECTED GOAL (xG) E INDICI DERIVATI (EXPECTED GOAL AGAINST E EXPECTED GOAL DIFFERENCE)

2.1 Nozioni sugli xG

“L’indice xG è una misura della probabilità che ha un determinato tiro di essere trasformato in goal.” (Goal Italia, Federico Casotti, 2018)¹⁹.

È un indice in scala da 0 a 1 determinato sulla base di migliaia di tiri effettuati nel corso degli anni nei top 5 campionati europei (Italia, Germania, Inghilterra, Spagna, Francia) che calcola la probabilità che un determinato tiro, da una determinata posizione, in un determinato momento della partita possa terminare in rete oppure no.

Ogni tiro ha così un suo valore di “xG”, ovvero la probabilità (eventualmente espressa in percentuale) che ha quel tiro di tramutarsi in goal.

Ad esempio, supponiamo che da una determinata posizione siano stati effettuati 10 mila tiri (aventi le stesse caratteristiche), due mila di questi si sono trasformati in goal, avremo un 20% di possibilità di successo e un conseguente 0,2 xG. Il coefficiente attribuito all’indice riproduce la probabilità di successo del “giocatore medio”.

Il valore dell’indice è sempre compreso tra 0 e 1, infatti ad un tiro che parte dalla linea di porta avversaria verrà attribuito un xG indicativo di 0,99, (da prassi si considera una “*big chance to score*” quando l’indicatore assume un valore superiore a 0,50).

La situazione più oggettiva per capire il funzionamento dell’indicatore si ha studiando i calci di rigore, poiché è l’unica situazione di gioco a presentarsi sempre con le stesse caratteristiche, alla quale viene, quindi, assegnato un xG pari, all’incirca, a 0,79.²⁰

Ci sono diversi modelli per classificare l’xG, tra cui quelli forniti da UNDERSTAT, OPTA e dalla rivista “Ultimo uomo”

¹⁹ <https://www.goal.com/it/notizie/cosa-sono-gli-expected-goals-domande-e-risposte/p53bt2s14p2m1gdb5rfq42pc5>

²⁰ <https://www.goal.com/it/notizie/cosa-sono-gli-expected-goals-domande-e-risposte/p53bt2s14p2m1gdb5rfq42pc5>

2.2 I parametri per definire xG

In questa relazione finale ho scelto di utilizzare come modello di riferimento xG quello fornito dalla rivista “*Ultimo uomo*”.

Prima di definire le variabili che influenzano i coefficienti da attribuire ad ogni tiro bisogna distinguere le tipologie conclusioni, ovvero:

- 1- Tiri effettuati direttamente da calcio piazzato (punizione/rigore)
- 2- Tiri definiti “regolari”, quindi di piede
- 3- Colpi di testa

All’interno di queste categorie si individuano diversi fattori che intervengono sulla probabilità di realizzazione.

La prima variabile da considerare, quando si deve valutare la probabilità di un tiro di diventare goal, è la posizione dalla quale ogni tiro viene calciato.

La posizione è composta da due elementi, la distanza dalla porta e l’angolo di tiro da cui viene calciata la palla in relazione alla porta avversaria.

È generalmente noto, che da una distanza ravvicinata un calciatore ha maggior probabilità di segnare, ma anche l’angolo di tiro riveste un ruolo decisivo, poiché a parità di distanza le *chance* di realizzare una rete differiscono in maniera significativa in relazione al fatto che il tiro venga effettuato da una posizione centrale oppure defilata.²¹

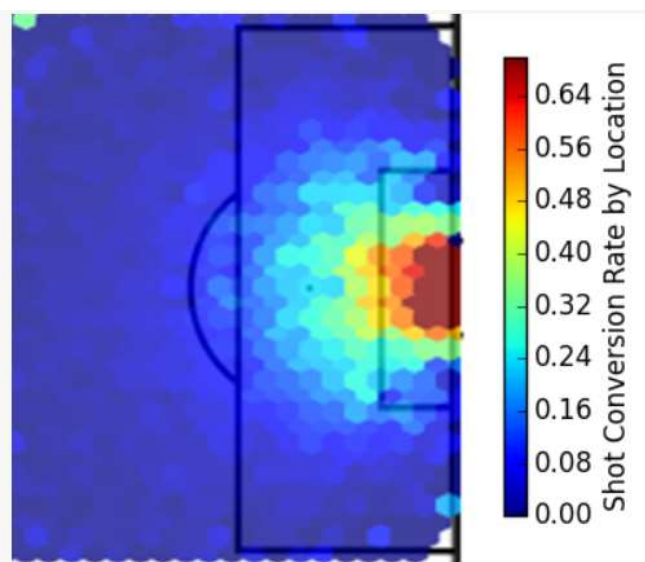


Figura 2.1: shot conversion rate by location (Fonte: Ultimo Uomo, SKY SPORT).

²¹ <https://www.ultimouomo.com/expected-goals/>

Il primo passo per analizzare la correlazione tra distanza e angolo di tiro sta nell'esaminare il database di Christopher Long, il quale contiene le coordinate x e y di ciascun tiro.

Il campo da calcio quindi viene raffigurato come un piano cartesiano con punto d'origine (0; 0) in una delle bandierine del calcio d'angolo.

“Le coordinate estratte dal database hanno rappresentato i dati input del modello matematico e hanno permesso di ricavare un'equazione con cui calcolare la probabilità di fare goal, una volta conosciuta distanza e angolo di tiro.”²²

Questa equazione realizza così un modello di tipo “continuo”, cioè non si rifà ad un frazionamento in zone di gioco sostituendo la vecchia teoria di Paul Riley che divideva il campo in 46 settori all'interno dei quali ogni tiro assumeva la stessa probabilità di successo, definendo così un modello di tipo “discreto”.²³



Figura 1.2: Modello SPAM di Paul Riley (Fonte: Different game).

Un altro fattore da tenere in considerazione è il valore dell'assist, per realizzare una rete, come già detto, conta moltissimo la posizione, ma anche ricevere il giusto assist.

²² <https://www.ultimouomo.com/expected-goals/>

²³ <https://www.ultimouomo.com/expected-goals/>

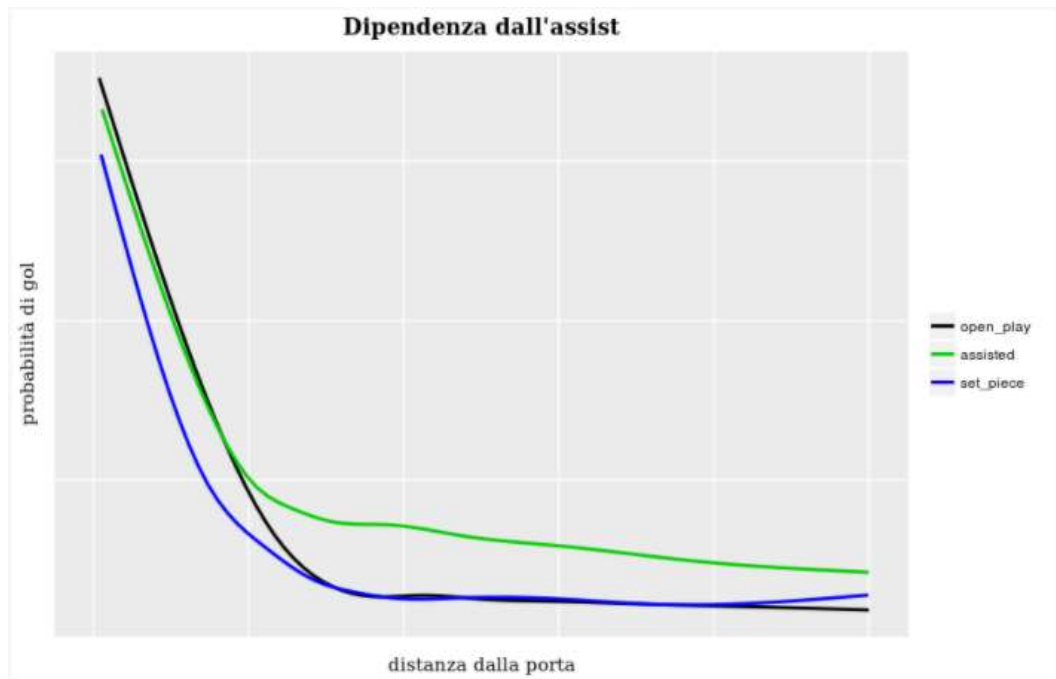


Figura 2.3: Dipendenza dall'assist (Fonte: Ultimo Uomo, SKY SPORT).

La figura 2.3 ci permette di comparare come varia la probabilità di goal in funzione della distanza dalla in tre diverse situazioni di gioco.

La linea nera rappresenta la percentuale di realizzazione in un'azione "regolare", non avvenuta successivamente ad un calcio piazzato e non favorita da un compagno, ad esempio un giocatore che calcia verso la porta dopo aver fatto un'azione solitaria. A distanze elevate la probabilità di segnare una rete diminuisce sensibilmente

La linea blu riproduce la possibilità di realizzazione per un tiro successivo ad un calcio piazzato, a distanze ravvicinate fare goal attraverso un calcio piazzato è assai meno probabile rispetto al caso precedente. Possiamo facilmente individuare la difficoltà di segnare da un calcio d'angolo, dove tutti i giocatori avversari si trovano in area di rigore. A distanze elevate le due linee sostanzialmente si sovrappongono.

La linea verde evidenzia i valori di xG successivi ad un assist. Al di sopra di una certa distanza, ricevere un assist aumenta in maniera consistente la probabilità di segnare un goal.

Il risultato in corso di partita (game state) modifica l'attitudine delle squadre in campo, a parità di altre condizioni, solitamente chi è in vantaggio ha le occasioni migliori.

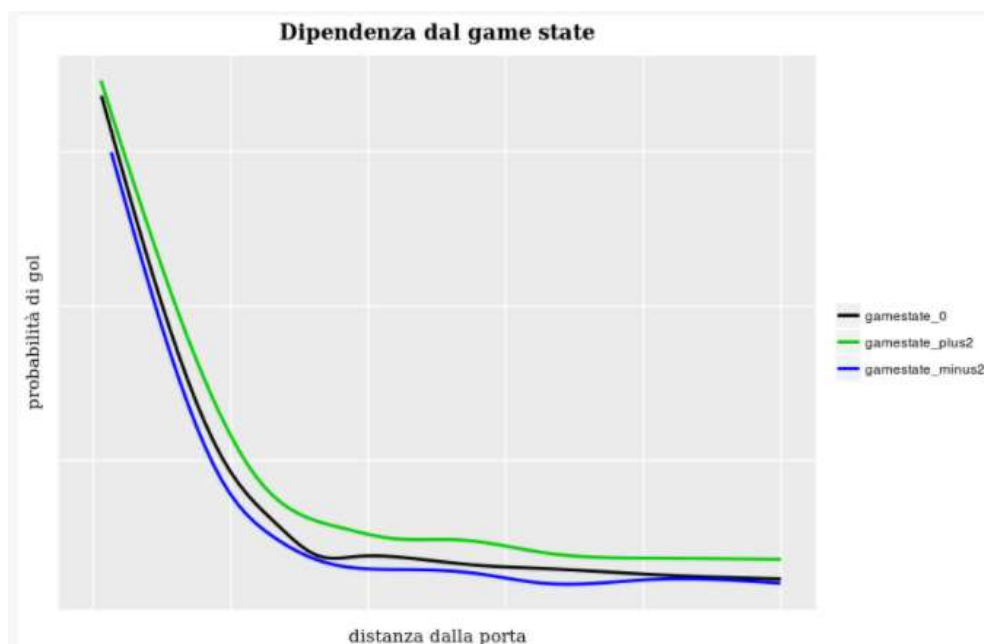


Figura 2.4: Dipendenza dal game state (Fonte: Ultimo uomo, SKY SPORT).

La figura 2.4 mette in relazione come varia la probabilità di segnare una rete in tre diverse situazioni, `gamestate_0` (le due squadre sono sul risultato di parità), `gamestate_plus2` (squadra in vantaggio di una o più reti), `gamestate_minus2` (squadra in svantaggio di una o più reti)

La linea verde nel grafico denota come a parità di altre condizioni (per esempio teniamo fissa la variabile distanza dalla porta) vi è maggior probabilità di fare goal delle squadre in vantaggio di una o più reti nel punteggio.

Questo è un discorso qualitativo, quindi si tende a sottolineare come le squadre in svantaggio cerchino (soprattutto nei minuti finali) di calciare di più, ma nella confusione delle ultime azioni cercano soluzioni con poca possibilità di andare a segno (linea blu).

Anche il fattore campo ha una certa influenza sull' xG, nonostante esistano studi che dimostrano come il fattore campo negli ultimi anni stia via via perdendo di importanza questa variabile persiste in termini di goal attesi.

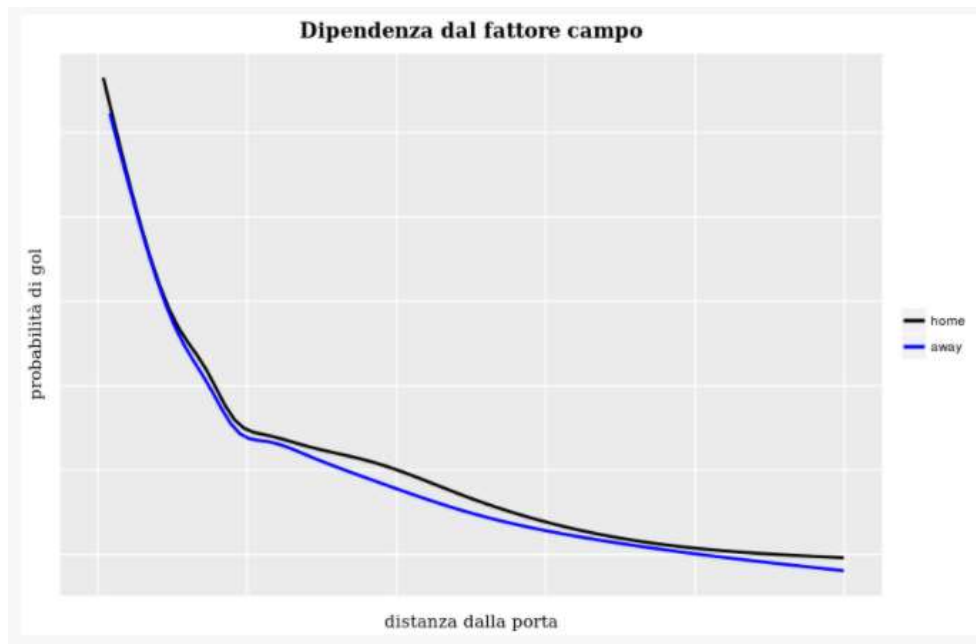


Figura 2.2: Dipendenza dal fattore campo (Fonte: *Ultimo uomo*, SKY SPORT).

Dalla figura 2.5 si nota che a parità di altre condizioni, le squadre in casa (soprattutto a livello europeo) hanno un leggero vantaggio rispetto alle avversarie.

Questo ultimo fattore non viene più considerato nell'arco del 2020 dove a causa COVID-2019 tutti gli incontri sono stati disputati a porte chiuse.

2.3 Campi di utilizzo dell'indice.

L'utilizzo primario dell'indice è quello di analizzare e controllare le prestazioni sia del giocatore singolo sia come squadra, infatti, sommando l'xG complessivo di un particolare giocatore o di una determinata squadra fornisce un'indicazione di quanti goal quel giocatore o quella squadra avrebbero dovuto segnare in media.²⁴

Ad esempio un attaccante definito "cinico" sarà un calciatore in grado di tramutare in goal anche occasioni statisticamente sfavorevoli.

Nel caso ipotizzato ci aspetteremo un numero di goal totali a fine stagione maggiore rispetto al numero di goal attesi (dati dalla somma degli xG di tutti i suoi tiri effettuati).

²⁴ <https://www.optasports.com/services/analytics/advanced-metrics/>

L'indice, inoltre, può anche fornire un'interpretazione sull'andamento futuro della squadra. Portiamo come esempio il caso Juventus 2015/2016.

La squadra torinese iniziò nel peggiore dei modi il campionato, trovandosi dopo 10 partite in zona retrocessione, in realtà andando ad analizzare gli xG delle prime giornate, si scopre che la squadra stava producendo lo stesso numero di occasioni del girone di ritorno, che la vide protagonista al punto che la squadra si aggiudicò lo scudetto.



Figura 2.3: xG e xGA, Juventus stagione 2015/2016 (Fonte: GOAL).

Evidentemente nelle prima parte del campionato c'era sfortuna, poca lucidità, bravura dei portieri o altri fattori che non possiamo in qualche modo preventivare.

L' xG può quindi aiutare a capire il reale rendimento di una squadra al di là dei risultati e delle posizioni in classifica, che spesso possono rivelarsi legate a elementi casuali difficilmente controllabili.

A tal proposito, facendo riferimento ad una tesi di laurea del 2011, menzionata anche dal Giornale di Brescia, si è notato che durante una partita di calcio intervengono 1200 variabili che possono modificare il risultato di una partita.

Delle 1.200 variabili individuate, Roberta Tengattini (autrice della tesi) ha fatto una scrematura andando a individuare otto fattori per determinare il risultato: tre per la squadra di casa (giocate aeree, giocate pericolose a palla bassa e indice di difesa) e cinque per la squadra in trasferta (passaggi lunghi oltre il centrocampo, indice di pericolosità in attacco, tentativi di contropiede, fase difensiva a protezione della porta e barriera difensiva a protezione dell'area).²⁵

Andando ad eliminare gli elementi casuali, come il meteo ed errori arbitrali la classifica di serie A durante la stagione 2010/2011 presenta notevoli differenze.

La classifica reale...			...e secondo il modello		
1	Milan	82	1	Inter	80
2	Inter	76	2	Milan	77
3	Napoli	70	3	Udinese	72
4	Udinese	66	4	Juventus	65
5	Lazio	66	5	Roma	62
6	Roma	63	6	Genoa	60
7	Juventus	58	7	Palermo	58
8	Palermo	56	8	Lazio	57
9	Fiorentina	51	9	Catania	56
10	Genoa	51	10	Cagliari	55
11	Chievo	46	11	Fiorentina	51
12	Parma	46	12	Napoli	48
13	Catania	46	13	Lecce	47
14	Cagliari	45	14	Brescia	46
15	Cesena	43	15	Sampdoria	45
16	Bologna	42	16	Bari	41
17	Lecce	41	17	Cesena	39
18	Sampdoria	36	18	Parma	35
19	Brescia	32	19	Chievo	33
20	Bari	24	20	Bologna	30

 in Champions League	 ai Preliminari di Champions
 in Europa League	 retrocesse in Serie B

Figura 2.4: Classifica reale e classifica rielaborata della stagione 2010/2011 (Fonte: Giornale di Brescia).

La figura 2.7 mostra come secondo il modello l'Inter avrebbe vinto lo scudetto ai danni del Milan, mentre il Brescia avrebbe raggiunto la salvezza "guadagnando" ben 26 punti in più rispetto alla classifica reale.

²⁵ <https://bodai.unibs.it/bdsports/wp-content/uploads/sites/2/2021/01/GdB2012.pdf>

2.4 I limiti dell'indice

La principale critica rivolta all'indicatore xG è la mancanza nel calcolo di un indicatore della pressione difensiva, in sostanza, l'indice non tiene in considerazione dove si trovano i difensori al momento del tiro.

Esistono già studi che vanno in questa direzione ma, al contrario del basket in NBA, nel mondo del pallone non viene impiegata in maniera sistematica la tracciabilità spazio-temporale dei movimenti.

Un altro limite significativo dell'indice consiste nel fatto che durante la partita non tutte le azioni pericolose si traducono in un tiro verso la porta, quindi i goal attesi tendono a sottostimare la pericolosità complessiva dell'azione.

Per ovviare a questi limiti esistono alcuni indici specifici in grado di calcolare la probabilità che un determinato assist si tramuti in goal in questo caso parliamo di Expected assist (xA), esistono anche veri e propri indici di progressione della palla nel campo da gioco, tuttavia questi modelli sono ancora in via sperimentale e non ancora applicati in maniera professionale al calcio.

Premesso che l'xG è una statistica sicuramente perfezionabile, la loro superiorità rispetto a tutti gli altri indici pubblici basati sui tiri è innegabile.

Dal 2014, Opta organizza gli OptaPro Analytics Forum, che rappresentano un'opportunità sia per i semplici interessati sia per i match analyst professionisti: durante il Forum 2019 tenutosi a Londra e successivamente a Colonia in Germania, si sono iscritti i rappresentanti di oltre 70 società di calcio.

Dal 2006 annualmente negli USA si svolge anche la Sloan Sports Analytics Conference, organizzata dalla MIT Sloan, la Business School del Massachusetts Institute of Technology di Boston. La conferenza è copresieduta da Daryl Morey, Presidente delle operazioni di basket dei Philadelphia 76ers e Jessica Gelman, CEO di KAGR (Kraft Analytics Group)²⁶.

Secondo l'obiettivo dell'istituto, la SSAC è un luogo di incontro tra gli esperti del mondo dello sport, gli studenti della scuola e i semplici

²⁶ https://en.wikipedia.org/wiki/MIT_Sloan_Sports_Analytics_Conference

appassionati ad oggi è diventata la sede principale per le discussioni sull'analisi dei dati sportivi. A differenza degli OptaPro Forum, la SSAC si occupa di diversi sport, in particolare la conferenza tratta l'analisi dei dati di MLB, NBA, NFL, NHL, MLS e Premier League anche se negli ultimi anni il calcio sta conquistando sempre più spazio. Possiamo quindi affermare che questi due eventi rappresentano il futuro della statistica nel modo del calcio e non solo.

La nuova sfida della statistica legata al calcio, è quella di creare modelli che non prescindano da un evento (restando nel campo dell'xG abbiamo detto che, anche se ci fosse un'azione pericolosa ma che non culmina attraverso un tiro in porta, il modello non la prenderebbe in considerazione).²⁷²⁸

Durante l'OptaPro Analytics Forum del 2019 sono stati presentati 2 modelli a parer mio molto interessanti, riassunti da Alfredo Giacobbe nella rivista *“Ultimo uomo”*.

1. *“Il modello presentato da Javier Fernandez, Data Scientist del Barcellona, utilizza i dati di posizione di tutti i 22 uomini in campo, registrati istante per istante. In ogni momento, un calciatore può decidere se avanzare portando palla, se effettuare un passaggio verso un compagno o se tentare un tiro. Considerando la probabilità che ciascuna di queste tre azioni possa compiersi istante per istante in ogni zona di campo, e misurando la loro efficacia presunta, il modello di Fernandez calcola un Expected Possession Value. L'indice EPV stima la probabilità che ogni possesso possa scaturire in un gol per chi ha la palla (con il valore EPV che oscilla tra 0 e 1), oppure in un gol per chi non ha la palla in caso di turnover (cioè di una palla persa, con il valore EPV che oscilla tra -1 e 0. L'approccio proposto è estremamente complesso dal punto di vista matematico, ma la ricchezza delle informazioni spazio-temporali che è possibile ricavare sul gioco è impagabile. Ogni singolo giocatore, quando è in possesso del pallone, può essere valutato nella sua capacità di generare valore (quantificato nella differenza positiva nel EPV dopo un suo passaggio), a fronte del rischio che è in grado di sostenere (la*

²⁷ <https://www.ultimouomo.com/statistiche-calcio-novita/>

²⁸ https://en.wikipedia.org/wiki/MIT_Sloan_Sports_Analytics_Conference

differenza negativa nel EPV in caso di turnover dopo un suo errore). Allo stesso modo, anche un giocatore che non è in possesso del pallone, ma che detta un passaggio smarcandosi con un movimento corretto nello spazio eseguito con i tempi giusti, può essere valutato nella sua capacità di generare valore senza palla.

2. *Il secondo modello, presentato dal Data Analyst dell'Huddersfield Town Mladen Sormaz, si concentra ancora di più su quello che accade lontano dalla palla. In questo caso, indipendentemente dal fatto che il pallone venga servito o meno a un determinato giocatore, viene valutata la sua capacità di disordinare le linee avversarie con una corsa senza palla.²⁹*

La forma difensiva di riferimento di una squadra può essere nota a priori, o valutata ex-post mediante le posizioni medie tenute in fase di non possesso. Come cambia spazialmente questa forma rispetto al riferimento, in reazione ad una corsa senza palla di un offendent, dà la misura del danno creato dalla corsa stessa.

Dal confronto tra la forma ideale e quella reale viene calcolato il punteggio che misura l'efficacia della corsa senza palla, variabile da 0 (forma ideale mantenuta inalterata) a 1 (forma avversaria completamente disorganizzata). Accoppiando poi questa misura del danno indotto nella squadra che si difende con la quantità di spazio in controllo della squadra in possesso palla, agli autori è stato possibile verificare una correlazione tra la qualità della corsa e la creazione del pericolo. Cioè più è alto lo scompiglio che una corsa senza palla crea nelle linee avversarie, maggiore sarà la probabilità che l'azione offensiva terminerà con un tiro. Gli autori per ora hanno solo conteggiato i tiri effettuati in corrispondenza di valori elevati del loro indice, ma già pensano di integrare un modello di xG.”³⁰

²⁹ <https://www.ultimouomo.com/statistiche-calcio-novita/>

³⁰ <https://www.ultimouomo.com/statistiche-calcio-novita/>

2.5 Expected goal against

L'indice *Expected goal against* (xGA) è una misura della probabilità che un determinato tiro avversario venga trasformato in goal.³¹

L'indice xGA rappresenta il corrispettivo dell'xG, esaminato nella propria metà campo. Questo indice aiuta a valutare la qualità della fase difensiva, una grande difesa presuppone che si limitino la quantità e la qualità dei tiri avversari.

I goal effettivamente subiti dipendono anche dalla precisione delle squadre incontrate e dalla prestazione del proprio portiere.

Ad esempio, supponiamo che alla fine di una partita la squadra A abbia prodotto 10 tiri pari a $0,20xG + 0,13xG + 0,30xG + 0,37xG + 0,10xG + 0,50xG + 0,25xG + 0,15xG + 0,5xG + 0,5xG = 3xG$ complessivi.

Questo vuol dire che la squadra A avrebbe potuto realizzare in media 3 gol nel corso di quella partita.

Similmente, la squadra avversaria nel nostro caso squadra "B" potrebbe anch'essa aver calciato verso la porta.

Questi tiri rappresentano l'indice xG prodotto dalla squadra B, mentre dal punto di vista della squadra A essi rappresentano i tiri concessi all'avversario, ovvero xGA.

Nell'esempio, supponiamo la squadra A, nella stessa partita in cui ha prodotto complessivamente $3xG$ abbia concesso complessivamente $2xGA$. Al contrario la squadra B ha prodotto $2xG$ e concesso $3xGA$.

Da un punto di vista strettamente ed esclusivamente della prestazione la squadra A ha prodotto di più degli avversari ed è lecito aspettarsi una vittoria.

Nel basket sarà così quasi sicuramente, nel calcio, per il fatto che il gol a differenza del canestro si configura come "evento raro", la variabilità della previsione è più elevata e la correlazione tra prestazioni e risultato nella singola partita è meno frequente, salvo che la differenza tra xG non sia abissale.³²

³¹ <https://www.goal.com/it/notizie/cosa-sono-gli-expected-goals-domande-e-risposte/p53bt2s14p2m1gdb5rfq42pc5>

³² <https://www.calcioefinanza.it/2018/03/19/statistica-applicata-al-calcio-expected-goals/>

2.6 Expected goal difference

Gli Expected goal difference (xGD) corrispondono alla differenza reti (i goal fatti meno quelli subiti) comparata alla differenza degli xG (cioè la differenza tra le occasioni da goal prodotte misurate in xG, e le occasioni concesse misurate in xGA).³³

Per capire meglio l'xGD propongo un'analisi riferita alle prime 14 giornate del campionato di calcio 2020/2021.

Inizialmente creiamo una tabella riferita ai goal effettivamente realizzati e all'indice xG.

POSIZIONE	TEAM	GOAL	XGOAL
1	AC MILAN	32	32,15
2	INTER	34	29,59
3	ROMA	31	34,99
4	SASSUOLO	26	20,55
5	NAPOLI	27	23,41
6	JUVE	25	28,19
7	ATALANTA	28	26,08
8	LAZIO	22	16,41
9	VERONA	15	20,25
10	BENEVENTO	17	14,25
11	SAMPDORIA	23	18,28
12	UDINESE	14	19,46
13	BOLOGNA	21	19,99
14	FIorentina	16	20,11
15	CAGLIARI	21	16,48
16	PARMA	13	14,01
17	SPEZIA	19	16,01
18	GENOA	14	12,81
19	CROTONE	13	11,65
20	TORINO	22	16,37

Tabella 2.1: Goal e xG anno 2020/2021 (Fonte: nostre elaborazioni).

Il secondo passaggio è creare una tabella analoga alla prima, ma riferita ai goal effettivamente subiti e all'indice xGA.

³³ <https://www.calcioefinanza.it/2018/03/19/statistica-applicata-al-calcio-expected-goals/>

POSIZIONE	TEAM	GOAL AGAINST	XGOAL AGAINST
1	AC MILAN	16	16,97
2	INTER	17	19,24
3	ROMA	20	16,62
4	SASSUOLO	17	20,67
5	NAPOLI	12	9,75
6	JUVE	13	12,75
7	ATALANTA	20	15,60
8	LAZIO	23	19,70
9	VERONA	14	21,51
10	BENEVENTO	23	25,72
11	SAMPDORIA	24	23,34
12	UDINESE	17	10,33
13	BOLOGNA	27	25,32
14	FIorentina	21	22,82
15	CAGLIARI	27	28,61
16	PARMA	25	21,62
17	SPEZIA	29	24,95
18	GENOA	27	23,33
19	CROTONE	29	25,36
20	TORINO	32	26,80

Tabella 2.2: Goal against vs xGA stagione 2020/2021 (Fonte: nostre elaborazioni).

Il terzo ed ultimo passaggio consiste nel trovare per sottrazione la differenza reti effettiva e quella attesa.

POSIZIONE	TEAM	GOAL DIFF.	XGOAL DIFF.
1	AC MILAN	16	15,18
2	INTER	17	10,35
3	ROMA	11	18,37
4	SASSUOLO	9	-0,12
5	NAPOLI	15	13,66
6	JUVE	12	15,44
7	ATALANTA	8	10,48
8	LAZIO	-1	-3,29
9	VERONA	1	-1,26
10	BENEVENTO	-6	-11,47
11	SAMPDORIA	-1	-5,06
12	UDINESE	-3	9,13
13	BOLOGNA	-6	-5,33
14	FIorentina	-5	-2,71
15	CAGLIARI	-6	-12,13
16	PARMA	-12	-7,61
17	SPEZIA	-10	-8,94
18	GENOA	-13	-10,52
19	CROTONE	-16	-13,71
20	TORINO	-10	-10,43

Tabella 2.3: Goal difference vs xGD stagione 2020/2021 (Fonte: nostre elaborazioni).

Analizzando la figura 2.3 possiamo notare che alcune squadre presentano delle differenze marcate tra la differenza reti reale e quella attesa.

La prima squadra che prendiamo in considerazione è l'Inter, la differenza reti reale è di 17 mentre quella calcolata dall'indice è di 10,35, questa discrepanza è dovuta ai goal segnati, in quanto, come possiamo vedere nella prima tabella, i goal effettivamente realizzati sono 34, mentre quelli attesi 29.

Possiamo quindi dedurre che l'Inter è riuscita a trasformare in reti anche situazioni dal punto di vista statistico non favorevoli, questo può essere dettato dalle qualità degli attaccanti (Romelu Lukaku con 11 goal in 14 partite è uno degli attaccanti più prolifici del campionato) oppure da un periodo particolarmente fortunato.

Il secondo esempio che andiamo a proporre è la sorpresa di questo campionato, ovvero il Sassuolo.

La squadra emiliana si trova attualmente al quarto posto, nonostante l'ambizione di inizio campionato fosse la salvezza in serie A.

Analizzando la tabella degli xGD mostra una differenza reti reale di +9 goal, mentre quella attesa di -0,12.

La prima differenza sostanziale riguarda i goal segnati (26) e quelli attesi (20,55). Non avendo il Sassuolo attaccanti particolarmente forti, possiamo ricondurre probabilmente questa over-performance offensiva al frutto della casualità e del momento particolarmente fortunato.

Oltre ai goal fatti, a sostegno della mia tesi porto i goal subiti; quelli effettivamente subiti ammontano a 17, mentre quelli attesi a 21, anche in questo caso si parla di un over-performance difensiva.

Secondo il mio parere, quindi, l'andamento della squadra Sassuolo non sarà tale da confermare, a fine anno, il posizionamento attuale.

Un caso contrario a quelli precedentemente analizzati è l'Udinese, studiando i dati notiamo la presenza di una under performance.

Infatti, la differenza reti reale è di -3, mentre quella attesa di +9,13.

Possiamo notare come la squadra friulana segni meno e subisca più di quanto previsto. Tale situazione sarà dovuta a un caso sfortunato? Alla mancanza di attaccanti e difensori di qualità?

Per provare a dare una risposta ho cercato di studiare i dati dell'anno precedente, i goal realizzati dall'Udinese nella stagione 2019/2020 sono stati 37 mentre l'indice xG è stato di 45.24, anche in questa stagione la squadra ha mostrato difficoltà realizzative in zona goal.

Avendo mantenuto la stessa composizione della squadra in attacco, è lecito supporre che non sia frutto di un caso sfortunato, ma di una carenza a livello qualitativo degli attaccanti.

CAPITOLO 3:

EXPECTED POINTS

3.1 Nozione *expected points*

L'indice denominato *expected points* (xPTS) assegna un punteggio da 0 a 3 ad ogni squadra in ogni partita³⁴.

Nella misurazione dell'indicatore xPTS non si assegnano tre punti (in caso di vittoria), un punto (nel caso di parità) o zero punti (nel caso di sconfitta) come avviene nella realtà, ma per esempio, alla squadra vincente potrebbero essere assegnanti 2 punti e alla squadra perdente 0,7 o 0,5 punti.

L'indice non prende in considerazione il risultato effettivo, infatti potrebbe anche verificarsi un caso in cui la squadra perdente abbia prodotto molte occasioni non riuscendo a segnare e abbia perso nonostante avesse subito pochissimi tiri. In questo caso estremo la squadra vincente potrebbe vedersi assegnato addirittura un indice xPTS minore di quello avversario.

L'obiettivo primario dell'indicatore è quello di rielaborare una classifica basata sulle prestazioni delle squadre.

3.2 Criteri per l'assegnazione del punteggio

Il metodo più preciso per calcolare l'indice consiste nell'assumere che l'esito di un tiro con indice xG sia modellabile come una variabile di Bernoulli di parametro $p = xG$ che assume il valore 1 se il tiro termina in un goal e 0 altrimenti.

In una partita, per entrambe le squadre, si associa ad ogni tiro una variabile casuale. Successivamente si simula una partita estraendo casualmente l'esito di ogni tiro dalla variabile di Bernoulli che lo descrive. Sommando gli esiti si ottiene, per ogni squadra, il numero di goal della simulazione e, confrontando tale valore per le due squadre, si definisce l'esito della partita in vittoria, pareggio o sconfitta della squadra di casa (1, X, 2). Tale procedimento si ripete un gran numero di volte e al termine della simulazione si stimano le

³⁴ <https://www.calcioefinanza.it>

probabilità di vittoria, pareggio e sconfitta con le rispettive frequenze relative nella simulazione.

L'indice xPTS è quindi dato dalla somma tra la percentuale di vittoria moltiplicata *3 punti e la percentuale di pareggio moltiplicata *1 punto.

In particolare;

$xPTS \text{ (squadra A)} = (\text{numero di vittorie squadra A} / \text{numero di simulazioni}) * 3 + (\text{numero di pareggi} / \text{numero di simulazioni}) * 1 + (\text{numero di vittorie squadra B} / \text{numero di simulazioni}) * 0$

$xPTS \text{ (squadra B)} = (\text{numero di vittorie squadra B} / \text{numero di simulazioni}) * 3 + (\text{numero di pareggi} / \text{numero di simulazioni}) * 1 + (\text{numero di vittorie squadra A} / \text{numero di simulazioni}) * 0$

Espresso in termini generali, l'indice xPTS rappresenta il valore atteso di una variabile casuale discreta che assume valori {0;1;3} con probabilità di sconfitta, pareggio e vittoria simulata attraverso il metodo "Monte Carlo".

A titolo esemplificativo propongo 3 configurazioni realizzate con Excel sulla base di 10000 simulazioni.

Configurazione 1:

xG TIRI	Squadra A	Squadra B	Numero vittorie squadra A	2308	23,08%
xG TIRO 1	0,1	0,9	Numero vittorie squadra B	4292	42,92%
xG TIRO 2	0,6	0,8	Numero di pareggi	3400	34,00%
xG TIRO 3	0,2	/	Numero simulazioni	10000	100,00%
xG TIRO 4	0,3	/	Expected points squadra A	1,03	
xG TIRO 5	0,4	/	Expected points squadra B	1,63	
SOMMA	1,6	1,7			

Tabella 3.4: Configurazione 1 xPTS (Fonte: nostre elaborazioni).

La configurazione 1 della tabella 3.1 mostra il risultato complessivo del calcolo di 10k simulazioni.

Viene ipotizzato che la squadra A nel corso della partita abbia effettuato cinque tiri, mentre la squadra B due (ad ogni tiro di entrambe le squadre viene associato l'xG che rappresenta la probabilità di successo espressa da 0 a 1).

Attraverso il foglio di lavoro Excel ad ogni tiro viene generato un valore estratto da una variabile casuale uniforme (compresa tra 0 e 1) che potrà essere maggiore o minore dell'xG.

Nel caso il valore fosse minore vorrà dire che il tiro si è trasformato in goal e la cella assumerà valore "1", nel caso contrario verrà assegnato valore "0".

Questo procedimento consente di generare l'esito di una variabile casuale di Bernoulli di parametro $p=Xg$ utilizzando la funzione disponibile su Excel, che estrae valori casuali uniformemente distribuiti nell'intervallo $[0,1]$.

Nel nostro caso, sulla base di 10k simulazioni si sono riscontrate 2308 vittorie della squadra A, con una percentuale del 23,08% ($2308/10000 * 100$), 4292 vittorie della squadra B (42,92%) e 3400 pareggi (34,00%).

L'indice xPTS è stato così calcolato:

xPTS squadra A: $23,08\% * 3 + 34,00\% * 1 = 1,03$

xPTS squadra B: $42,92\% * 3 + 34,00\% * 1 = 1,63$

Configurazione 2:

xG TIRI	Squadra A	Squadra B		
xG TIRO 1	0,1	0,15	Numero vittorie squadra A	3344 33,44%
xG TIRO 2	0,6	0,45	Numero vittorie squadra B	3835 38,35%
xG TIRO 3	0,2	0,25	Numero di pareggi	2821 28,21%
xG TIRO 4	0,3	0,15	Numero simulazioni	10000 100,00%
xG TIRO 5	0,4	0,6	Expected points squadra A	1,29
SOMMA	1,6	1,7	Expected points squadra B	1,43

Tabella 3.5: Configurazione 2 (Fonte: nostre elaborazioni).

La tabella 3.2 denota come la squadra B a parità di condizioni (lasciando invariato il comportamento della squadra A e mantenendo lo stesso valore xG complessivo dell'1,7) abbia statisticamente meno *chance* di vittoria

realizzando un maggior numero di tiri complessivi ma dal coefficiente minore.

Passando da una probabilità stimata di vittoria pari al 42,92% di vittoria (tabella 3.1) a una probabilità stimata pari al 38,35% nella configurazione 2.

Viceversa, per quanto riguarda la squadra A, le probabilità di vittoria aumentano in maniera considerevole (circa del 10%) concedendo all'avversario più tiri statisticamente non favorevoli rispetto a poche occasioni ma dal coefficiente elevato.

Questo confermerebbe che nel calcio sia più rilevante la qualità dei tiri rispetto alla quantità.

Configurazione 3:

xG TIRI	Squadra A	Squadra B
xG TIRO 1	0,35	0,15
xG TIRO 2	0,45	0,05
xG TIRO 3	0,15	
xG TIRO 4	0,75	
xG TIRO 5	0,12	
SOMMA	1,82	0,2

Numero vittorie squadra A	8698	86,98%
Numero vittorie squadra B	154	1,54%
Numero di pareggi	1148	11,48%
Numero simulazioni	10000	100,00%
Expected points squadra A	2,72	
Expected points squadra B	0,16	

Tabella 6.3: Configurazione 3 (Fonte: nostre elaborazioni).

A inizio capitolo si è detto che l'indice xPTS non considera il risultato effettivo della partita e che, quindi, si potrebbe presentare un caso estremo in cui alla squadra vincitrice venga assegnato un punteggio inferiore rispetto alla squadra perdente.

La configurazione 3 si pone l'obiettivo di visualizzare in maniera concreta questa situazione "particolare".

Nell'esempio proposto, qualunque fosse il risultato maturato sul terreno di gioco, l'indice xPTS assegnerebbe 2,72 punti alla squadra A e 0,16 alla squadra B (anche nel caso quest'ultima vincessesse la partita).

Sono state studiate dagli analisti un gran numero di situazioni reali e si è constatato che l'indice xPTS ha una relazione stretta con la differenza tra xG e xGA. Sulla base di queste analisi è stata proposta la classificazione di “*OPTA*”

EXPECTED GOALS DIFFERENCE (XG – XGA)	EXPECTED POINTS
Fino a -1.5	0.1
-1.5 – -1	0.3
-1 – -0.5	0.5
-0.5 – 0	0.7
0 – 0.5	1.5
0.5 – 1	2
1 – 1.5	2.3
Oltre a 1.5	2.7

Tabella 3.7: Distribuzione xPTS (Fonte:OPTA).

La tabella 3.4 consente di approssimare il valore dell'indice xPTS (senza dover svolgere ogni volta la simulazione), in maniera più immediata, ma meno precisa, partendo dalla differenza tra xG e xGA.

3.3 Scostamenti dell'indice dalla realtà

Per come è concepito l'indice xPTS, esso tende ad essere sempre inferiore ai punti effettivi delle squadre di vertice, dato che questi club vincono molte partite conquistando i 3 punti, mentre nella classifica rielaborata ciò può non accadere.

Al Contrario l'xPTS tende ad essere superiore ai punti effettivamente realizzati per le squadre di bassa classifica, le quali per il motivo inverso ai top club, anche se sconfitte in maniera netta potrà non venire attribuito zero al valore xPTS ma 0.1 o 0.3.

I punti reali si differenziano da quelli attesi anche per altri fattori, per esempio dalle caratteristiche di ogni squadra, dal tipo di gioco e dai valori individuali di ciascun team.

3.4 Campi di utilizzo

L'indice xPTS è utilizzato dallo staff tecnico di un team (composto da allenatore, match analyst, preparatore atletico ecc..) per analizzare e far crescere la qualità complessiva del gioco, indipendentemente dalla presenza di singoli giocatori in grado di risolvere la partita da soli.³⁵

La principale differenza tra una valutazione fatta attraverso xG e xPTS sta nel fatto che, questo secondo deve essere analizzato sul lungo periodo, in quanto ha poco senso valutare uno scostamento tra punti reali ed attesi prendendo come base di riferimento poche partite.

Lo strumento degli xPTS è utile anche per i direttori sportivi e dirigenti per programmare la campagna acquisti e prevedere la stagione successiva come possiamo vedere dall'esempio Manchester city stagione 2016/2017.

Esempio numerico applicato all'indice xPTS:

POSIZIONE	TEAM	G			XG		
		GOAL	AGAINST	POINTS	XGOAL	AGAINST	XPTS
1	CHELSEA	85	33	93	61,8	28,62	75,74
2	TOTTENHAM	86	26	86	70,07	33,78	75,37
3	MANCHESTER CITY	80	39	78	79,85	29,61	85,41
4	LIVERPOOL	78	42	76	66,63	37,65	69,83
5	ARSENAL	77	44	75	63,58	47,07	62,21
6	MAN UNITED	54	29	69	57,57	31,62	70,89
7	EVERTON	62	44	61	49,63	45,78	55,78
8	SOUTHAMPTON	41	48	46	49,98	43,08	55,69
9	BOURNEMOUTH	55	67	46	50,02	62,99	43,17
10	WEST BROMWICH	43	51	45	38,21	47,9	45,59

³⁵ <https://understat.com>

11	WEST HAM	47	64	45	48,25	59,67	46,04
12	LEICESTER	48	63	44	44,34	51,3	48,12
13	STOKE CITY	41	56	44	47,87	48,67	52,28
14	CRYSTAL PALACE	50	63	41	45,4	50,03	48,45
15	SWANSEA	45	70	41	39,5	65,13	36,49
16	BURNLEY	39	55	40	35,59	55,027	38,36
17	WATFORD	40	68	40	36,25	52,62	40,62
18	HULL	37	80	34	33,84	71,77	30,73
19	MIDDLSBRUGH	27	53	28	30,93	55,3	37,25
20	SUNDERLAND	29	69	24	34,22	65,66	32,24

Tabella 3.5: Classifica premier league stagione 2016/2017 (Fonte: UNDERSTAT).

La tabella 3.5 riporta la classifica del campionato inglese della stagione del 2016/2017, oltre a appresentare i classici dati (Goal fatti, subiti e punti effettivi) mette in evidenza gli indici analizzati in questo capitolo e nel capitolo precedente (expected goals, expected goals against ed expected points).

Come possiamo vedere dalla classifica finale, il Chelsea guidato dal tecnico italiano Antonio Conte, si laureò campione d’Inghilterra, ottenendo un punteggio molto elevato (93 punti) e con un cospicuo margine sulla seconda classificata (+6 sul Tottenham).

Mentre la squadra Manchester City, a detta di molti favorita per la vittoria del titolo si piazzò “solamente” al terzo posto, andando al di sotto delle aspettative.

Osservando i dati pubblici, o per lo meno quelli resi visibili dalle trasmissioni sportive, essi legittimano la vittoria del Chelsea, infatti fu una delle squadre che marcarono il maggior numero di goal (85) e mostrò anche una gran solidità difensiva, subendo solo 33 goal.

Facendo una riflessione più nello specifico, e grazie agli indici studiati nel corso di questa relazione possiamo notare come il Chelsea abbia *over-performato* da un punto di vista offensivo realizzando 85 reti a fronte delle 62 attese.

Per quanto riguarda il Manchester City è evidente che durante la stagione abbia subito più goal di quanti ne sarebbe stato lecito aspettarsi.

Anche per quanto riguarda il “nuovo” indice xTS possiamo osservare una netta contrapposizione dei due club.

POSIZIONE	TEAM	G GOAL	AGAINST	POINTS	XG XGOAL	AGAINST	XPTS	PTS DIFF
1	MANCHESTER CITY	80	39	78	79,85	29,61	85,41	-7,41
2	CHELSEA	85	33	93	61,8	28,62	75,74	17,26
3	TOTTENHAM	86	26	86	70,07	33,78	75,37	10,63
4	MAN UNITED	84	29	69	57,57	31,62	70,89	-1,89
5	LIVERPOOL	78	42	76	66,63	37,65	69,83	6,17
6	ARSENAL	77	44	75	63,58	47,07	62,12	12,88

Tabella 1.6: Classifica premier league basata attraverso xPTS (Fonte: nostre elaborazioni).

La tabella 3.6 mostra la classifica basata attraverso xPTS, (qui riporto solo uno spaccato delle prime posizioni) possiamo osservare come il Manchester City da terzo classificato sia in vetta alla graduatoria e il Chelsea da primo assoluto si piazza secondo con approssimativamente lo stesso punteggio del Tottenham (circa 75 punti).

Come abbiamo già introdotto nella parte iniziale del capitolo, l'indice xPTS per quanto concerne le squadre di testa è presumibile che sia minore dei punti effettivi.

Quindi ancor più della differenza del Chelsea di ben 17 punti, è fondamentale segnalare come il Manchester City abbia una differenza punti negativa di oltre 7 punti.

Questa analisi ci porta a dire che il Chelsea nel corso di questa stagione ha ottenuto più punti di quelli attesi, ma sicuramente il Manchester City ha raccolto meno di quello che ha prodotto durante la stagione.

Abbiamo detto che l'indice aiuta sia gli allenatori che i direttori sportivi per programmare e prevedere la stagione successiva, ed ora andiamo a vedere l'importanza specifica di tali dati.

Il direttore sportivo del Manchester city durante la “finestra” degli scambi ha acquistato un nuovo portiere, ritenendo che l'*under performance* difensiva della squadra fosse dovuta alle scarse qualità tecniche dell'estremo difensore. Mentre i dirigenti del Chelsea, decisero di sostituire il loro attaccante principale, questa scelta si rivelò completamente errata da un punto di vista statistico, perché cambiare il centroavanti di una squadra che ha segnato molto di più di quanto atteso?

POSIZIONE	TEAM	G			XG		
		GOAL	AGAINST	POINTS	XGOAL	AGAINST	XPTS
1	MANCHESTER CITY	106	27	100	91,43	24,51	91,09
2	MAN UNITED	68	28	81	59,04	43,54	62,33
3	TOTTENHAM	74	36	77	68,91	35,86	76,02
4	LIVERPOOL	84	38	75	77,49	35,75	79,38
5	CHELSEA	62	38	70	59,16	34,03	68,46
6	ARSENAL	74	51	63	72,27	48,75	65,9

Tabella 3.7: Classifica Premier league stagione 2017/2018 (Fonte: UNDERSTAT).

La stagione successiva mostra come l'indice xPTS sia stato “buon profeta” a livello di previsione e utile consigliere ai direttori.

Come avevamo detto in precedenza, il Manchester City, già nella stagione precedente avrebbe meritato il titolo. Supportato poi dalla buona decisione del cambio portiere ha ridotto il numero di goal subiti, inoltre durante la stagione 2017/2018 a differenza del campionato 2016/2017 i goal subiti sono stati in linea con l'xGA (goal effettivamente subiti 27, indice xGA 24,51).

Per quanto riguarda il Chelsea, già gli xPTS ci avevano segnalato una possibile flessione nel corso della stagione successiva (ricordiamo che la classifica rielaborata poneva la squadra londinese al secondo/terzo posto a pari merito con il Tottenham), inoltre, come sospettato precedentemente il

cambio di attaccante non è stato una mossa vincente, infatti, i goal attesi delle due stagioni sono rimasti pressochè identici, ma a differenza dell'anno precedente i goal effettivamente realizzati sono stati ben 23 in meno.

Il secondo esempio che andiamo ad analizzare attraverso l'indice xPTS è il caso Manchester United durante la stagione 2018/2019.

La squadra allenata dal portoghese Josè Mourinho reduce dal secondo posto nella stagione precedente iniziò il campionato malissimo, tanto che la società decise di esonerare il tecnico già a metà stagione.

A sorpresa di tutti, i dirigenti, decisero di affidare la squadra a Gunnar Solskear, fino ad allora allenatore delle giovanili del Manchester United.

Il neo allenatore collezionò una serie di vittorie importanti, tanto che diventò il primo allenatore della storia del club a vincere tutte le prime 7 partite giocate.

La società a fronte di queste continue vittorie si decise a offrirgli un contratto di 4 anni con un ingaggio annuo da top allenatore.³⁶

Sfortunatamente per la squadra, dopo questa sequenza di vittorie e di risultati positivi iniziarono i problemi e una serie di sconfitte che fecero concludere il Manchester United al sesto posto.

Grazie all'aiuto degli xPTS andremo ad analizzare come la flessione della squadra fosse pronosticabile, e come la scelta dei dirigenti avrebbe dovuto essere fatta in maniera più attenta e non solamente basata sulle vittorie iniziali.

POSIZIONE	TEAM	POINTS	XPTS
1	MANCHESTER UNITED	29	21,09
2	LIVERPOOL	24	24,26
3	MANCHESTER CITY	24	25,05
4	ARSENAL	22	17,95
5	TOTTENHAM	21	18,79

Tabella 3.8: Classifica premier league dal 21/12/2018 al 28/02/2019 (Fonte: nostre Elaborazioni).

³⁶ https://it.wikipedia.org/wiki/Ole_Gunnar_Solskj%C3%A6rhttps://understat.com/

La tabella 3.8 mostra i punti effettivi e gli xPTS dal 21/12/2018 (prima partita del nuovo allenatore) al 28/02/2019 (prima sconfitta sotto la gestione Solskjaer).

In questi due mesi il lavoro del neo tecnico è stato sicuramente positivo, sia i punti effettivi che l'xPTS del Manchester United sono elevati, anche se è chiaro che siamo di fronte ad una evidente *over performance*.

Infatti, una differenza di quasi 8 tra punti reali ed attesi nell'arco di sole 11 partite significano che in media la squadra di Manchester ha ottenuto quasi un punto in più a partita rispetto alle aspettative.

Questo segnale avrebbe dovuto far porre delle domande ai dirigenti sul reale andamento della squadra, i quali, o avrebbero dovuto rinforzare la rosa nel periodo di calciomercato invernale (01.01.2019 – 31.01.2019), oppure aspettare la fine della stagione prima di proporre un contratto lungo e molto oneroso al nuovo tecnico.

POSIZIONE	TEAM	POINTS	XPTS
1	MANCHESTER CITY	30	24,55
2	LIVERPOOL	28	21,78
3	CHELSEA	19	21,24
4	CRYSTAL PALACE	19	13,74
...	...	...	...
14	MANCHESTER UNITED	11	13,84

Tabella 3.9: Classifica premier league dal 05/03/2019 al 07/05/2019 (Fonte:nostre elaborazioni).

La tabella 3.9 riporta i punti e l'indice xPTS dell'ultima parte della stagione, dal 05/03/2019 al 07/05/2019.

Come prevedibile la squadra in queste ultime 10 partite non ha mantenuto il rendimento delle prime 11.

Nel periodo qui analizzato, con soli 11 punti in 10 partite il Manchester United ha avuto un rendimento da squadra di bassa classifica, comparando l'indice xPTS con i punti realizzati la differenza è solamente di circa -3 punti (11 reali, circa 14 quelli attesi).

Ciò significa che la squadra non ha attraversato solamente un momento sfortunato, ma la coerenza tra punti reali e xPTS ci mostra come questo piazzamento sia stato meritato.

CAPITOLO 4:

ANALISI STAGIONE CALCISTICA 2017/2018

4.1 Premessa iniziale

In questo capitolo finale verranno analizzate le performance delle squadre partecipanti al campionato italiano durante la stagione 2017/2018, inizialmente considerando alcune situazioni chiave (possesso palla, capacità offensiva e forza difensiva); per poi passare ad una rielaborazione della classifica di serie A attraverso l'indice xPTS, con particolare riferimento a due squadre: Juventus e Napoli.

Il calcio italiano sta vivendo una fase storica unica, infatti dalla stagione 2011/2012 ad oggi la società Juventus ha conquistato nove campionati consecutivi, nella maggior parte dei casi con un ampio margine di vantaggio sulla seconda classificata.

Ho scelto di analizzare l'annata 2017/2018 perché secondo la mia opinione è stata la più interessante da un punto di vista statistico, infatti, per la prima volta dopo alcuni anni l'assegnazione del campionato ha visto sfidarsi in un "testa a testa" due grandi squadre, la Juventus, che ne è uscita vittoriosa, e il Napoli, sconfitto (Tabella 4.1).

Posizione	Squadra	Punti
1	Juventus	95
2	Napoli	91
3	Roma	77
4	Inter	72
5	Lazio	72
6	Milan	64
7	Atalanta	60
8	Fiorentina	57
9	Torino	54
10	Sampdoria	54
11	Sassuolo	43
12	Genoa	41
13	Udinese	40
14	Chievo	40
15	Bologna	39
16	Cagliari	39
17	SPAL 2013	38
18	Crotone	35
19	Verona	25
20	Benevento	21

Tabella 4.1: Classifica stagione 2017/2018 (Fonte: Lega serie A).

4.2 Il Possesso palla

Nel calcio si possono individuare due “filosofie” di gioco, che contrappongono squadre definite proattive, ovvero con un’elevata percentuale di possesso palla, e squadre reattive, caratterizzate, invece, da una percentuale inferiore.

Ma esiste una correlazione tra possesso palla e risultati?

Da un report del CIES Football Observatory condotto durante la stagione 2016/2017 considerando 35 campionati differenti, è stato rilevato che le squadre vincitrici dei rispettivi tornei hanno registrato un possesso palla medio del 55%.

L’indagine evidenzia come una strategia difensivista non porti risultati sul lungo periodo.

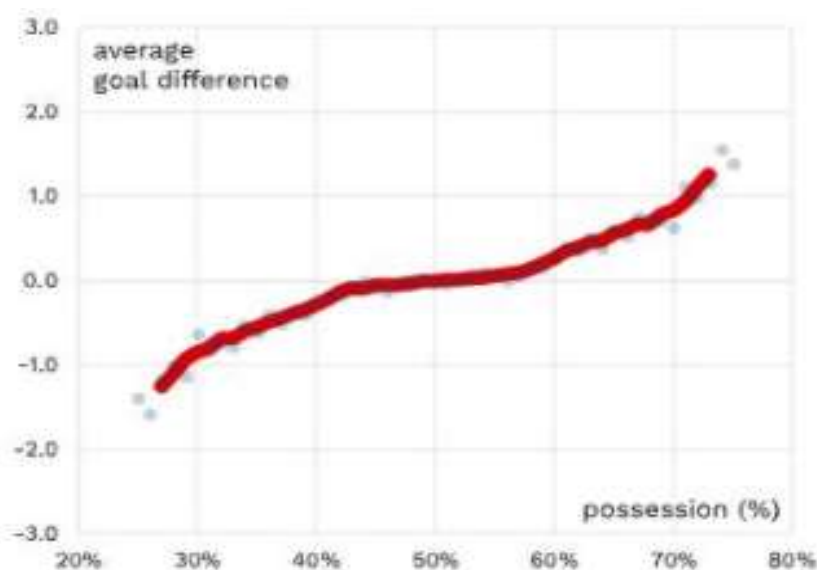


Figura 4.1: Percentuale di possesso palla e differenza reti di 35 campionati (Fonte: CIES)

Il grafico di Figura 4.1 mostra che a valori elevati della percentuale di possesso palla si ha in media un impatto positivo sulla differenza reti durante una partita.

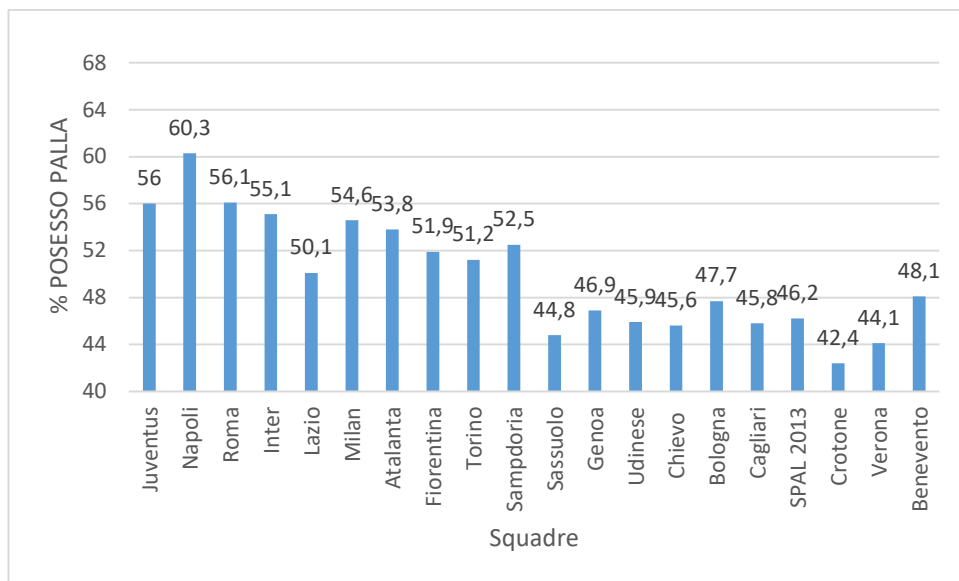


Figura 4.2: Percentuale di possesso palla di ogni squadra durante la stagione 2017/2018 (Fonte: nostre elaborazioni).

Anche per quanto riguarda il campionato preso in considerazione in questa relazione finale si può notare dal grafico di Figura 4.2 (in cui le squadre sono ordinate secondo la classifica finale) una certa relazione tra percentuale di possesso e piazzamento finale.

A spiccare in vetta di questa speciale graduatoria c'è il Napoli con una percentuale di oltre il 60% mentre la Juventus si stabilizza al 56%.

Il dato sul possesso palla complessivo può darci solo un'indicazione parziale della pericolosità di una squadra. Grazie ai dati messi a disposizione dal sito *"fbref.com"* ho ricavato percentuale di possesso palla che viene prodotta nella metà campo offensiva (definita zona pericolosa).

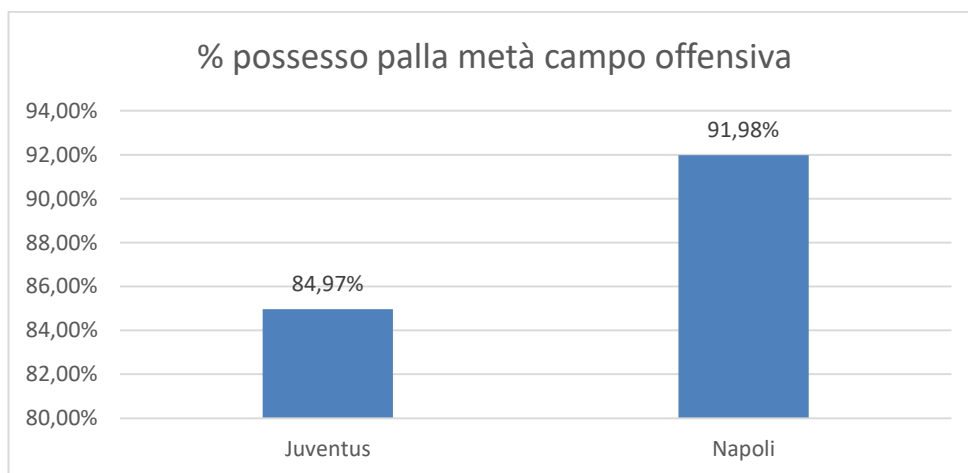


Figura 4.3: % percentuale possesso palla nella metà campo offensiva rispetto al totale (Fonte: nostre elaborazioni, dati forniti da fbref.com).

La figura 4.3 mostra come il Napoli abbia prodotto circa il 92% del possesso palla complessivo nella metà campo avversaria. (Rispetto alla Juventus che si ferma all'85%)

Unendo l'analisi dei grafici 2 e 3 possiamo giungere alla conclusione che la Squadra Napoli durante la stagione 2017/2018 ha realizzato sia un maggior possesso palla in termini assoluti sia un possesso di maggiore pericolosità.

4.3 Valutazione capacità offensiva ed xG

La valutazione della capacità offensiva di una squadra può essere interpretata studiando molteplici dati; in questa relazione ho scelto di effettuare questa analisi mettendo al centro della discussione l'evento "tiro".

Squadre	Goal	TIRI	% Gol/tiri
Juventus	82	546	15,0%
Napoli	73	651	11,2%
Roma	56	667	8,4%
Inter	64	584	11,0%
Lazio	86	557	15,4%
Milan	54	633	8,5%
Atalanta	57	582	9,8%
Fiorentina	53	624	8,5%
Torino	53	460	11,5%
Sampdoria	56	474	11,8%
Sassuolo	28	496	5,6%
Genoa	31	421	7,4%
Udinese	45	445	10,1%
Chievo	36	396	9,1%
Bologna	37	353	10,5%
Cagliari	33	378	8,7%
SPAL 2013	37	374	9,9%
Crotone	38	385	9,9%
Verona	30	362	8,3%
Benevento	32	470	6,8%
MEDIA	49,05	492,9	9,9%

Tabella 4.2: Goal, tiri totali e rapporto percentuale goal/tiri (Fonte:UNDERSTAT).

La tabella 4.2 mostra i goal, i tiri totali (tiri verso la porta più tiri fuori esclusi i rigori) e il rapporto goal/tiri.

Quest'ultimo indice si pone l'obiettivo di individuare quanto una squadra sia stata "cinica" sotto porta avversaria.

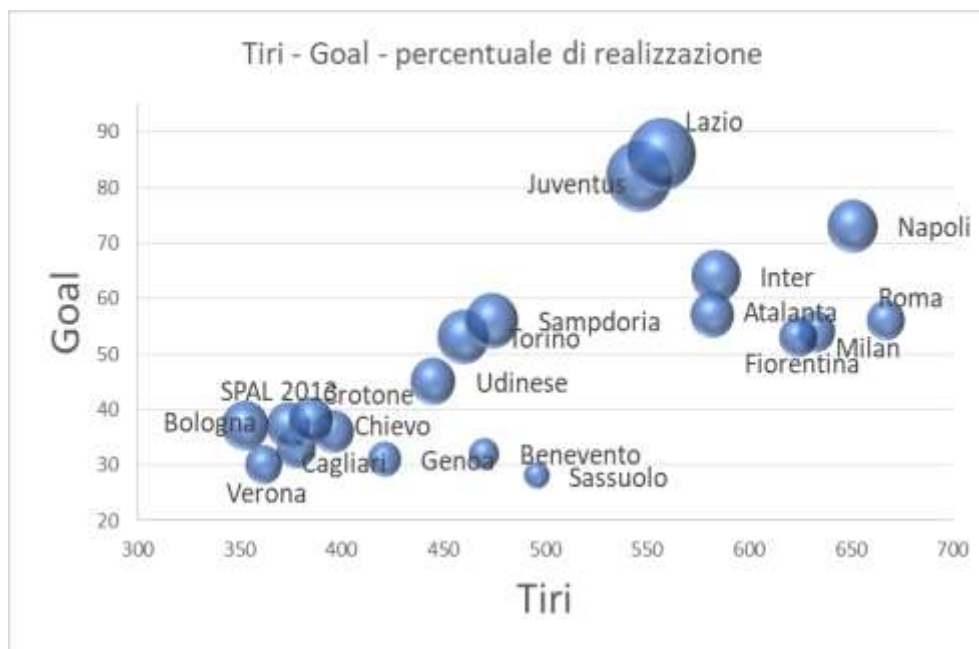


Figura 4.4: Tiri - Goal - percentuale di realizzazione (Fonte: nostre elaborazioni).

Il grafico a dispersione in figura 4.4 permette di visualizzare concretamente i dati della tabella 4.2

La collocazione delle squadre all'interno del piano cartesiano dipende dalle due variabili; Tiri (asse delle ascisse) e Goal (asse delle ordinate), mentre la grandezza del "punto" è determinata in proporzione alla percentuale di realizzazione.

La Lazio con 86 Goal realizzati è stata la squadra più prolifica della stagione (detenendo, inoltre, il primato sul rapporto goal/tiri, pari al 15.4%), non a caso il suo miglior attaccante *Ciro Immobile* si è aggiudicato la classifica marcatori con 29 goal.

Il Sassuolo con solo 28 goal ha praticato il peggior attacco del campionato nonostante i tiri in porta siano stati 496 (media campionato di 492).

Detto ciò non ci stupisce vedere che la squadra emiliana ha prodotto il più basso rapporto Goal/tiri con il solo 5,6%.

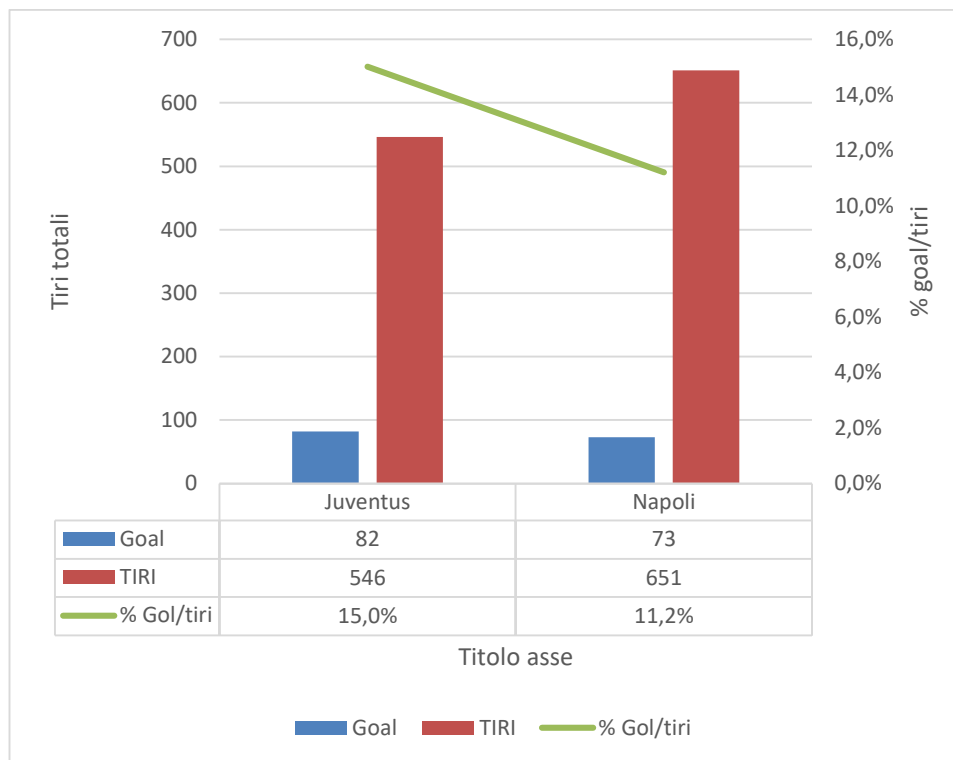


Figura 4.4: Confronto capacità offensiva di Juventus e Napoli (Fonte: nostre elaborazioni).

Dal grafico 4 si può notare la forza della Juventus nel trasformare in reti i tiri effettuati, registrando un tasso goal/tiri del 15% (secondo migliore del campionato).

Il Napoli, nonostante un maggior numero di tiri ha realizzato 9 goal in meno della squadra torinese, mostrando difficoltà a convertire in goal le occasioni prodotte.

Come abbiamo già affermato nel capitolo 2, non tutti i tiri hanno la stessa probabilità di essere convertiti in goal, ecco perché è utile effettuare un'analisi tramite xG per ottenere delle informazioni più precise rispetto al rapporto Goal/tiri.

Squadre	Goal	XG	DELTA
Juventus	82	59,23	22,77
Napoli	73	70,45	2,55
Roma	56	65,63	-9,63
Inter	64	60,83	3,17
Lazio	86	66,51	19,49
Milan	54	50,88	3,12
Atalanta	57	64,29	-7,29
Fiorentina	53	54,12	-1,12
Torino	53	46,59	6,41
Sampdoria	56	53,14	2,86
Sassuolo	28	47,15	-19,15
Genoa	31	42,21	-11,21
Udinese	45	45,2	-0,2
Chievo	36	34,42	1,58
Bologna	37	36,11	0,89
Cagliari	33	38,27	-5,27
SPAL 2013	37	34,61	2,39
Crotone	38	33,24	4,76
Verona	30	33,76	-3,76
Benevento	32	38,21	-6,21

Tabella 4.8: Goal vs xG (Fonte: nostre UNDERSTAT).

La tabella 4.3 conferma attraverso un'analisi xG che il Sassuolo, con un Delta di - 19.15, è la squadra con la più elevata under-performance offensiva del campionato.

Anche la Roma evidenzia problematiche in zona goal (Delta pari a - 9,63); già nella tabella 2 si poteva notare un posizionamento del suo rapporto goal/tiri (8.4%) sotto la media del campionato (9.9%).

Lazio e Juventus si confermano le squadre più “spietate” del campionato con un importante *over-performance* offensiva (rispettivamente +19,49 e + 22,77).

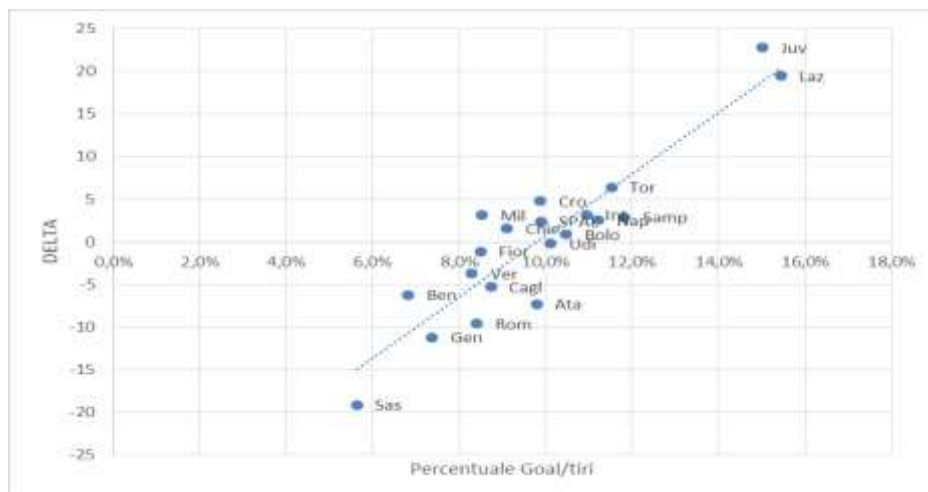


Figura 4.5: Relazione tra Delta e percentuale Goal/tiri (Fonte: nostre elaborazioni).

Dalla *trendline* del grafico in Figura 4.5 constatiamo una certa correlazione lineare positiva tra rapporto Goal/tiri e il Delta (goal effettivi – xG).

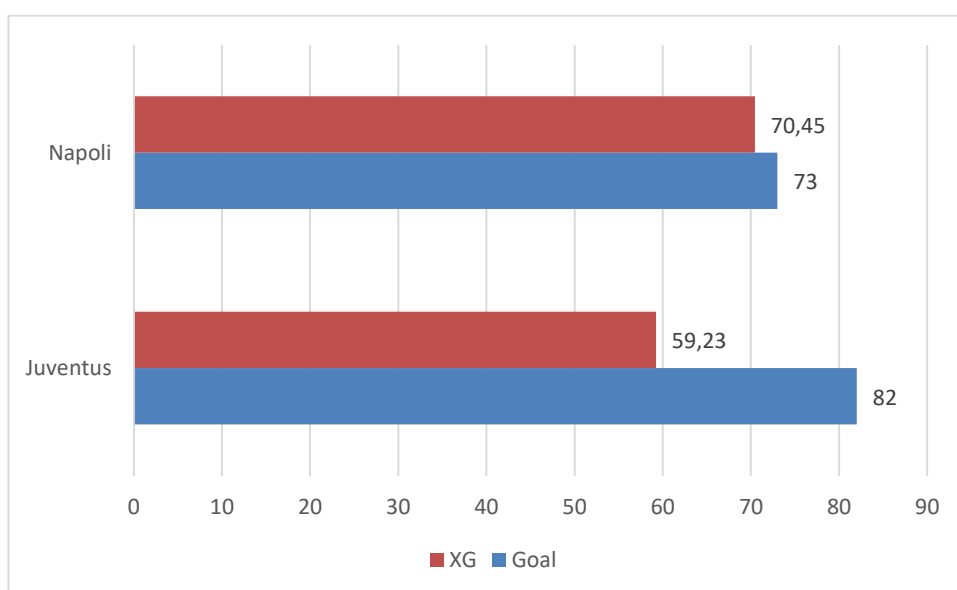


Figura 4.6: Confronto Juventus-Napoli attraverso xG (Fonte: nostre elaborazioni).

Il grafico 4.6 mostra chiaramente la differenza di performance offensiva tra Juventus e Napoli; Da cosa è stata determinata?

Per provare a dare una risposta ho analizzato i dati dei giocatori più rappresentativi in zona offensiva di ciascuna delle due squadre come mostrato in Tabella 4.4

Giocatore	Squadra	Goal	XG	DELTA
Paulo Dybala	Juventus	22	13,17	8,83
Gonzalo Higuain	Juventus	16	13,41	2,59
Mario Mandzukic	Juventus	5	5,8	-0,8
Douglas Costa	Juventus	4	3,07	0,93
Jose Callejon	Napoli	10	10,09	-0,09
Arkadiusz Milik	Napoli	5	4,75	0,25
Dries Mertens	Napoli	18	18,79	-0,79
Lorenzo Insigne	Napoli	8	14,67	-6,67

Tabella 9.4: Goal ed xG dei singoli attaccanti di Juve e Napoli (Fonte: UNDERSTAT).

La Tabella 4.4 denota come se, tutto sommato, tra gli attaccanti delle due squadre non vi è una rilevante differenza di performance, per Dybala e Insigne bisogna fare delle considerazioni diverse.

Infatti, i due attaccanti, pur avendo un xG personale abbastanza simile, ovvero 13 per lo Juventusino e 14 per il napoletano, non possono essere paragonati dal punto di vista dei goal realizzati, poiché quelli di Dybala superano di ben 14 le reti dell'avversario.

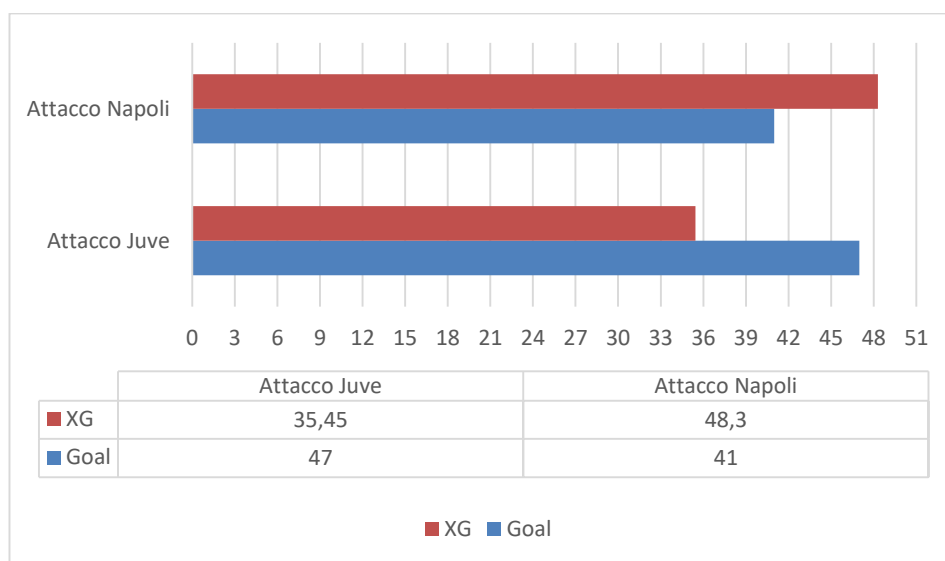


Figura 4.7: Confronto tra attacco Juventus e Napoli sulla base di xG personali (Fonte: nostre elaborazioni).

Dopo l'analisi di questi dati siamo in grado di rispondere all'interrogativo che ci siamo posti all'inizio di questo paragrafo, potendo ragionevolmente affermare che la differenza di performance a livello offensivo tra Juventus e Napoli sia stata in larga parte determinata dall'*over-performance* individuale di Dybala (e in minor misura di Higuain) e dall'*under-performance* di Insigne.

4.3 Valutazione forza difensiva.

In questa relazione finale la capacità difensiva delle squadre viene analizzata inizialmente attraverso l'indice xGA il quale si pone l'obiettivo di analizzare la forza difensiva in senso stretto (numero di tiri in porta subiti e pericolosità di tali conclusioni) e successivamente la capacità individuale del portiere calcolando la percentuale di parate su tiri in porta ricevuti.

Squadre	Goal subiti	XGA	DELTA
Juventus	24	28,58	-4,58
Napoli	29	25,27	3,73
Roma	28	38,46	-10,46
Inter	30	37,33	-7,33
Lazio	49	42,57	6,43
Milan	42	36,78	5,22
Atalanta	39	36,5	2,5
Fiorentina	46	42,15	3,85
Torino	46	49,11	-3,11
Sampdoria	60	62,26	-2,26
Sassuolo	59	50,63	8,37
Genoa	43	42,87	0,13
Udinese	63	50,71	12,29
Chievo	59	55,28	3,72
Bologna	52	48,8	3,2
Cagliari	61	57,91	3,09
SPAL 2013	59	60,18	-1,18
Crotone	66	59,43	6,57
Verona	78	72,71	5,29
Benevento	89	77,31	11,69

Tabella 4.10: Goal subiti vs xGA (Fonte: UNDERSTAT).

La Juventus con 24 goal subiti ha praticato la miglior difesa del campionato (tabella 4.5) alla quale si contrappone il Benevento che con 89 goal al passivo ha messo in atto la peggior difesa della stagione 2017/2018.

Se prendiamo in considerazione l'indice xGA non mancherebbero le sorprese, infatti il Napoli con un indicatore pari al 25.27 risulterebbe essere la squadra più solida del campionato.

Un caso di particolare interesse è quello della Roma, che, nonostante un indice di goal attesi subiti abbastanza elevato (38.46) ha incassato solamente 28 goal, piazzandosi seconda nella speciale classifica delle migliori difese.

Questa straordinaria over-performance della Roma (Delta pari a -10.46), come vedremo prossimamente, non è dovuta a una forza della linea difensiva o dall'imprecisione degli attaccanti, avversari ma all'abilità del proprio portiere Allison, che non a caso è stato premiato come il migliore del campionato 2017/2018.

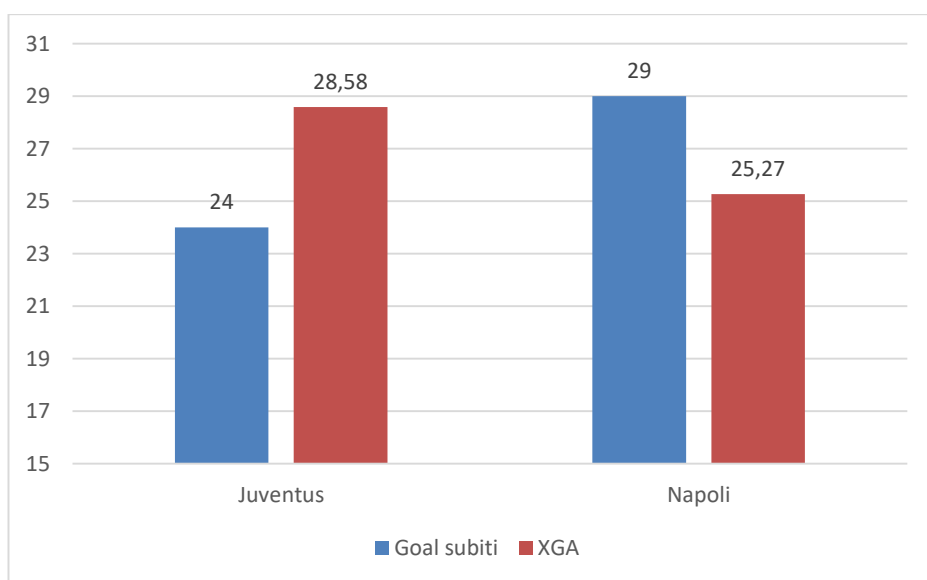


Figura 4.8: Confronto goal subiti e xGA di Juventus e Napoli (fonte: nostre elaborazioni).

Tornando al confronto tra Juventus e Napoli il grafico in figura 4.8 mostra che la vincitrice del campionato ha realizzato una *over-performance*, mentre il Napoli anche in questa occasione ha mostrato una *under-performance* a livello difensivo.

Come detto in precedenza, la fase difensiva può essere esaminata sotto due differenti aspetti, tra cui il numero di tiri subiti e il loro grado di pericolosità, che evidenziano la forza della difesa, mentre la forza del portiere è definita dal rapporto parate/tiri.

Squadre	Tiri in porta subiti	Parate	% parate/tiri
Juventus	94	73	77,7%
Napoli	99	71	71,7%
Roma	138	110	79,7%
Inter	119	94	79,0%
Lazio	153	112	73,2%
Milan	131	92	70,2%
Atalanta	118	80	67,8%
Fiorentina	152	105	69,1%
Torino	164	123	75,0%
Sampdoria	193	140	72,5%
Sassuolo	177	122	68,9%
Genoa	170	133	78,2%
Udinese	163	107	65,6%
Chievo	180	126	70,0%
Bologna	164	116	70,7%
Cagliari	162	108	66,7%
SPAL 2013	192	144	75,0%
Crotone	193	138	71,5%
Verona	257	186	72,4%
Benevento	203	130	64,0%
Media			72,0%

Tabella 4.6: Classifica tiri in porta subiti, parate e % parati/tiri (Fonte:fbref.com).

A differenza dell'analisi sulla capacità offensiva (in quel caso consideravamo tutti i tiri) la tabella 7 mostra solamente i tiri subiti verso la porta.

La tabella evidenzia una certa coerenza con l'analisi xGA.

Considerando nuovamente la Roma per cui i dati precedenti mostravano un basso numero di goal effettivi subiti ma un elevato indice di goal attesi previsti, possiamo osservare dalla tabella 6 come la tale squadra abbia subito 138 tiri verso la porta (circa 40 tiri in più rispetto a Napoli e Juventus). Ciò significa in linea generale che la difesa romanista non è stata a livello delle prime due classificate. Il fatto che, nonostante l'elevato numero di tiri subiti abbia avuto la seconda miglior difesa, può essere attribuito al merito del proprio portiere, che è riuscito a parare il 79.7% dei tiri, ampiamente sopra la media campionato (72%.)

Il Benevento è stata la squadra con più goal subiti (89, tabella 5); si può ragionevolmente pensare che questo pessimo dato sia stato causato sia dalla

linea difensiva che ha concesso 203 tiri sia dalla scarsa abilità del portiere che, con il 64% dei tiri parati, è collocato notevolmente sotto la media.

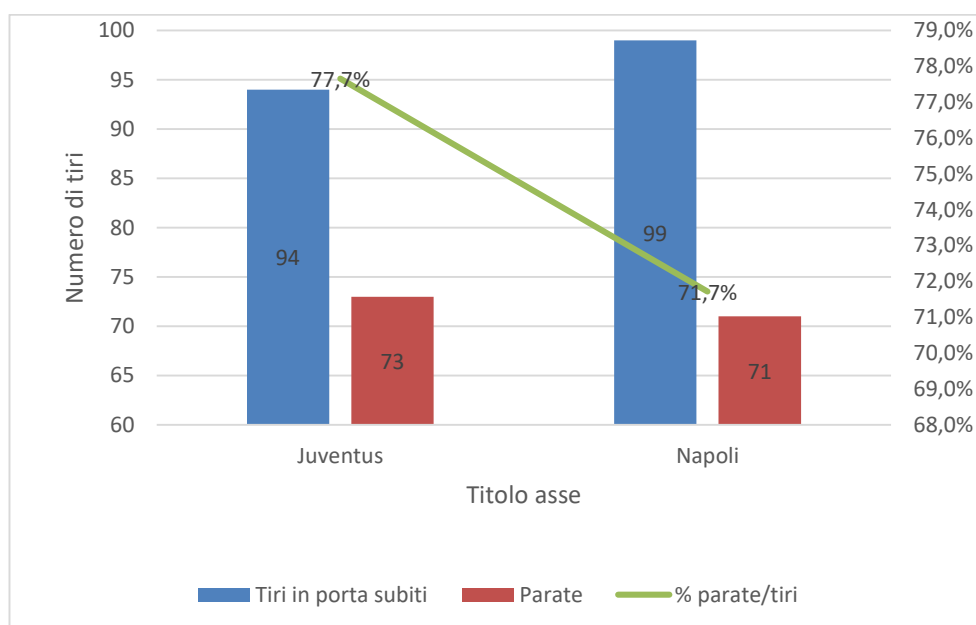


Figura 4.9: Confronto capacità difensiva tra Juventus e Napoli (Fonte: nostre elaborazioni).

Durante l'analisi del grafico 4.8 si era detto che la Juventus aveva realizzato una *over-performance* difensiva al contrario del Napoli.

Il grafico in figura 4.9 palesa come il portiere del Napoli abbia una percentuale di parate non solo minore rispetto al portiere avversario ma anche rispetto alla media del campionato.

Ciò potrebbe ragionevolmente essere stata la causa principale dell'*under-performance* difensiva.

Al contrario, il portiere Juventino con il 77.7 % di parate ha determinato l'*over-performance*.

Attraverso l'analisi dei dati si può arrivare alla conclusione che, anche in questo caso la bravura del singolo ha influito notevolmente sul risultato finale.

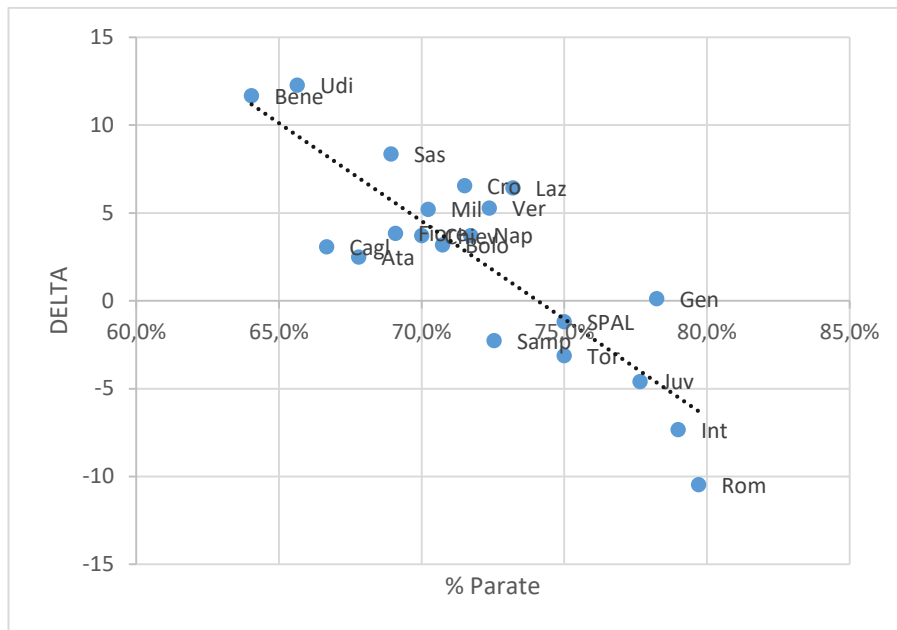


Figura 4.10: Relazione tra Percentuale di parate e Delta (Fonte: nostre elaborazioni).

Il grafico in figura 4.10 mette in relazione la percentuale di parate e il Delta (goal effettivamente subiti – xGA).

Tracciando la linea di tendenza, si nota (in maniera più debole rispetto alla valutazione fatta sulla fase offensiva) una correlazione lineare negativa tra le due variabili.

4.4 Rielaborazione classifica finale attraverso l'indice xPTS

Posizione	Squadre	XPTS	DELTA (XPTS- PTS)	Posizione classifica reale
1	Napoli	82,23	-8,77	2
2	Juventus	73,51	-21,49	1
3	Atalanta	69,88	9,88	7
4	Roma	69,24	-7,76	3
5	Lazio	66,05	-5,95	5
6	Inter	65,25	-6,75	4
7	Milan	62,35	-1,65	6
8	Fiorentina	60,41	3,41	8
9	Torino	51,84	-2,16	9
10	Sassuolo	51,7	8,7	11
11	Genoa	50,38	9,38	12
12	Udinese	47,34	7,34	13
13	Sampdoria	46,87	-7,13	10
14	Bologna	43,01	4,01	15
15	Cagliari	40,41	1,41	16
16	Chievo	38,66	-1,34	14
17	Crotone	35,44	0,44	18
18	SPAL	35,41	-2,59	17
19	Verona	31,06	6,06	19
20	Benevento	30,09	9,09	20

Tabella 4.7: Classifica rielaborata attraverso xPTS (Fonte: UNDERSTAT).

La rielaborazione della classifica secondo l'indice xPTS è stata realizzata attraverso la simulazione delle realizzazioni delle distribuzioni di probabilità di Bernoulli relative ai singoli tiri, applicata con il metodo Monte Carlo descritto nel terzo capitolo.

La tabella 4.7 evidenzia i principi teorici descritti nel capitolo precedente. Infatti, come già accennato, l'indice xPTS tende ad essere minore rispetto ai punti effettivamente realizzati dalle squadre di fascia alta, mentre accade il contrario per le squadre posizionate nelle zone bassa di classifica.

In particolar modo, esaminando le squadre che hanno ottenuto il diritto di accesso alle coppe europee (prime 7 classificate) solamente l'Atalanta ha evidenziato un saldo positivo tra punti attesi ed effettivi.

Al contrario, se si considerano esclusivamente le squadre retrocesse in serie B (ultime 3 classificate) si evidenzia la prevalenza di un Delta positivo o comunque prossimo al valore 0.

Coerentemente all'analisi proposta in questo capitolo, non ci sorprende veder ribaltate le prime due posizioni in classifica.

Tenendo presente che l'indice xPTS si pone l'obbiettivo di rielaborare la classifica in relazione alle performance e non sui risultati effettivi, possiamo affermare che il Napoli rispetto alla Juventus sotto il punto di vista della prestazione di squadra è stato superiore in tutte le situazioni esaminate.

In particolare, per quanto riguarda la fase offensiva, il Napoli ha realizzato un xG complessivo superiore a quello Juventino, ciò significa, che la squadra napoletana ha creato più occasioni statisticamente favorevoli rispetto agli avversari.

Anche in zona difensiva, attraverso l'analisi dell'indice xGA, il Napoli si è mostrato migliore, concedendo complessivamente meno “buone” occasioni agli avversari rispetto alla Juventus.

Supponendo che la componente fortuna all'interno di un intero campionato agisca come una variabile a media nulla (a volte a favore, altre contro) e che quindi, le eventuali *over/under-performance* dei calciatori siano attribuibili a fattori di gioco e non a elementi derivanti dalla favorevole o avversa sorte; Si può ragionevolmente giungere alla conclusione che, nonostante la supremazia a livello collettivo del Napoli, le singole capacità individuali (in attacco e del portiere) dei calciatori Juventini hanno determinato l'esito della competizione.

Perché, se è vero che il calcio è uno sport di squadra, anche le capacità individuali di ogni calciatore possono decidere partite e campionati.

CONCLUSIONE

In questa relazione finale è stato approfondito il tema della statistica applicata al calcio, al fine di misurare le *performance* di giocatori e squadre.

Lo scopo del lavoro era comprendere il funzionamento e il significato dell'indice xG , ovvero la probabilità di trasformarsi in rete di un determinato tiro, e dei suoi derivati (xGA , xGD e $xPTS$).

Il lavoro è stato suddiviso in quattro capitoli: il primo tratta dell'importanza del binomio statistica-sport, con particolare attenzione al gioco del basket e alla sua evoluzione nel mondo del calcio.

Nel secondo capitolo vengono definiti gli indici xG , xGA e xGD , con i quali, per la prima volta, si ha la possibilità di attribuire un coefficiente di realizzazione ad ogni singolo tiro, differenziandosi dalle tradizionali statistiche che, invece, si limitano a conteggiare i tiri.

Questo permette di venire a conoscenza del numero di goal che il calciatore o la squadra media avrebbe dovuto realizzare, secondo l'indice xG , e subire, attraverso l' xGA , e la differenza reti attesa, calcolata tramite l' xGD .

Il terzo capitolo illustra l'indicatore $xPTS$, ovvero i punti attesi ad ogni partita, a scopo illustrativo sono stati simulati, mediante il metodo Monte Carlo, alcuni scenari attraverso il foglio di lavoro Excel, per mostrare il funzionamento e l'andamento di questo indicatore.

Il quarto ed ultimo capitolo presenta un'analisi della stagione calcistica 2017/2018, focalizzando, principalmente, l'attenzione sulla "lotta scudetto" tra Juventus e Napoli, con l'obiettivo di elaborare una classifica basata sulle prestazioni e non sui risultati effettivi.

Sono state individuate tre situazioni di gioco: possesso palla, valutazione offensiva (esaminata tramite l'indice xG) e capacità difensiva (valutata attraverso l'indice xGA), grazie alle quali indagare sulle cause che hanno portato ad uno scostamento tra la classifica reale e quella attesa.

Questo lavoro mi ha consentito di trarre le seguenti conclusioni: l'indice xG , nonostante abbia rappresentato un enorme passo in avanti nella statistica applicata al calcio, non è ancora in grado di fornire dati sui quali fare completamente affidamento, questo perché esistono più modelli, che considerano differenti variabili, portando a risultati diversi.

Inoltre, il problema principale, che accomuna tutti i modelli, risiede nell'escludere la pressione difensiva al momento del tiro, ossia il posizionamento degli avversari.

L'indice xG viene utilizzato, anche, come base per il calcolo dell'indicatore $xPTS$, per cui le limitazioni appena descritte si riflettono, conseguentemente, nei risultati ottenuti da quest'ultimo.

Affinché questo indice possa rappresentare una vera svolta nel mondo della statistica applicata al calcio, sarebbe opportuno, come nell'NBA, monitorare, secondo dopo secondo, la tracciabilità di tutti e 22 giocatori in campo.

Firma

A handwritten signature in dark ink, reading "Andrea Morosdi". The signature is written in a cursive, flowing style.

SITOGRAFIA

A.C.MILAN, *Le sedi*,
<https://www.acmilan.com/it/>.

Casotti F., *Cosa sono gli Expected Goals*”, in “Goal Italia”, 2018
<https://www.goal.com/it/notizie/cosa-sono-gli-expected-goals-domande-e-risposte/p53bt2s14p2m1gdb5rfq42pc5>.

Diffentgame, *xG model*, 3 Aprile 2019,
<https://differentgame.wordpress.com/xg-shot-maps-and-tables/>.

Giacobbe A., *Un nuovo modello di Expected Goals*, in “Ultimo uomo”, 8 Settembre 2016,
<https://www.ultimouomo.com/expected-goals/>.

Giacobbe A., *Quali sono le nuove statistiche nel calcio?* in “Ultimo uomo”, 29 Marzo 2019,
<https://www.ultimouomo.com/statistiche-calcio-novita/>.

Jeckelen K., *FBref.com*,
<https://fbref.com/it/>.

Kuper S., *Storie di calcio*, in “Calcionomica”,
<https://storiedicalcio.altervista.org/blog/calcio-e-numeri.html>.

Magro G., *“Brescia la discesa dalla A? Ingiusta lo dice l’algoritmo”* in “Giornale di Brescia”, 24 Maggio 2012,
<https://bodai.unibs.it/bdsports/wp-content/uploads/sites/2/2021/01/GdB2012.pdf>.

Mincuo A., *Prestazioni vs risultati* in “Calcio e Finanza”, 19 Marzo 2018
<https://www.calcioefinanza.it/2018/03/19/statistica-applicata-al-calcio-expected-goals/>.

MIT SLOAN ANALYTICS CONFERENCE
<https://www.sloansportsconference.com/>.

NBA, *Statistiche individuali*,
<https://it.global.nba.com/statistics/>.

OPTA SPORTS,
<https://www.optasports.com/>.

Piccione L., *Si può misurare il calcio?* In “Rivista Undici”, 3 Aprile 2019,
<https://www.rivistaundici.com/2019/04/03/calcio-statistiche/>.

Wikipedia, *L'arte di vincere*,
https://it.wikipedia.org/wiki/L%27arte_di_vincere.

Wikipedia, *Jaap Stam*,
https://it.wikipedia.org/wiki/Jaap_Stam.

Wikipedia, *Teorema del limite centrale*,
https://it.wikipedia.org/wiki/Teoremi_centrali_del_limite.

Zuccolotto P., Manisera M., *Scuola dello sport*, xxxvii,117, pag. 41.
<https://bodai.unibs.it/bdsports/>.

RINGRAZIMENTI

Giunti alla fine vorrei riservare un ringraziamento particolare alla mia relatrice Zuccolotto Paola, che mi ha seguito con infinita disponibilità e pazienza, fornendomi indispensabili consigli e conoscenze durante tutto il percorso di stesura dell'elaborato.

Infine, un grazie va alla mia famiglia, che mi ha supportato sia moralmente sia economicamente nel mio percorso accademico.