



CALCIO E STATISTICA

Chi ha detto che calcio e statistica
non possono coesistere?

Andrea Moscheni, Matteo Parise, Lorenzo Luigi Scarsella, Leonardo Perozzo



Chi ha detto che calcio e statistica non hanno niente in comune? In questa relazione vogliamo mostrare come si possano utilizzare i metodi propri della statistica per effettuare previsioni su alcuni fenomeni legati al calcio, quali il numero di reti segnate complessivamente per partita.

Mostreremo, inoltre, come vengono decise le quote nelle scommesse legate agli eventi calcistici.





Alternanza scuola-lavoro

L'alternanza scuola-lavoro (A.S.L.) – introdotta con la cosiddetta “Buona Scuola” – ha tra le sue finalità quella di indirizzare la didattica verso un metodo di apprendimento in sintonia con le esigenze del mondo esterno. Nel fare questo coinvolge anche gli adulti, siano essi insegnanti (tutor interni) o referenti della realtà ospitante (tutor esterni).

È importante sottolineare che i partner educativi esterni non sono obbligatoriamente imprese ed aziende; possono essere, infatti, anche associazioni sportive, enti culturali e istituzionali e ordini professionali. L’obiettivo è quello di sviluppare in modo condiviso alcune esperienze coerenti alle attitudini e alle passioni degli studenti.

I percorsi realizzati durante l’A.S.L. sono co-progettati, attuati, verificati ed opportunamente valutati sulla base di opportune convenzioni con imprese, associazioni di rappresentanza, con enti pubblici e privati, inclusi quelli del terzo settore, che siano disposti ad accogliere gli studenti per periodi di apprendimento in situazione lavorativa, che non si configurano, però, in rapporto individuale di lavoro (D.Lgs. 15/4/2005, n. 77).

I licei

Da molti anni ormai, il mondo degli istituti tecnico-professionali si è affacciato all’esterno della scuola, investendo tempo ed energie in percorsi formativi che prevedessero collaborazioni con enti ed aziende del territorio. La novità, potremmo dire una vera rivoluzione copernicana, è l’estensione di questo tipo di formazione anche ai licei, per i quali però l’obiettivo da perseguire è anche quello di orientare gli studenti verso la prosecuzione dei loro studi. Le attività sono obbligatorie per tutti gli studenti e, nel caso dei licei, prevedono attualmente lo svolgimento di 200 ore nell’arco del triennio.

Chi è lo Studente RicercAttore?

È innanzitutto uno *Studente* che si affaccia per la prima volta nel mondo della *Ricerca Scientifica* e vi recita un ruolo da *Attore protagonista*.

Fare scienza non significa chiaramente solo studiare modelli matematici preconfezionati da altri, ma “sporcarsi le mani” raccogliendo dati ed elaborandoli al fine di ottenere delle conclusioni che confermino o smentiscano un’ipotesi preliminare di ricerca.

Come spesso accade per un ricercatore scientifico, può succedere che lo studente in ASL abbia la sensazione di brancolare nel buio, faticosi a comprendere il significato dei dati in suo possesso; ma non è mai solo, lavora in equipe e collabora con altri, discute, dibatte.

Qui si innesta anche l’azione del tutor che deve riuscire a fornire una stampella in caso di necessità, senza intervenire direttamente nelle dinamiche del gruppo orientandone la ricerca; deve fornire stimoli e strumenti affinché gli studenti sviluppino durante il progetto di ASL le necessarie competenze d’indagine, risvegliando in loro la curiosità e sostenendone la fiducia di poter capire.

È necessario risvegliare anche la creatività in campo scientifico, cioè aggredire i problemi utilizzando – al bisogno – approcci diversi, attivare connessioni nel pensiero, renderlo flessibile ed in grado di adattarsi alle esigenze del progetto di ricerca, formulare ipotesi e validarle alla luce dei dati raccolti.

È chiaro che, per poter ottenere un simile risultato, non ci si può basare su rigidi schematismi, ma si devono stimolare i ragazzi a formulare congetture; e questo non lo si può certo inventare in pochi minuti: è necessario che la didattica venga rimodellata perché i ragazzi si abituino a proporre idee e nuovi punti di vista.

L’evoluzione scientifica sempre più frenetica che permea di sé la società moderna ci spinge a promuovere una cultura scientifica basata sulla ricerca, l’informazione e la conoscenza.

La scuola deve agire per consolidare quanto più possibile le conoscenze scientifiche degli studenti, per consentire loro di orientarsi al meglio in un mondo che cambia così rapidamente.

Per poterci riuscire non si può rinunciare a ricorrere alla Statistica, una disciplina che acquisisce sempre maggiore rilievo vista la necessità di elaborare grandi moli di dati in vari ambiti della scienza, e di “*farli parlare*” aiutandoci a comprenderne il significato.

Uno degli obiettivi di questo progetto è stato, quindi, quello di fornire ai ragazzi quelle basi di statistica necessarie ad orientarsi in ambito scientifico che, però, faticano a trovare spazio nella normale didattica in classe.

Un ricercatore, poi, presenta i risultati della sua ricerca scrivendo articoli che rispettino un preciso codice comunicativo e tenendo conferenze e seminari.

L'attività di ricerca del nostro progetto culmina, quindi, con la stesura di un articolo divulgativo e la presentazione pubblica dei risultati nel corso di alcune conferenze tematiche.

Questo consente ai ragazzi di trovare una modalità comunicativa efficace sia in forma scritta che orale.

In cosa consiste, in sostanza, il nostro progetto di A.S.L.?

Come abbiamo visto, gli studenti sono stati impegnati in un *project work* nel quale hanno simulato l'attività di un gruppo di ricerca: hanno analizzato dati provenienti da sensori presenti in Laboratorio di Fisica o da banche dati open italiane e straniere, in ambito sportivo e socio-sanitario. I dati raccolti sono stati elaborati utilizzando gli strumenti della statistica descrittiva ed hanno consentito loro di scrivere delle relazioni come questa, secondo un protocollo di divulgazione scientifica concordato assieme al nostro partner principale (l'Associazione Nemesis). In questa fase gli studenti hanno anche esercitato un “controllo incrociato” sulle relazioni dei vari gruppi, svolgendo attività di “*blind peer reviewer*”, tipica dei comitati di redazione delle riviste scientifiche. L'ultima fase del progetto ha previsto l'esposizione pubblica dei risultati conseguiti, mediante alcune serate di divulgazione aperte alla cittadinanza.

I nostri partner

Ringraziamo innanzitutto l'Associazione Nemesis, che ha tra i suoi obiettivi la diffusione della cultura scientifica nella società civile, con particolare riferimento al mondo della scuola, e che ospita nel suo sito una sezione che raccoglie i lavori preparati dagli studenti.

Non meno importante il supporto fornito dal laboratorio Big & Open Data Innovation (<http://bodai.unibs.it/>) dell'Università di Brescia. Nato grazie al supporto di Fondazione Cariplo con l'obiettivo di creare gruppi di lavoro che studiano e sviluppano – attraverso specifici progetti di ricerca – nuovi metodi, tecniche e strumenti per raccolta, gestione e analisi di dati multidisciplinari, merita la nostra gratitudine per averci fornito, tramite il progetto BDsports, il *know how* necessario per sviluppare il progetto.

Un ringraziamento doveroso va infine al Comune di Thiene, che ci ha messo a disposizione gli spazi idonei a tenere le serate in cui i ragazzi hanno potuto mettersi alla prova come divulgatori.

CALCIO E STATISTICA

INDICE DEI CONTENUTI

Introduzione	5
1. Reti segnate e subite durante una stagione	6
2. Analisi di serie storiche	8
2.1 Serie A Tim	8
2.2 Premier League	9
2.3 Liga spagnola	10
2.4 Bundesliga	11
2.5 Ligue 1	12
3. Calcio e distribuzione di Poisson	13
4. Il test del Chi-quadro	19
5. Come si calcolano le quote nelle scommesse?	23
5.1 Differenza tra probabilità e quote proposte	23
5.2 Metodo della lavagna	24
5.3 Differenza tra quote reali e quote proposte	25
5.4 Come calcolare una vincita	25
5.5 Bonus	26

INTRODUZIONE

Chi ha detto che calcio e statistica non hanno niente in comune? Durante il nostro progetto abbiamo analizzato alcuni elementi relativi ai più popolari campionati di calcio europei (Bundesliga, Serie A, Ligue 1, Liga, Premier Ligue). Abbiamo scelto questo argomento in quanto riprende un nostro interesse comune. Ciò che abbiamo fatto è stato ripercorrere i risultati ottenuti da alcune squadre nell'ultima stagione (2017/2018), notando in che modo i goal segnati e subito incidono nella classifica finale; il tutto mette in risalto quelle che sono le diverse modalità di gioco nei vari stati. Inoltre prendendo in considerazione gli esiti di ogni competizione negli ultimi vent'anni, abbiamo notato in che modo anche il tempo influisce nell'affermazione dei vari stili di gioco. Abbiamo notato anche che la distribuzione delle frequenze dei goal segnati complessivamente in ciascuna partita somiglia alla distribuzione di Poisson ed abbiamo analizzato la situazione mediante il test del chi-quadro, confermando – almeno in parte – la nostra intuizione.

Di questo tipo sono gli studi svolti dai bookmakers per calcolare le quote delle scommesse, di cui abbiamo parlato nell'ultimo capitolo del nostro lavoro.

Speriamo che il lavoro risulti interessante e di aver fatto parlare i dati.

1. Reti segnate e subite durante una stagione

Abbiamo deciso di prendere in esame i cinque campionati di calcio più popolari in Europa nella stagione 2017/2018 (serie A TIM, Bundesliga, Ligue1, Premier League e Liga), e vedere se dall'analisi dei goal segnati e subiti dalle squadre partecipanti nelle varie giornate fosse possibile notare qualcosa di interessante.

Dal momento che risulterebbe eccessivamente laborioso prendere in considerazione tutte le squadre partecipanti ai vari campionati, abbiamo deciso di considerare la prima classificata di ogni campionato, la prima delle squadre retrocesse ed una squadra di metà classifica scelta a piacere. Per esemplificare il lavoro svolto, riportiamo i dati relativi al campionato francese di Ligue 1.

Abbiamo preso in considerazione il Paris-Saint-Germain (campione di Francia 2018), il Metz (prima squadra retrocessa) ed il Montpellier (classificatasi al 10° posto), ricavando i dati presentati in Tab. 1.

		GIORNATA	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
P.S.G.	ANDATA	SEGNATI	2	3	6	3	5	2	0	6	2	2	3	5	4	2	2	1	3	4	3
		SUBITI	0	0	2	0	1	0	0	2	1	2	0	0	1	1	0	2	1	1	1
	RITORNO	SEGNATI	1	6	1	4	3	1	5	3	2	5	2	2	1	7	1	2	2	0	0
		SUBITI	0	0	2	0	0	0	2	0	0	0	1	1	1	1	0	2	2	2	0
MONTPELLIER	ANDATA	SEGNATI	1	0	1	1	0	1	0	1	2	1	0	1	0	3	0	1	1	1	2
		SUBITI	0	1	1	2	1	0	0	1	0	0	1	1	0	0	0	1	1	3	0
	RITORNO	SEGNATI	0	1	2	0	2	1	1	0	1	1	2	3	0	1	0	0	2	1	1
		SUBITI	0	1	1	4	1	0	1	0	1	1	2	1	0	3	1	1	0	1	1
METZ	ANDATA	SEGNATI	1	0	0	0	1	1	0	0	1	1	0	0	0	0	0	1	1	3	3
		SUBITI	3	2	1	1	5	0	1	1	3	2	2	3	0	2	3	3	1	1	0
	RITORNO	SEGNATI	1	3	1	2	3	0	0	2	1	0	1	2	0	2	1	1	1	0	0
		SUBITI	1	0	3	1	6	1	1	2	1	5	1	2	5	1	1	3	2	2	4

Tab. 1 – Risultati delle squadre selezionate, partita per partita

Per ognuna delle tre squadre abbiamo analizzato il numero di goal segnati e subiti nelle diciannove giornate di andata e nelle altrettante giornate di ritorno. In base ai dati riportati in Tab. 1 abbiamo calcolato la frequenza relativa dei goal segnati per partita dalle tre squadre, compilando la Tab. 2 ed abbiamo costruito gli istogrammi presentati in Fig. 3.

P.S.G.	GOAL	0	1	2	3	4	5	6	7	8
	FREQUENZA	3	6	11	7	3	4	2	1	1
MONTPELLIER	GOAL	0	1	2	3					
	FREQUENZA	16	14	4	4					
METZ	GOAL	0	1	2	3					
	FREQUENZA	12	18	6	2					

Tab. 2 – Frequenza di goal segnati per partita

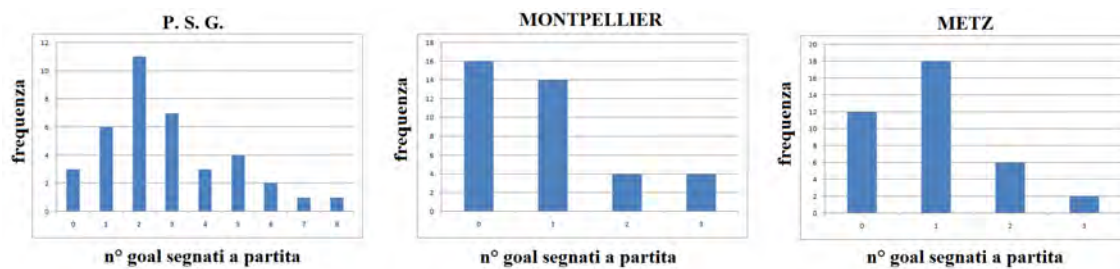


Fig. 3 – Frequenze di reti segnate a partita a confronto

In modo analogo abbiamo elaborato i dati relativi alle reti subite da ciascuna delle squadre considerate. La Tab. 4 riporta le frequenze dei goal subiti dalle tre squadre durante la stagione.

P.S.G.	GOL	0	1	2				
	FREQUENZA	18	11	9				
MONTPELLIER	GOL	0	1	2	3	4		
	FREQUENZA	14	19	2	2	1		
METZ	GOL	0	1	2	3	4	5	6
	FREQUENZA	4	14	8	7	1	3	1

Tab. 4 – Frequenza di goal subiti per partita

Confrontando l'andamento dei grafici in Fig. 5 con quelli in Fig. 3 possiamo notare che il P.S.G. ha non solo l'attacco più prolifico, ma anche la difesa più forte, visto che in 18 partite su 38 non ha subito alcuna rete.

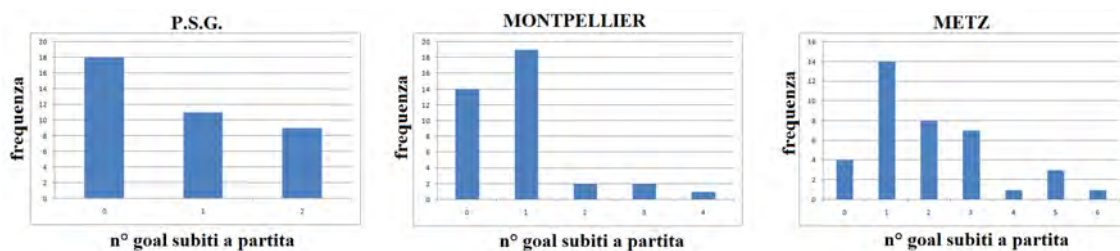


Fig. 5 – Frequenze di reti subite a partita a confronto

Analizzando la media dei goal segnati e subiti per partita dalle tre squadre considerate si ottengono ulteriori informazioni interessanti:

- 1) il P.S.G. ha segnato mediamente 2.84 goal a partita, subendone 0.76;
- 2) il Montpellier ha una media di realizzazione leggermente inferiore a quella del Metz (0.89 contro 0.94), ma ha una difesa sicuramente più efficace; la media dei goal incassati dal Montpellier è infatti 0.87 contro i 2 di media del Metz.

Questa differenza nelle performance delle squadre può consentirci di spiegare la diversa posizione in classifica; in particolare, con i numeri del Metz difficilmente ti salvi a fine stagione.

2. Analisi di serie storiche

In un secondo momento abbiamo esaminato la serie storica della media annua dei goal segnati nei campionati più importanti d'Europa, ottenendo la tabella 6.

STAGIONE	SERIE A	PREMIER LEAGUE	LIGA	BUNDESLIGA	LIGUE 1
1997/98	2,8	2,7	2,7	2,8	1,9
1998/99	2,8	2,5	2,6	2,8	1,9
1999/2000	2,9	2,8	2,6	2,9	2,0
2000/01	2,8	2,6	2,9	2,9	2,0
2001/02	2,6	2,6	2,5	2,9	1,8
2002/03	2,6	2,6	2,4	2,7	2,2
2003/04	2,7	2,7	2,7	3,0	2,3
2004/05	2,6	2,6	2,6	2,9	2,2
2005/06	2,5	2,5	2,6	2,8	2,1
2006/07	2,5	2,5	2,5	2,7	2,3
2007/08	2,6	2,6	2,6	2,7	2,3
2008/09	2,5	2,5	2,9	2,9	2,2
2009/10	2,8	2,8	2,7	2,8	2,4
2010/11	2,8	2,8	2,7	3,0	2,3
2011/12	2,8	2,8	2,8	2,9	2,5
2012/13	2,8	2,8	2,7	2,9	2,5
2013/14	2,8	2,8	2,8	3,2	2,4
2014/15	2,6	2,6	2,7	2,8	2,5
2015/16	2,7	2,7	2,7	2,8	2,3
2016/17	2,8	2,8	2,9	2,9	2,6
2017/18	2,7	2,7	2,7	2,8	2,7

Tab. 6 – Confronto tra le medie di goal segnati annualmente nei principali campionati europei

Analizziamo ora uno ad uno i singoli campionati.

2.1 Serie A Tim

Analizzando i dati riportati in Tab. 6 abbiamo costruito il grafico della Serie A (Fig. 7), da cui si può notare che negli ultimi 20 anni la media goal non è mai stata superiore a 3. Il picco è stato registrato nell'annata 1999/2000 con una media di 2,93 goal per partita. Nelle altre 19 stagioni esaminate, la media è sempre stata compresa tra 2,5 e 2,7.

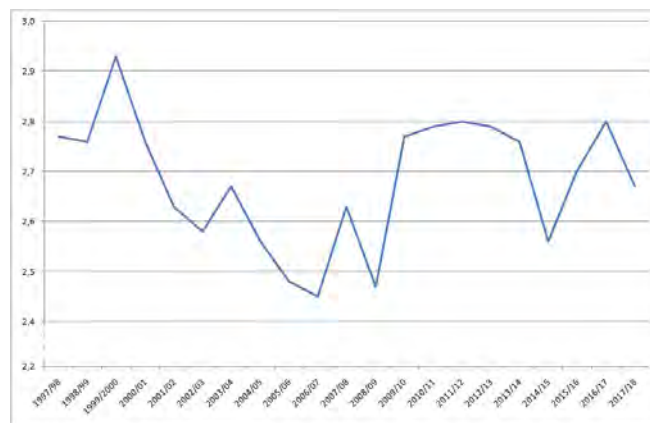


Fig. 7 – Andamento della media goal segnati annualmente in Serie A

Come possiamo spiegare il picco nella media?

La motivazione è principalmente matematica.

Nella stagione 1999/2000 le squadre iscritte al campionato italiano non erano nè 18, come accadeva spesso in passato, nè tantomeno 20, come accade ora. Le squadre iscritte erano 17. A causa di questo inusuale numero di partecipanti, il più basso degli ultimi 20 anni, le partite totali del campionato furono 256 a differenza delle 380 attuali o delle 306 che si disputavano prima della riforma del 2004/2005 che ha portato a 20 il numero di squadre iscritte.

Disputando un numero inferiore di partite, aumenta la probabilità di ottenere una media goal alta. Infatti meno partite si disputano, più la media goal può essere condizionata dalle partite in cui si segna molto (per esempio partite che terminano con 6/7 goal totali). Ma non è questa l'unica spiegazione. In quella stagione, infatti, le squadre considerate "medio-piccole" segnarono quasi quanto le squadre considerate "grandi".

La media goal infatti non cresce molto se tutte le squadre segnano poco e solo poche squadre riescono ad arrivare a fine campionato con un alto numero di goal realizzati. La media cresce se le squadre di media classifica realizzano un maggior numero di goal, mentre diminuisce se quest'ultime si avvicinano al numero di goal segnati dalla squadre vicine alla zona retrocessione.

Come già detto, la media goal annua nelle restanti 19 annate è rimasta costante. Questo è dovuto

principalmente al fatto che le squadre nel campionato italiano prediligono la fase difensiva a quella offensiva. Difatti l'Italia è conosciuta nel mondo come la patria dei grandi portieri (Zoff, Buffon...) e dei grandi difensori (Maldini, Baresi, Bonucci, Materazzi...). Questa tattica difensiva, pur regalando grandi gioie al calcio italiano (nel mondiale del 2006, l'Italia prese solamente 2 goal), comporta partite meno spettacolari e con un numero di goal non troppo alto, basti pensare che il record di goal di un singolo giocatore in un singolo campionato, siglato da Higuain con 36 goal, è un numero di reti che molti singoli giocatori riescono a fare ogni anno nei campionati stranieri (Messi, Cristiano Ronaldo, Cavani...).

Nel campionato italiano sarà quindi molto difficile vedere grandi picchi di media goal. Neppure nelle ultime stagioni, quando squadre come Roma, Napoli e Juventus hanno sfiorato i 100 goal in campionato, la media è stata alta, poichè molte squadre hanno concluso il campionato con pochi goal all'attivo (molte squadre hanno chiuso il campionato segnando addirittura meno di un goal di media a partita).

2.2 Premier League

Prima di esaminare la colonna relativa alla Premier League (Tab. 6) ci saremmo aspettati di incontrare medie goal abbondantemente sopra il 3, abituati come siamo alle pirotecniche partite d'oltremarina. Invece la media goal del campionato inglese, pur risultando quasi sempre maggiore alla media goal della Serie A, non ha mai avuto alti picchi (Fig. 8), tanto che negli ultimi 20 anni non ci si è mai nemmeno avvicinati al picco di 2,93 registrato in Serie A. Il motivo di questi risultati è probabilmente il fatto che negli ultimi 20 anni la Premier ha avuto sempre 20 squadre iscritte e mai 17/18 come si è verificato nel campionato italiano. Come abbiamo già osservato, aumentando il numero di partite, la media difficilmente è molto alta poichè diminuisce il "peso" relativo delle gare in cui si segna di più.

Dal 2010/2011 in poi, la media è quasi sempre in costante aumento. Negli ultimi anni le "big" inglesi hanno fatto grandi investimenti per migliorare la fase offensiva (basti pensare agli acquisti di Aguero e Salah o alla maturazione calcistica di grandi attaccanti come Kane e Hazard) e spesso hanno fatto grandi buchi nell'acqua negli acquisti in difesa (basti pensare a Zouma, comprato dal Chelsea per 35 milioni e quasi mai utilizzato). Le difese sempre meno forti, gli attacchi sempre più prolifici e gli arrivi di grandi allenatori (Guardiola, Conte, Sarri) hanno portato ad un innalzamento nella media goal per partita, che ha reso il campionato inglese il più spettacolare al mondo agli occhi dei molti appassionati di calcio.



Fig. 8 – Andamento della media goal segnati annualmente in Premier League

2.3 Liga spagnola

Osservando la terza colonna della Tabella 6, si può notare come solamente 2 volte negli ultimi 20 anni la media goal sia stata inferiore a 2,5. Particolare è il picco registrato, in negativo, nella stagione 2002/2003 quando la media goal è stata di circa 2,3 (Fig. 9).

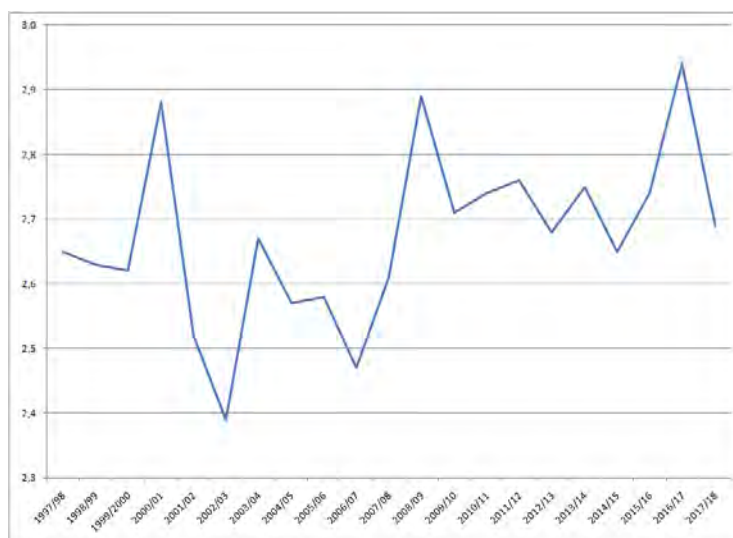


Fig. 9 – Andamento della media goal segnati annualmente in Liga spagnola

La motivazione è probabilmente che il Barcellona, uno dei grandi team spagnoli, chiuse la stagione addirittura al sesto posto, segnando molto poco. Anche molti altri team segnarono mol-

to poco, con molte squadre di bassa classifica che non raggiunsero nemmeno la media di un goal a partita.

Nelle altre stagioni la media goal è stata sempre piuttosto alta, sfiorando spesso una media annua di 3 goal a partita. Il calcio spagnolo non è famoso per le grandi difese, ma per i grandi attacchi. Le tattiche molto offensive hanno fatto sì che anche squadre come il Real Madrid possano subire 3 goal a partita. Inoltre l'innalzamento della media goal è dovuto alla grande crescita dell'Atletico Madrid, che negli ultimi anni è diventata la terza forza del campionato spagnolo, e alla presenza di molti top player in attacco (Suarez, Neymar, Bale, Isco, Griezmann, Benzema, Coutinho) e dei due migliori giocatori del mondo, Messi e Cristiano Ronaldo, capaci di segnare anche 40 goal a campionato.

Siamo propensi a ritenere che la media goal continuerà ad aumentare, se le big spagnole continueranno a crescere ed arriveranno buoni giocatori anche nelle squadre di media classifica.

2.4 Bundesliga

La Bundesliga negli ultimi 20 anni ha sempre avuto una media goal molto alta, arrivando sempre oltre i 2,5 goal a partita (Tab. 6).

Anche in questo caso è determinante il numero di partite in una stagione. Essendo 18 le squadre iscritte, il numero totale di partite in un campionato è 306. Quindi in Italia, Francia, Inghilterra e Spagna si giocano 72 partite in più e la media cala inevitabilmente.

Il picco in Germania si è registrato nel 2013/2014 con una media di 3,1 goal a partita (Fig. 10).



Fig. 10 – Andamento della media goal segnati annualmente in Bundesliga

Tale picco non è dovuto solo al fatto che molte squadre di alta classifica abbiano segnato molto (Bayern 94 goal, Borussia Dortmund 80) ma anche al fatto che solamente due squadre hanno segnato meno di 40 goal (Nurnberg e Braunschweig, classificate rispettivamente all'ultimo ed al penultimo posto). Possiamo ritenere che la media goal quell'anno si sia alzata anche grazie all'arrivo di Guardiola al Bayern Monaco. L'allenatore spagnolo ha rivoluzionato il calcio mondiale con i suoi schemi innovativi, che sono stati copiati e studiati da tutti gli allenatori delle squadre tedesche. Tutto questo ha portato ad un miglioramento globale del calcio tedesco

al punto che la vittoria della Germania nei mondiali del 2014 in Brasile non sembra affatto un evento fortuito.

Lo stile di gioco tedesco, che predilige l'attacco alla difesa, ha fatto il resto, permettendo che la media goal annuale della Bundesliga rimanesse molto alta e rendendo la Bundesliga uno dei campionati più seguiti al mondo.

Siamo portati a credere che in futuro la media goal in Germania rimarrà sempre piuttosto alta, anche grazie alla crescita di squadre come Borussia Dortmund, Lipsia e Schalke che sono diventate temibili anche nelle competizioni europee, aumentando ancora di più il fascino del calcio tedesco.

2.5 Ligue 1

L'andamento della media goal della Ligue 1 è quello che ha portato i risultati più bassi durante le nostre ricerche (Tab. 6). Addirittura per ben due stagioni consecutive la media è stata inferiore ai due goal a partita e fino al 2010/2011 non è mai andata oltre ai 2,5 goal.

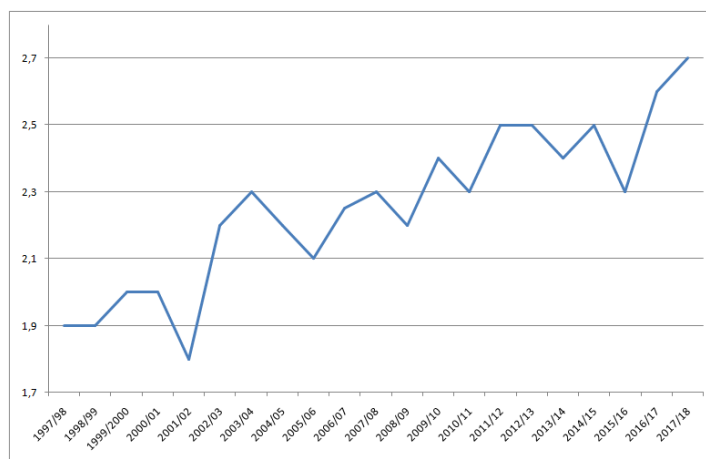


Fig. 11 – Andamento della media goal segnati annualmente in Ligue 1

Il motivo del cambiamento di rotta registrato negli ultimi anni è senza dubbio la crescita esponenziale del PSG che, grazie ai soldi del nuovo proprietario proveniente dal Qatar, ha potuto comprare grandissimi campioni come Pastore, Di Maria, Verratti, Ibrahimovic, Cavani, Neymar ed è diventato potenzialmente uno dei top club europei pur senza riuscire a conseguire grandi risultati internazionali. Importante per la media goal è stata la crescita del Monaco, ricca compagine monegasca che ha avuto in squadra giovani talentuosi come Mbappe e Lemar ma ha potuto fare anche grandi investimenti, come Falcao. Inoltre stanno crescendo Lione e Marsiglia, con quest'ultima che ha sfiorato la vittoria dell'Europa League quest'anno.

Prima di questa crescita francese, la media goal era molto bassa principalmente per motivi calcistici. La maggior parte dei campioni nati in Francia non hanno quasi mai giocato nella Ligue 1, preferendo giocare in altri grandi club europei (la Juventus aveva in rosa Zidane, Trezeguet ed Henry), e quindi il livello del calcio francese non è mai stato molto alto.

Non siamo particolarmente fiduciosi di veder crescere molto la media goal in futuro, in quanto solamente il PSG ha una rosa con attaccanti di prim'ordine, mentre il livello delle restanti squadre rimane molto basso.

3. Calcio e distribuzione di Poisson

Dopo aver analizzato le serie storiche abbiamo considerato un altro aspetto interessante: il numero di goal segnati complessivamente per partita.

In un campionato di calcio ci sono x partite che terminano con 0 goal (l'unico risultato possibile è lo 0-0), y partite che terminano con 1 goal (risultati possibili: 1-0 e 0-1), z partite che terminano con 2 goal (risultati possibili: 2-0, 1-1 e 0-2) e così via fino al numero massimo di goal a partita che c'è stato in quel campionato.

Definiamo gli eventi:

- $E(0)$: "0 goal segnati nella partita"
- $E(1)$: "1 solo goal segnato nella partita"
- $E(2)$: "2 goal segnati nella partita"
- ...
- $E(n)$: " n goal segnati nella partita"

Per studiare tali eventi è utile analizzarne le frequenze relative, ovvero vedere quante volte si è verificato $E(0)$, quante volte si è verificato $E(1)$, ..., nel campionato che si sta studiando.

L'istogramma riportato in figura 12 rappresenta le frequenze ottenute dagli eventi considerati.

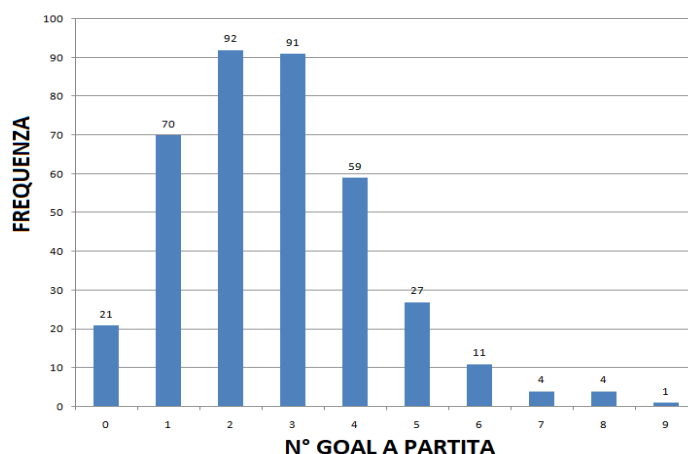


Fig. 12 – Frequenza assoluta del numero di reti segnate complessivamente per partita (dati Ligue 1 – 2017/18)

Il nostro obiettivo è quello di cercare una distribuzione teorica alla quale si possa fare riferimento per i dati reali, ottenuti dallo studio di ciascun campionato. In altri termini, lo scopo di questo studio è verificare se i dati del campionato A approssimano in modo più o meno preciso l'andamento di una distribuzione teorica di riferimento.

È qui che entra in scena la distribuzione di Poisson.

Questa distribuzione consente di studiare i fenomeni legati ad eventi rari, come ad esempio il numero di errori di stampa in una pagina di un libro o il numero di bambini nati con sindrome di Down in un ospedale nell'arco di un mese.

Questa distribuzione ha delle caratteristiche ben precise:

- il verificarsi di un evento non ha alcuna influenza sulla probabilità che quell'evento si verifichi una seconda volta (se nella 1° giornata di un campionato la partita tra le squadre A e B termina con 2 goal, nella successiva giornata la partita disputata dalla squadra A potrà

terminare con 4 goal così come con 0 goal; dunque il risultato della partita della 1° giornata non influenza il risultato della partita della 2° giornata), dunque gli eventi sono indipendenti tra loro;

- la probabilità che un evento si verifichi in un certo intervallo aumenta con il crescere dell'ampiezza di tale intervallo (nella pagina di un libro ci sono 680 parole e la probabilità di trovare un errore è dello 0,6%, se nella stessa pagina ci fossero invece 1045 parole la probabilità di trovare un errore diventerebbe dello 0,92%);
- la probabilità che un evento si verifichi più di una volta nello stesso intervallo è trascurabile, perché non è necessaria per la nostra analisi (se dovessimo studiare il numero di chiamate ricevute in un giorno da un callcenter non ci interesserebbe la probabilità che arrivino 4 chiamate in 20 minuti).

Prendiamo in analisi la variabile aleatoria X che indica il numero di volte che si è verificato un evento raro all'interno di un intervallo. Vale il seguente **teorema**:

La probabilità che la variabile aleatoria X assuma un preciso valore x è data dall'espressione:

$$P(X = x) = \frac{e^{-m} \cdot m^x}{x!} \quad (1)$$

dove il parametro m indica il numero medio di volte in cui l'evento si è realizzato nell'intervallo scelto.

Esempio:

Da una rilevazione risulta che il numero di incidenti sul lavoro avvenuti in un'azienda, in un mese, è una variabile casuale distribuita secondo Poisson con un valore medio di 1,5. Per calcolare la probabilità che in un mese non ci siano incidenti basta applicare la (1) con $x = 0$ e $m = 1,5$:

$$P(X = 0) = \frac{e^{-1,5} \cdot 1,5^0}{0!} \approx 0,22 = 22\%$$

La distribuzione di Poisson può essere ben rappresentata da un istogramma; la forma del grafico che si ottiene acquisisce maggior simmetria rispetto ad un valore centrale all'aumentare della media m (vd. Fig. 13).

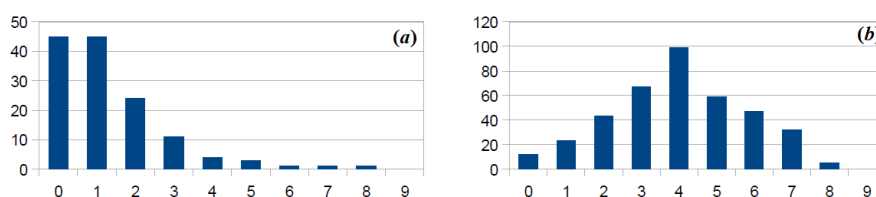


Fig. 13 – Gli istogrammi rappresentano la distribuzione di Poisson con parametro $m = 1,07$ (a) e $m = 4$ (b)

La distribuzione di Poisson è una distribuzione discreta ed assume valori nei numeri naturali. Quando il valore del parametro m non è troppo basso (almeno 10) la distribuzione può essere approssimata ad una curva gaussiana; l'unica differenza fondamentale è che la variabile di Poisson è una distribuzione discreta mentre la gaussiana è continua.

Cosa c'entra la distribuzione di Poisson con il gioco del calcio?

Analizzando gli istogrammi delle frequenze del numero di goal segnati per partita, abbiamo osservato che la forma (Fig. 12) assomiglia molto a quella tipica della distribuzione di Poisson.

Il nostro obiettivo è verificare se questa distribuzione ci può fornire il modello teorico a cui fare riferimento per l'analisi del fenomeno. Per farlo abbiamo analizzato i cinque maggiori campionati europei di calcio:

- Bundesliga (Germania);
- Premier League (Inghilterra);
- Ligue 1 (Francia);
- Liga (Spagna);
- Serie A Tim (Italia).

Riportiamo, a titolo di esempio, il lavoro svolto su 2 campionati: la Ligue 1 e la Serie A Tim, relativamente alla stagione 2017/18.

Dopo aver analizzato tutte le partite dei due campionati esaminati, abbiamo calcolato la frequenza con cui le partite sono terminate con n goal complessivi, compilando la tabella 14.

N° DI GOAL PER PARTITA	FREQUENZA ASSOLUTA	
	LIGUE 1	SERIE A TIM
0	21	28
1	70	74
2	92	80
3	91	96
4	59	43
5	27	40
6	11	12
7	4	5
8	4	2
9	1	===

Tab. 14 – Frequenze di reti segnate complessivamente per partita in Ligue 1 e Serie A Tim

Come si può facilmente notare, il numero complessivo di partite che terminano sullo 0-0 sono davvero poche nell'intero arco della stagione. Allo stesso modo, sono rari i casi in cui la partita termina con un elevato numero di goal segnati. La maggior parte delle gare termina con 2 o 3 reti complessive.

Abbiamo, poi, costruito gli istogrammi delle frequenze (Fig. 15), notando la forma “a campana”. Non possiamo fare riferimento alla distribuzione gaussiana poiché quella che stiamo esaminando è una distribuzione discreta.

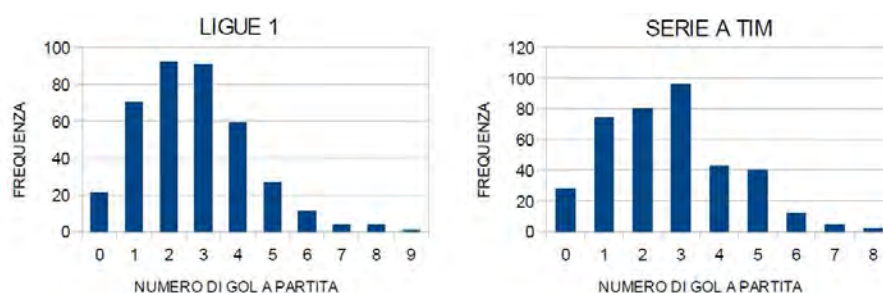


Fig. 15 – Distribuzione delle frequenze delle gare che finiscono con n reti complessive

Come già affermato, il nostro obiettivo è ora quello di trovare (se c'è) una distribuzione teorica di riferimento che possa venir presa come modello per rappresentare il numero di reti complessivamente segnate in ciascuna gara.

Osservando con attenzione l'andamento degli istogrammi in fig. 15 si può avvertire una certa somiglianza con l'andamento della distribuzione di Poisson, soprattutto nel caso della Ligue 1.

Per verificare se la distribuzione di Poisson possa essere presa come modello per descrivere il fenomeno considerato, bisogna innanzitutto calcolare la probabilità che una partita termini con x goal (nell'intero arco del campionato) e confrontarla con la probabilità teorica data dalla formula, spiegata nella prima parte del capitolo.

Come potremmo spiegare meglio quanto appena detto?

Proviamo a riflettere su questo esempio (relativo al campionato di Ligue 1): immaginiamo di porre 380 biglie in un grande scatolone e di volerne estrarre una. Su ogni biglia è rappresentata una cifra da 0 a 9. Le 380 biglie corrispondono alle 380 partite svolte nel campionato preso in esame e la cifra riportata sulla superficie della biglia corrisponde al numero di goal che ci sono stati in quella partita. Più precisamente, nello scatolone sono state inserite 21 palline col numero 0, 70 col numero 1, 92 col numero 2, ecc. Estraiamo casualmente una biglia dallo scatolone; per determinare la probabilità che esca – ad esempio – una biglia con la cifra 3 dobbiamo calcolare il rapporto:

$$P(X = 3) = \frac{\text{casi favorevoli}}{\text{casi possibili}} = \frac{\text{n° biglie con la cifra 3}}{\text{n° totale biglie}} \quad (2)$$

Il risultato ottenuto mediante la (2), cioè la frequenza relativa del verificarsi dell'evento considerato, dovrà essere confrontato con la probabilità calcolata mediante la (1).

Per farlo, però, ci serve conoscere la media dei goal segnati nell'arco di un'intera stagione in Francia ed in Italia (che corrisponde al parametro m della distribuzione di Poisson).

Visto che il numero di partite da esaminare è elevato, conviene utilizzare la formula della media ponderata:

$$m = \frac{\sum_{i=0}^k i \cdot f_i}{\sum_{i=0}^k f_i} = \frac{0 \cdot f_0 + 1 \cdot f_1 + 2 \cdot f_2 + \dots}{f_0 + f_1 + f_2 + \dots} \quad (3)$$

ove i rappresenta il numero di reti segnate complessivamente in ciascuna partita ed f_i la frequenza assoluta di quel dato.

○ LIGUE 1: $m_1 = \frac{\text{n° totale di reti segnate}}{\text{n° totale di partite giocate}} = \frac{1033}{380} = 2.71842105$

○ SERIE A TIM: $m_2 = \frac{\text{n° totale di reti segnate}}{\text{n° totale di partite giocate}} = \frac{1017}{380} = 2.67631579$

La tabella 16 riporta i dati relativi ai due campionati. I valori empirici della probabilità di ciascun evento sono calcolati come rapporto tra il numero di gare in cui sono stati segnati n goal

ed il numero totale di partite giocate nel campionato in esame, partendo dai dati riportati in Tab. 14; abbiamo ottenuti, invece, i valori teorici applicando la (1).

N° DI GOAL PER PARTITA	LIGUE 1		SERIE A TIM	
	PROBABILITÀ		PROBABILITÀ	
	VALORI EMPIRICI	POISSON	VALORI EMPIRICI	POISSON
0	0.05526	0.06598	0.07368	0.06882
1	0.18421	0.17936	0.19474	0.18417
2	0.24211	0.24379	0.21053	0.24645
3	0.23947	0.22090	0.25263	0.21986
4	0.15526	0.15013	0.11316	0.14711
5	0.07105	0.08162	0.10526	0.07874
6	0.02895	0.03698	0.03158	0.03512
7	0.01053	0.01436	0.01316	0.01343
8	0.01053	0.00488	0.00526	0.00449
9	0.00263	0.00147	===	===

Tab. 16 – Distribuzioni (empirica e teorica) a confronto per il campionato di Ligue 1 e di Serie A Tim

Osservando attentamente il contenuto della tabella 16 si può notare che alcuni valori nelle colonne delle “probabilità” si assomigliano molto.

Per quanto riguarda la Ligue 1 la frequenza relativa delle partite in cui si sono segnati complessivamente 2 goal è del 24,21% ed il valore della probabilità determinata mediante il modello di Poisson è del 24,38%; la differenza tra i due valori è molto esigua (0,17%).

Anche per quanto riguarda la Serie A Tim ci sono alcuni valori molto simili. La frequenza relativa delle partite in cui si sono segnati complessivamente 7 goal è dell' 1,32% ed il valore della probabilità determinata mediante il modello di Poisson è 1,34%.

Nella Ligue 1 lo scarto massimo tra la frequenza relativa e la probabilità data dal modello di Poisson è dell' 1,86% nel caso di partite in cui siano stati segnati complessivamente 3 goal.

A prima vista, quindi, il modello di Poisson sembra essere un buon modello di riferimento per lo studio del fenomeno che abbiamo preso in considerazione.

Per la Serie A Tim lo scarto massimo tra la frequenza relativa e la probabilità determinata mediante il modello di Poisson è del 3,60%, nel caso di partite in cui siano stati segnati complessivamente 2 goal. In questo caso il modello sembra comunque essere un modello abbastanza buono anche se non quanto quello offerto per la Ligue 1.

Fatte queste considerazioni generali, per avere una visione d'insieme migliore della situazione che stiamo studiando e per provare a trarre delle conclusioni abbiamo deciso di realizzare dei grafici che rappresentino la distribuzione teorica e quella empirica (Fig. 17).

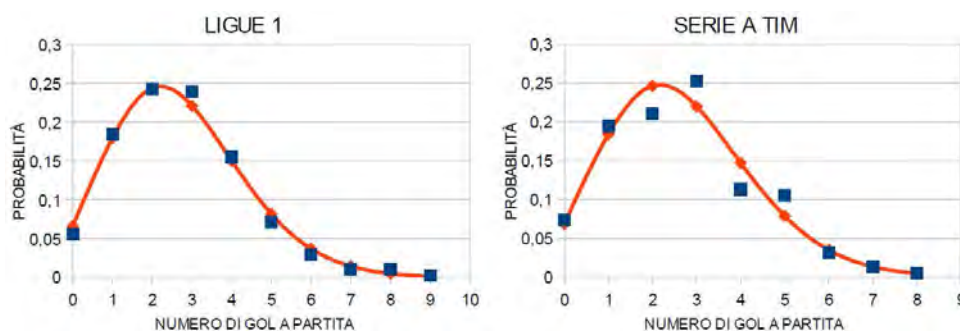


Fig. 17 – Distribuzioni empiriche e teoriche a confronto

In figura 17 abbiamo rappresentato dei grafici a dispersione in cui abbiamo riportato in rosso l'andamento delle distribuzioni di Poisson, mentre i puntini blu rappresentano le probabilità ottenute dai dati delle partite dei campionati francese ed italiano del 2017/18.

Per la Ligue 1 la poissoniana sembra essere un'ottima curva di riferimento per il fenomeno che stiamo studiando, ovvero il numero di goal in una partita. Infatti, a parte un caso in cui lo scostamento non è molto marcato, notiamo che i puntini blu sono esattamente sopra la linea rossa. Siamo portati a ritenere che la distribuzione di Poisson sia un buon modello di riferimento almeno per quanto riguarda la Ligue 1.

Guardando il grafico della Serie A Tim possiamo osservare, invece, che sono parecchi i puntini blu a non essere sovrapposti alla linea rossa; in questo caso, quindi, la distribuzione di Poisson non sembra essere modello idoneo a rappresentare il fenomeno.

Come fare a decidere se un modello sia affidabile o meno?

Per capire se il modello sia valido o meno esiste un test, chiamato test del chi-quadro, che consente di ottenere matematicamente il grado di affidabilità del modello esaminato; nel prossimo capitolo lo utilizzeremo per provare a trarre le conclusioni del nostro studio.

4. Il test del Chi-quadro

Nel capitolo precedente abbiamo osservato una certa somiglianza tra la distribuzione empirica del numero di reti segnate in ciascuna partita del campionato (italiano e francese) 2017/18 con la distribuzione di Poisson. Una volta ipotizzata questa relazione è utile vedere anche il grado di certezza con cui possiamo considerare affidabili i dati ottenuti e quindi anche quanto il modello da noi adottato (Poisson) possa ritenersi valido. Per dirla più semplicemente, si cerca di determinare quanto i fenomeni reali assomigliano a quelli teorici.

Il metodo più efficace per fare questa prova è il cosiddetto test del chi-quadro, basato sul confronto tra le frequenze empiriche registrate e quelle teoriche.

Prima di introdurlo, proviamo a chiarire il concetto di frequenza teorica e quello di frequenza empirica. La prima corrisponde al numero di volte che ci aspettiamo che un evento si verifichi considerando la sua probabilità calcolata a priori.

Facciamo un esempio: lanciando una moneta regolare, la probabilità che esca testa è uguale a quella che esca croce, e corrisponde a $\frac{1}{2}$ =50%. Consideriamo ora la probabilità dello stesso evento dal punto di vista empirico. Lanciando 20 volte la moneta, ci aspettiamo che per 10 volte (50% dei lanci) esca testa, e le altre 10 esca croce. Sono proprio questi valori a corrispondere alla frequenza (assoluta) teorica. Tuttavia questa condizione è un'evenienza ideale, è difficile che tutto ciò si verifichi davvero, con un numero di lanci così basso.

Potrebbe accadere, infatti, che nella serie di 20 lanci, 15 volte esca testa e 5 croce. La frequenza empirica si calcola una volta che la serie di lanci è conclusa, e corrisponde al numero di volte in cui si verifica veramente un evento (testa o croce). In questo caso, testa è uscita nel 75% dei casi e croce nei restanti 25%.

È possibile fare una considerazione simile anche per il caso dei campionati che abbiamo analizzato, considerando la frequenza con cui una partita si conclude con $0,1,2,\dots,n$ goal. Riportiamo di seguito (Tab. 18) i dati del campionato di Ligue 1 francese, per la stagione 2017/18.

N° GOAL	0	1	2	3	4	5	6	7	8	9
FREQUENZE ASSOLUTE	21	70	92	91	59	27	11	4	4	1

Tab. 18 – Frequenze di reti segnate complessivamente per partita in Ligue 1

In questo caso, però, il calcolo della probabilità teorica che in una partita si segni un certo numero di goal risulta più complesso.

Dal momento che l'evento "si è segnato un goal" è piuttosto raro in una partita di calcio, abbiamo ritenuto ragionevole utilizzare la distribuzione Poisson, di cui abbiamo esaminato le caratteristiche generali nel cap. 3.

Tramite la distribuzione di Poisson è possibile estrapolare la probabilità teorica che in una partita si segnino n goal. Ricordando che, per la definizione classica di probabilità, risulta:

$$P(E) = \frac{\text{numero dei casi favorevoli}}{\text{numero dei casi possibili}} \quad (4)$$

per ottenere la frequenza teorica (numero ideale di casi favorevoli), basterà poi invertire la (4) ottenendo così il numero dei casi favorevoli come prodotto tra la probabilità teorica e il numero totale dei casi possibili, che coincide con il numero di partite giocate nell'intero campionato (380).

N° GOAL	PROBABILITÀ (POISSON)	FREQUENZA TEORICA
0	0.06597885	25.07196275
1	0.17935829	68.15615138
2	0.24378568	92.63855839
3	0.22090404	83.94353581
4	0.15012755	57.04846874
5	0.08162198	31.01635169
6	0.03698048	14.0525839
7	0.01436122	5.457262846
8	0.00487998	1.854392276
9	0.00147398	0.560113223

Tab. 19 – Frequenze teoriche di reti segnate complessivamente per partita (Ligue 1)

La tabella 19 riporta le frequenze teoriche calcolate (per il campionato di Ligue 1) sulla base della distribuzione di Poisson; per poter applicare il test del Chi-quadro è necessario che le frequenze teoriche non siano inferiori a 5 (vd. le celle evidenziate in Tab. 19).

È necessario, quindi, raggruppare più modalità e sommarle come in tabella 20.

N° GOAL	PROBABILITÀ (POISSON)	FREQUENZA TEORICA
0	0.06597885	25.07196275
1	0.17935829	68.15615138
2	0.24378568	92.63855839
3	0.22090404	83.94353581
4	0.15012755	57.04846874
5	0.08162198	31.01635169
6	0.03698048	14.0525839
7	0.01436122	7.871768345

Tab. 20 – Frequenze teoriche "corrette" per l'applicazione del test Chi-quadro

Per applicare il test del Chi-quadro dobbiamo confrontare tra loro le frequenze empiriche, calcolate in base ai dati del campionato (raccolti in Tab. 18), e quelle teoriche (Tab. 20).

Per calcolare il Chi-quadro applichiamo la formula

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - n \cdot p_i)^2}{n \cdot p_i} \quad (5)$$

dove f_i corrisponde alle frequenze relative sperimentali, $n \cdot p_i$, al prodotto tra le probabilità teoriche e il numero di partite, quindi la frequenza teorica, e k al numero di distribuzioni possibili, quindi i probabili goal segnati nell'arco della sfida (0,1,2,...,n).

Applicando la formula si ottengono i termini della somma (Tab. 21) da cui si ricaverà il chi-quadro.

N° GOAL	TERMINI DELLA SOMMA
0	0,661331576
1	0,049882185
2	0,007888146
3	0,593180718
4	0,066758571
5	0,520083118
6	0,663100006
7	0,161705301

Tab. 21 – Termini della somma (5)

Per calcolare il Chi-quadro basta ora sommare i termini della seconda colonna della tabella 21; così facendo otteniamo $\chi^2 \approx 2.72$.

Rimane ora da capire a quale percentuale di affidabilità corrisponde questo valore.

Per farlo è necessario il calcolo dei gradi di libertà ν della distribuzione di chi-quadro, da mettere in relazione con il risultato ottenuto. Ma come calcolarlo?

Nel caso in cui nella distribuzione delle frequenze teoriche ci si sia serviti solamente di dati non campionari (teorici), i gradi di libertà corrispondono a:

$$\nu = k - 1 \quad (6)$$

dove k corrisponde al numero di elementi della distribuzione. Nel nostro caso k risulta essere il numero di possibili esiti di ogni partita (in termini di goal complessivamente segnati). Considerando che nella Ligue 1 francese nella stagione 2017/2018 si è segnato in ogni partita un numero di goal compreso tra 0 e 9, il nostro k risulta essere il numero di elementi compresi tra 0 e 9 inclusi, quindi 10.

Nel caso in cui invece nella distribuzione si siano utilizzati dati campionari, i gradi di libertà si calcolano in modo diverso:

$$\nu = k - 1 - m \quad (6)$$

dove m corrisponde al numero di dati campionari di cui ci si è serviti.

Durante il nostro studio abbiamo utilizzato un unico dato sperimentale: la media dei goal a partita, calcolata sulla base di tutti gli esiti delle partite della stagione 2017/18. È stato infatti grazie a questo valore che è stato possibile ricavare la distribuzione di Poisson.

Per noi, quindi, $m = 1$. Risulta facile ora il calcolo dei gradi di libertà. Applicando la (6) otteniamo, infatti:

$$\nu = 10 - 1 - 1 = 8$$

Una volta determinato il numero di gradi di libertà, utilizzando la tavola del Chi-quadro abbiamo determinato il grado di affidabilità della nostra ipotesi.

Considerando 8 gradi di libertà, il nostro chi-quadro risulta compreso tra 1,65 e 2,73, ciò significa che la percentuale di significatività dei valori sperimentali è compresa tra il 95% e il 99%.

I dati empirici ricavati quindi sono molto affini a quelli ideali e possiamo perciò affermare che il modello teorico adottato (Poisson) è valido per il campionato di Ligue 1.

Il valore del Chi-quadro per la distribuzione relativa al campionato italiano è circa 10.72 ed il numero di gradi di libertà è 7; in base a questi valori la probabilità di essere nella regione di non accettazione del test è compresa tra il 10% ed il 25%.

Come si vede il risultato ottenuto è peggiore, ma non da scartare necessariamente l'ipotesi di aderire al modello poissoniano.

5. Come si calcolano le quote nelle scommesse?

Per sviluppare le nostre ricerche abbiamo utilizzato metodologie simili a quelle impiegate dai bookmakers per determinare le quote delle scommesse sugli eventi sportivi. Di cosa si tratta? Proviamo a vederlo assieme.

5.1 Differenza tra probabilità e quote proposte

Per calcolare la probabilità di un evento in una partita, i bookmakers prendono in considerazione le ultime partite delle squadre sulle quali si effettuano le scommesse.

Analizziamo a titolo di esempio la partita Milan-Sampdoria, partita del campionato italiano di Serie A disputatasi nella stagione 2008/2009.

BOOKMAKERS	1	X	2
A	1.60	3.40	5.60
B	1.61	3.76	6.89
C	1.57	3.70	6.25

Tab. 22 – Alcune quote proposte dai bookmakers per la partita Milan-Sampdoria del 19/10/2008

Per calcolare le probabilità degli eventi presenti nella Tabella 1 (1, X, 2) si sono esaminate le ultime 6 partite giocate in casa dal Milan prima del match preso in considerazione (4 vittorie, 1 pareggio, 1 sconfitta) e le ultime 7 in trasferta della Sampdoria (1 vittoria, 2 pareggi, 4 sconfitte). La probabilità dell'evento "1" (vittoria della squadra di casa) deriva dal rapporto tra la somma delle vittorie del Milan in casa nelle partite considerate e delle sconfitte dei blucerchianti nelle partite in trasferta e le partite totali di Sampdoria e Milan prese in considerazione:

$$P(1) = \frac{4+4}{6+7} = 0.61 = 61\%$$

Analogamente si calcola l'evento "X" (pareggio) e "2" (vittoria squadra in trasferta).

$$P(X) = \frac{1+1}{6+7} = 0.15 = 15\%$$

$$P(2) = \frac{1+1}{6+7} = 0.15 = 15\%$$

La quota derivante dalla probabilità (utilizzata come base di partenza) deriva dal rapporto tra 100 e la probabilità dell'evento.

La quota per l'"1" è quindi $\frac{100}{61\%} = 1.63$, quella del "2" è $\frac{100}{15\%} = 6.66$, come pure quella dell'"X".

Tali quote differiscono di molto dalle quote proposte. La variazione deriva da diversi fattori:

- Stato di forma delle due formazioni: si guarda l'andamento in classifica delle due squadre fino alla partita a cui si giunge (di seguito nella Fig. 23 è riportato lo stato di forma di Bologna e Sassuolo, in un grafico preso dal sito Snai);
- Assenze di alcuni giocatori (per infortuni o squalifiche);
- Importanza della partita (motivazione della squadra, rivalità);

- Quotisti: esperti di mercato che lavorano nelle agenzie di scommesse. Il loro obiettivo è quello di garantire sempre un guadagno da parte dell'agenzia, anche nel caso in cui lo scommettitore vinca. Il guadagno che i quotisti cercano di ottenere è pari all'8-10%. Per raggiungere tale obiettivo viene utilizzato il “metodo della lavagna”.

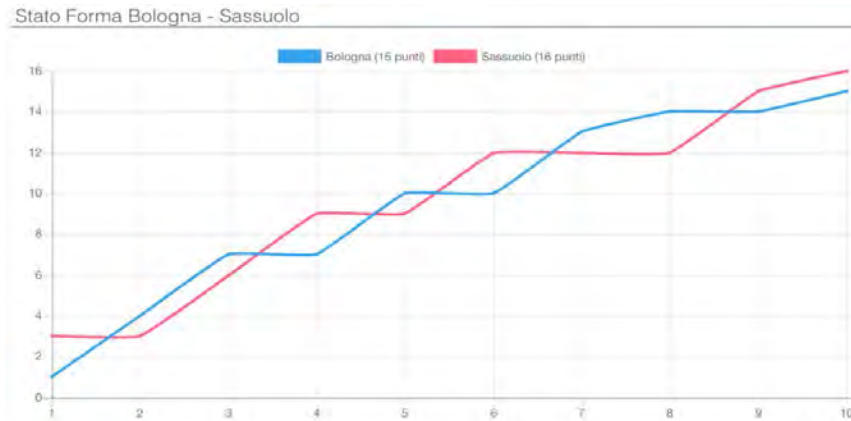


Fig. 23 – Stato di forma di Bologna e Sassuolo (fonte Snai)

5.2 Metodo della lavagna

Il metodo della lavagna è stato introdotto dai quotisti della SNAI per ottenere un guadagno, indipendentemente dalla vincita o meno dello scommettitore.

Prendendo in considerazione la partita precedente (Milan-Sampdoria del 2008/2009), i quotisti sono riusciti a guadagnare almeno l'8,9% da ogni scommessa.

Per calcolare tale cifra si parte dai reciproci della quota proposta dalle varie agenzie di scommesse; ad esempio:

$$1: \frac{1}{1.60} = 62.50\% \quad X: \frac{1}{3.40} = 29.41\% \quad 2: \frac{1}{5.60} = 17.85\%$$

Come si può notare, la somma dei reciproci è maggiore del 100%; infatti:

$$62.50 + 29.41 + 17.85 = 109,76$$

Tale cifra è chiamata **Lavagna**.

Da questa cifra si può ricavare a quanto ammonta il profitto dei bookmakers per ogni scommessa e quanto vince uno scommettitore in caso di esito positivo della scommessa.

Di seguito riportiamo la quota di retrocessione che viene sborsata dalle agenzie di scommesse in caso di vittoria dello scommettitore (a titolo di esempio, consideriamo ancora la gara Milan-Sampdoria):

$$\frac{100\%}{109.76\%} = 91.1\%$$

Come si calcola il profitto dei bookmakers?

Si ottiene sottraendo dal 100% la quota di retrocessione; in questo caso è pari a 8,9%. Infatti:

$$100\% - 91,1\% = 8,9\%.$$

La lavagna non è l'unico metodo utilizzato dalle agenzie per evitare perdite.

Ma non finisce qui. I bookmakers utilizzano anche alcune strategie per tentare di indurre all'errore gli scommettitori:

1. Si adeguano alla percezione degli scommettitori. Infatti anche in caso di pessimo stato di forma, viene sempre preferita la squadra ritenuta più forte (alla quale quindi si assegna

una quota bassa) per invogliare a piazzare una scommessa all'apparenza facile, ma con una bassa probabilità di vincita.

2. Modificano le quote. Nei giorni precedenti all'evento specifico, i bookmakers possono cambiare le quote di un determinato evento abbassandole o alzandole basandosi sul numero di scommesse già piazzate.

5.3 Differenza tra quote reali e quote proposte

Le agenzie di scommesse, per i motivi già citati, non presentano mai le quote reali agli scommettitori.

Come si calcola la quota reale?

Si prenda come esempio l'evento "1" in Milan-Sampdoria del 2008/2009, quotato a 1,30.

La probabilità di vincita in una quota di 1,30 è $\frac{100}{1.30} = 76.9\%$.

- Si calcola poi la *singola probabilità*, dividendo per 100 la somma dei reciproci delle quote delle scommesse che si possono piazzare (1,X,2). In questo caso la somma dei reciproci è 109,8%, la singola probabilità vale 1,098.
- Si calcola poi il rapporto tra la probabilità dell'evento e la singola probabilità:

$$\frac{76.9\%}{1.098} = 70.05\%$$

- Si divide 100 per il risultato ottenuto $\frac{100}{70.05\%} = 1.43$.

In tal modo si trova la quota reale che, come si può notare, è maggiore della quota proposta.

5.4 Come calcolare una vincita

Per calcolare la vincita bisogna distinguere tra la singola scommessa e la scommessa multipla.

- SCOMMESSA SINGOLA

Si prenda come esempio un evento quotato a 1,63 sul quale si puntano 100 €.

Vincita lorda = puntata × quota = 1,63 × 100 = 163 €

Vincita netta = vincita lorda – puntata = 163 – 100 = 63 €

- SCOMMESSA MULTIPLA

In una scommessa multipla si scommette su più eventi. Per vincere devono avere tutti esito positivo.

Si prendano come esempio tre eventi, quotati 1,50 1,63 e 3,29 sul quale si puntano 100 €.

Vincita lorda = puntata × (quota 1 × quota 2 × quota 3) = 100 × (1,50 × 1,63 × 3,29) = 804,4 €

Vincita netta = vincita lorda – puntata = 804,4 – 100 = 704,4 €

Questa vincita non prende in considerazione i bonus.

5.5 Bonus

I bonus vengono applicati quando in una scommessa in cui si punta su più di una partita sono presenti almeno 5 eventi nei quali la quota dell'evento giocato (per esempio, vittoria della squadra in trasferta) è maggiore di 1,25

N. EVENTI > BONUS	N. EVENTI > BONUS	N. EVENTI > BONUS	N. EVENTI > BONUS	N. EVENTI > BONUS
5 > + 5%	6 > + 10%	7 > + 16%	8 > + 22%	9 > + 28%
10 > + 34%	11 > + 41%	12 > + 48%	13 > + 55%	14 > + 63%
15 > + 71%	16 > + 80%	17 > + 89%	18 > + 98%	19 > + 108%
20 > + 118%	...	25 > + 179%	26 > + 193%	27 > + 207%
28 > + 223%	29 > + 239%	30 > + 256%		

Tab. 24 – BONUS SNAI (la percentuale rappresenta l'aumento nella vincita)

Acknowledgements

Research carried out in collaboration with the Big&Open Data Innovation Laboratory (BODal-Lab), University of Brescia (project nr. 03-2016, title "Big Data analytics in sports", www.bodai.unibs.it/bdsports/), granted by Fondazione Cariplo and Regione Lombardia.

Riferimenti bibliografici e sitografia

- [1] S. Borra, A. Di Ciaccio, *Statistica. Metodologie per le scienze economiche e sociali*, Mc Graw-Hill, Milano (2008)
- [2] P. Newbold, W.L. Carlson, B. Thorne, *Statistica*, 2° ed., Pearson Italia, Milano (2010)
- [3] R. Natalini, *Introduzione ai metodi numerici alle differenze finite per equazioni di evoluzione*, http://www.dmmm.uniroma1.it/pubblicazioni/doc/phd_quaderni/04-02-nat.pdf
- [4] F. Flandoli, *Il test chi quadro e la relativa distribuzione*, <http://users.dma.unipi.it/~flandoli/chisquare.pdf>
- [5] *Adattamento di modelli probabilistici a distribuzioni empiriche*, <http://www.edutecnica.it/calcolo/ada/ada.htm>
- [6] D. Costanzo, *Distribuzione di Poisson*, http://www.ecostat.unical.it/Costanzo/Didattica/Probabilit%C3%A0%20ed%20Inferenza%20Statistica/lucidi_10.pdf
- [7] G. Sanfilippo, *Distribuzione di Poisson*, http://www1.unipa.it/sanfilippo/pub/stad/Lezioni2011_2012/MaterialeDidattico/lezioni2011_2012_18_22.pdf
- [8] *Tavola della distribuzione CHI-QUADRATO*, http://www00.unibg.it/dati/corsi/40025/74822-tavola_chi2.pdf
- [9] www.superscommesse.it
- [10] www.artandars.it
- [11] www.bet365.com
- [12] www.eurosport.com

Image credits

<https://pixabay.com/it/>



Gruppo formato da: Andrea Moscheni, Matteo Parise, Leonardo Perozzo, Lorenzo Luigi Scarsella

Tutor scolastico: Meneghini Lorenzo

Progetto realizzato in collaborazione con l'Associazione Nemesis

Tutor Aziendale: dott.ssa Borriero Sonia

Editing in collaborazione con l'Associazione Nemesis

