

Penerapan Distribusi Poisson Bivariat pada Data Jumlah Gol Hasil Pertandingan Sepak Bola Liga 1 Indonesia Tahun 2018-2019

Alfan Siam Nuri*, Aceng Komarudin Mutaqin

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

*alfansiamnur787@gmail.com, aceng.k.mutaqin@gmail.com

Abstract. On every continent in the world there are several prestigious league competitions that roll on and become a place to meet the best players, for example on the continent of Europe there are some of the best leagues. For example the English Premier League (EPL) in England, La Liga in Spain, Serie A in Italy, the Bundesliga in Germany and Ligue 1 in France. In Indonesia itself there is a league competition which is currently known as the Indonesian Super League (ISL) or the Indonesian League 1 which is the highest caste football competition between clubs in Indonesia which is participated in by 18 teams. There are several univariate and bivariate distributions that can be used to model data on the number of goals scored by football teams. One such bivariate distribution is the bivariate Poisson distribution. The maximum likelihood estimator method is used to estimate the parameters of the discrete distribution. While the distribution fit test to be used is the chi-square test. The data that will be used contains information on the number of goals for the home team and the number of goals for the away team in the 2018-2019 Indonesian Super League matches. The results of the application show that the bivariate Poisson distribution is suitable for modeling bivariate data on the number of goals for the home team and the away team for the 2018-2019 Indonesian Super League.

Keywords: *Poisson Bivariate Distribution, Chi-Square Test, Indonesian Super League, Maximum Likelihood Estimation.*

Abstrak. Di setiap benua di dunia terdapat beberapa kompetisi liga bergengsi yang bergulir dan menjadi ajang bertemu pemain-pemain terbaik, misalnya benua Eropa terdapat beberapa liga terbaik. Contohnya *English Premier League (EPL)* di Inggris, *La Liga* di Spanyol, *Serie A* di Italia, *Bundes Liga* di Jerman dan *Ligue 1* di Prancis. Di Indonesia sendiri terdapat kompetisi liga yang pada saat ini dikenal dengan sebutan *Indonesian Super League (ISL)* atau Liga 1 Indonesia yang merupakan kompetisi sepak bola kasta tertinggi antar klub di Indonesia yang diikuti oleh 18 tim. Terdapat beberapa distribusi univariat dan bivariat yang dapat digunakan untuk memodelkan data jumlah gol tim sepakbola. Salah satu distribusi bivariat tersebut adalah distribusi Poisson bivariat. Metode penaksir kemungkinan maksimum digunakan untuk menaksir parameter distribusi diskrit tersebut. Sedangkan uji kecocokan distribusi yang akan digunakan adalah uji chi-kuadrat. Data yang akan digunakan tersebut berisi informasi jumlah gol tim kandang dan jumlah gol tim tandang pertandingan Liga 1 Indonesia tahun 2018-2019. Hasil penerapan menunjukkan bahwa distribusi Poisson bivariat cocok untuk memodelkan data bivariat jumlah gol tim kandang dan tim tandang Liga 1 Indonesia tahun 2018-2019.

Kata Kunci: *Distribusi Poisson Bivariat, Uji Chi-Kuadrat, Liga 1 Indonesia, Penaksir Kemungkinan Maksimum.*

A. Pendahuluan

Secara umum, sepak bola merupakan olahraga yang menggunakan bola dengan bahan kulit dan dimainkan oleh dua tim. Masing-masing tim memiliki 11 pemain inti dan beberapa pemain pengganti. Namun ada juga yang menyatakan bahwa sepak bola adalah permainan yang dilakukan dengan cara menendang bola yang memiliki tujuan memasukkan bola ke dalam gawang lawan. Dalam bahasa Inggris, sepak bola dikenal dengan sebutan *football* atau *soccer* dalam bahasa Inggris-Amerika Serikat.

Di setiap benua di dunia menyelenggarakan beberapa kompetisi liga bergengsi yang terus berlanjut dan menjadi ajang bertemunya para pemain terbaik, contohnya benua Eropa memiliki beberapa liga terbaik. Misalnya *English Premiere League (EPL)* di Inggris, *La Liga* di Spanyol, *Serie A* di Italia, *Bundes Liga* di Jerman dan *Ligue 1* di Prancis.

Di Indonesia sendiri, terdapat kompetisi liga yang pada saat ini dikenal dengan nama Liga Super Indonesia atau sekarang disebut Liga 1 Indonesia, yaitu kompetisi sepak bola antar klub tingkat atas di Indonesia yang diikuti oleh 18 tim.

Ada beberapa distribusi univariat yang dapat digunakan untuk memodelkan data jumlah gol tim sepakbola. Distribusi ini termasuk Poisson, binomial negatif, dan ZIP (*Zero Inflated Poisson*) (Ibrahim dan Mutaqin, 2021). Menurut Diniz (2017), beberapa model bivariat dipertimbangkan untuk memodelkan data bivariat jumlah gol (gol tim kandang dan gol tim tandang) dalam suatu pertandingan sepakbola. Salah satu distribusi bivariat tersebut adalah distribusi Poisson bivariat.

Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah dalam penelitian ini sebagai berikut: “bagaimana menerapkan distribusi Poisson bivariat pada data bivariat jumlah gol tim kandang dan tim tandang hasil pertandingan Liga 1 Indonesia tahun 2018-2019”. Selanjutnya, tujuan dalam penelitian ini adalah menerapkan distribusi Poisson bivariat pada data bivariat jumlah gol tim kandang dan tim tandang hasil pertandingan Liga 1 Indonesia tahun 2018-2019.

B. Metodologi Penelitian

Distribusi Poisson

Distribusi Poisson adalah suatu distribusi untuk peristiwa yang probabilitas kejadiannya kecil, dimana kejadiannya bergantung pada selang waktu tertentu atau di wilayah tertentu, dan hasil pengamatannya berupa data diskrit. Misalnya selang waktu tertentu yang dimaksud bisa berupa panjang, berapa menit, berapa hari, berapa minggu, berapa bulan bahkan berapa tahun. Wilayah tertentu yang dimaksud dapat berupa suatu garis, suatu luas, atau suatu volume (Walpole, 1995).

Peubah acak diskrit X dikatakan berdistribusi Poisson dengan fungsi massa peluang sebagai berikut:

$$f(x, y) = \begin{cases} p(x, \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} & ; \text{ untuk } x = 0, 1, 2, \dots \\ 0 & ; \text{ untuk } x \text{ lainnya} \end{cases}$$

Ekspektasi untuk distribusi Poisson adalah

$$E(X) = \lambda$$

Varians untuk distribusi Poisson adalah

$$Var(X) = \lambda$$

Misalkan X_1, X_2, \dots, X_n adalah suatu sampel acak berukuran n dari distribusi Poisson dengan parameter λ , dimana nilai dari sampel acak tersebut adalah x_1, x_2, \dots, x_n . Penaksir kemungkinan maksimum untuk parameter λ dari distribusi Poisson adalah rata-rata sampelnya, yaitu:

$$\hat{\lambda} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Distribusi Poisson Bivariat

Distribusi Poisson bivariat terjadi ketika terdapat peubah acak Z_1, Z_2 dan Z_3 berdistribusi Poisson dengan parameter λ_1, λ_2 dan λ_3 . Misalkan peubah acak X dan Y terbentuk dari peubah Z_1, Z_2 dan Z_3 yaitu $X = Z_1 + Z_3$ dan $Y = Z_2 + Z_3$ sehingga $E(X) = \lambda_1 + \lambda_3$ dan $E(Y) = \lambda_2 + \lambda_3$ dengan λ_3 merupakan kovarians antara X dan Y (Karlis and Ntzoufras, 2003).

Secara bersama-sama peubah acak X dan Y berdistribusi Poisson bivariat memiliki fungsi massa peluang sebagai berikut (Holgate, 1964):

$$p(x, y) = P(X = x, Y = y) = e^{-(\lambda_1 + \lambda_2 + \lambda_3)} \sum_{k=0}^{\min(x,y)} \frac{\lambda_1^{x-k} \lambda_2^{y-k} \lambda_3^k}{(x-k)! (y-k)! k!}$$

Penaksir kemungkinan maksimum dari $\lambda_1 + \lambda_3$ adalah $\sum_{i=1}^n \frac{x_i}{n}$ dan untuk $\lambda_2 + \lambda_3$ adalah $\sum_{i=1}^n \frac{y_i}{n}$. Sedangkan penaksir kemungkinan maksimum dari λ_1, λ_2 dan λ_3 adalah solusi dari 3 persamaan berikut:

$$\frac{1}{n} \sum_{i=1}^n \frac{p(x_i - 1, y_i)}{p(x_i, y_i)} = 1$$

$$\frac{1}{n} \sum_{i=1}^n \frac{p(x_i, y_i - 1)}{p(x_i, y_i)} = 1$$

dan

$$\frac{1}{n} \sum_{i=1}^n \frac{p(x_i - 1, y_i - 1)}{p(x_i, y_i)} = 1$$

Uji Chi-Kuadrat

Uji kecocokan distribusi chi-kuadrat adalah uji hipotesis statistik yang digunakan untuk menentukan apakah x_1, x_2, \dots, x_n adalah nilai dari sampel acak X_1, X_2, \dots, X_n yang berdistribusi $F(\cdot)$. Uji kecocokan distribusi dapat digunakan untuk menguji hipotesis berikut:

$H_0 : x_1, x_2, \dots, x_n$; nilai dari sampel acak berasal dari populasi yang berdistribusi tertentu dengan fungsi distribusi $F(\cdot)$

$H_1 : x_1, x_2, \dots, x_n$; nilai dari sampel acak bukan berasal dari populasi yang berdistribusi tertentu dengan fungsi distribusi $F(\cdot)$

Statistik uji untuk uji chi-kuadrat adalah:

$$\chi^2 = \sum_{x=1}^m \frac{(n_x - np_x)^2}{np_x}$$

dalam hal ini n adalah banyaknya pengamatan, n_x adalah banyaknya pengamatan untuk kategori x , p_x adalah peluang untuk kategori x dan m adalah banyaknya kategori. Statistik uji di atas berdistribusi chi-kuadrat dengan derajat bebas $m - r - 1$, dimana r menyatakan banyaknya parameter yang ditaksir dari distribusi. Dengan taraf nyata α , tolak H_0 jika nilai statistik uji yang ada pada Persamaan (2.3) lebih besar dari nilai kuantil distribusi chi-kuadrat atau tolak H_0 jika $\chi^2 \geq \chi_{(m-r-1)(1-\alpha)}^2$. Apabila terdapat kategori yang nilai harapan (np_x) < 5, maka kategori-kategori tersebut digabungkan untuk menghasilkan kategori baru dengan nilai harapan (np_x) ≥ 5 .

C. Hasil Penelitian dan Pembahasan

Data yang dipakai dalam penelitian ini adalah data sekunder yang didapat dari *rsssf.com* dan *flashscore.com* mengenai hasil-hasil pertandingan Liga 1 Indonesia tahun 2018-2019. Data yang digunakan berisi informasi mengenai hasil skor akhir setiap pertandingan tim kandang dan tim tandang. Distribusi yang diuji dikatakan cocok jika nilai statistik uji lebih kecil dari nilai kuantil

chi-kuadratnya.

Tabel 1. Data Bivariat Jumlah Gol Tim Kandang dan Tim Tandang

Pertandingan	Jumlah Gol Tim Kandang	Jumlah Gol Tim Tandang	Pertandingan	Jumlah Gol Tim Kandang	Jumlah Gol Tim Tandang	Pertandingan	Jumlah Gol Tim Kandang	Jumlah Gol Tim Tandang
1	3	1	201	4	1	277	1	1
2	1	2	202	6	0	278	1	0
3	2	1	203	0	2	279	0	1
4	1	1	204	1	2	280	3	2
5	1	5	205	2	2	281	4	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
110	1	3	245	1	0	301	3	0
111	4	2	246	1	2	302	5	2
112	2	2	247	1	2	303	2	2
113	0	2	248	1	1	304	5	2
114	4	1	249	1	0	305	0	2
115	2	4	250	0	0	306	2	3

Tabel 1 menyajikan data bivariat jumlah gol tim kandang dan tim tandang hasil pertandingan Liga 1 Indonesia tahun 2018-2019. Sebagai contoh, pada pertandingan ke 1 (kolom ke 1) hasil akhir dari pertandingan tersebut berakhir dengan skor 3-1 (kolom ke 2 untuk jumlah gol tim kandang dan kolom ke 3 untuk jumlah gol tim tandang), pada pertandingan ke 2 hasil akhir dari pertandingan tersebut berakhir dengan skor 1-2 dan seterusnya sampai pertandingan terakhir yaitu pertandingan ke 306 dengan hasil akhir dari pertandingan berakhir dengan skor 2-3.

Tabel 2. Frekuensi Data Bivariat Jumlah Gol Tim Kandang dan Tim Tandang

Jumlah Gol Tim Kandang	Jumlah Gol Tim Tandang					
	0	1	2	3	4	5
0	19	23	12	1	2	0
1	35	44	14	7	2	1
2	24	23	25	7	1	0
3	14	9	11	1	0	0
4	8	6	3	3	0	0
5	0	4	3	1	0	0
6	1	1	1	0	0	0

Tabel 2 menyajikan data frekuensi dari data bivariat jumlah gol tim kandang dan tim tandang untuk semua pertandingan Liga 1 Indonesia tahun 2018-2019. Terdapat 19 pertandingan yang berakhir dengan kedudukan 0-0 (tim kandang dan tim tandang tidak berhasil mencetak gol), terdapat 35 pertandingan yang berakhir dengan kedudukan 1-0 (tim kandang berhasil mencetak 1 gol dan tim tandang tidak berhasil mencetak gol), terdapat 23 pertandingan yang berakhir dengan kedudukan 0-1 (tim kandang tidak berhasil mencetak gol dan tim tandang berhasil mencetak 1 gol) dan seterusnya.

Tabel 3. Frekuensi Data Univariat Jumlah Gol Tim Kandang

Jumlah Gol Tim Kandang	Jumlah Pertandingan
0	57
1	103
2	80
3	35
4	20
5	8
6	3

Tabel 3 menyajikan data frekuensi dari data univariat jumlah gol tim kandang untuk semua pertandingan Liga 1 Indonesia tahun 2018-2019. Pada Liga 1 Indonesia tahun 2018-2019 jumlah gol tim kandang 0 ada sebanyak 57 pertandingan, jumlah gol tim kandang 1 ada sebanyak 103 pertandingan, jumlah gol tim kandang 2 ada sebanyak 80 pertandingan, jumlah gol tim kandang 3 ada sebanyak 35 pertandingan, jumlah gol tim kandang 4 ada sebanyak 20 pertandingan, jumlah gol tim kandang 5 ada sebanyak 8 pertandingan dan jumlah gol tim kandang 6 ada sebanyak 3 pertandingan.

Tabel 4. Frekuensi Data Univariat Jumlah Gol Tim Tandang

Jumlah Gol Tim Tandang	Jumlah Pertandingan
0	101
1	110
2	69
3	20
4	5
5	1

Tabel 4 menyajikan data frekuensi dari data univariat jumlah gol tim tandang untuk semua pertandingan Liga 1 Indonesia musim tahun 2018-2019. Pada Liga 1 Indonesia tahun 2018-2019 jumlah gol tim tandang 0 ada sebanyak 101 pertandingan, jumlah gol tim tandang 1 ada sebanyak 110 pertandingan, jumlah gol tim tandang 2 ada sebanyak 69 pertandingan, jumlah gol tim tandang 3 ada sebanyak 20 pertandingan, jumlah gol tim tandang 4 ada sebanyak 5 pertandingan dan jumlah gol tim tandang 5 ada sebanyak 1 pertandingan.

Tabel 5. Nilai-nilai yang Dibutuhkan untuk Perhitungan Statistik Uji Jumlah Gol Tim Kandang

Jumlah Gol Tim Kandang (x)	Jumlah Pertandingan (n_x)	Peluang Jumlah Gol Tim Kandang (\hat{p}_x)	Nilai Harapan Jumlah Gol Tim Kandang ($n\hat{p}_x$)	$\frac{(n_x - n\hat{p}_x)^2}{n\hat{p}_x}$
(1)	(2)	(3)	(4)	(5)
0	57	0,1914	59	0,0678
1	103	0,3164	97	0,3711
2	80	0,2616	80	0,0000
3	35	0,1442	44	1,8409
4	20	0,0596	18	0,2222
≥ 5	11	0,0268	8	1,125
Jumlah	306	1	306	3,627

Nilai statistik uji Chi-kuadrat ada dalam Tabel 5 kolom (5) baris terakhir, yaitu 3,627. Dengan taraf nyata 5%, nilai kuantil distribusi Chi-kuadrat dengan derajat bebas 4 ($6 - 1 - 1$) adalah 9,488. Terlihat bahwa nilai statistik ujinya lebih kecil dibandingkan dengan kuantilnya ($3,627 < 9,488$). Dengan demikian H_0 diterima dan disimpulkan bahwa data jumlah gol tim kandang Liga 1 Indonesia tahun 2018-2019 berasal dari populasi yang berdistribusi Poisson.

Tabel 6. Nilai-nilai yang Dibutuhkan untuk Perhitungan Statistik Uji Jumlah Gol Tim Tandang

Jumlah Gol Tim Tandang (x)	Jumlah Pertandingan (n_x)	Peluang Jumlah Gol Tim Tandang (\hat{p}_x)	Nilai Harapan Jumlah Gol Tim Tandang ($n\hat{p}_x$)	$\frac{(n_x - n\hat{p}_x)^2}{n\hat{p}_x}$
(1)	(2)	(3)	(4)	(5)
0	101	0,3368	103	0,0388
1	110	0,3665	113	0,0796
2	69	0,1994	61	1,0492
3	20	0,0723	22	0,1818
≥ 4	6	0,024	7	0,1429
Jumlah	306	1	306	1,4923

Nilai statistik uji Chi-kuadrat ada dalam Tabel 6 kolom (5) baris terakhir, yaitu 1,4923. Dengan taraf nyata 5%, nilai kuantil distribusi Chi-kuadrat dengan derajat bebas 3 ($5 - 1 - 1$) adalah 7,815. Terlihat bahwa nilai statistik ujinya lebih kecil dibandingkan dengan kuantilnya ($1,4923 < 7,815$). Dengan demikian H_0 diterima dan disimpulkan bahwa data jumlah gol tim tandang Liga 1 Indonesia tahun 2018-2019 berasal dari populasi yang berdistribusi Poisson.

Tabel 7. Taksiran dan Data Sesungguhnya Data Bivariat Jumlah Gol Tim Kandang dan Tim Tandang

Jumlah Gol Tim Kandang	Jumlah Gol Tim Tandang					
	0	1	2	3	4	5
0	21 (19)	22 (23)	11 (12)	4 (1)	1 (2)	0 (0)
1	33 (35)	36 (44)	19 (14)	7 (7)	2 (2)	0 (1)
2	26 (24)	29 (23)	16 (25)	6 (7)	2 (1)	0 (0)
3	14 (14)	16 (9)	9 (11)	4 (1)	1 (0)	0 (0)
4	5 (8)	7 (6)	4 (3)	2 (3)	0 (0)	0 (0)
5	2 (0)	2 (4)	1 (3)	1 (1)	0 (0)	0 (0)
6	0 (1)	1 (1)	0 (1)	0 (0)	0 (0)	0 (0)

Dalam Tabel 7 disajikan taksiran dan nilai sesungguhnya data bivariat jumlah gol tim kandang dan tim tandang Liga 1 Indonesia tahun 2018-2019. Terlihat bahwa taksiran dan nilai sesungguhnya tidak jauh berbeda. Hal ini menjadi satu bukti bahwa distribusi Poisson bivariat cocok untuk data bivariat jumlah gol tim kandang dan tim tandang Liga 1 Indonesia tahun 2018-2019.

D. Kesimpulan

Berdasarkan hasil analisis penerapan distribusi Poisson bivariat pada data bivariat jumlah gol tim kandang dan tim tandang hasil pertandingan sepak bola Liga 1 Indonesia tahun 2018-2019 dapat disimpulkan bahwa distribusi Poisson bivariat cocok untuk memodelkan untuk data jumlah gol tim kandang dan tim tandang.

Acknowledge

Alhamdulillah, for all the blessings that Allah SWT has given. I dedicate this thesis as a sign of my responsibility to my parents who always pray for, give advice, encourage and give support.

Don't forget to say thank you to my mentor and the people i love and love me for their help and support to me. May Allah will give you with a lot of goodness.

Daftar Pustaka

- [1] Diniz, M.A., dkk (2017). *Comparing probabilistic predictive models applied to football*. Journal of the Operational Research Society. Vol.70 (3).
- [2] Holgate, P. (1964). *Estimation for the bivariate Poisson distribution*. *Biometrika*, 51 (1-2) : 241-245.
- [3] Karlis, D. and Ntzoufras, I. (2003). *Analysis of sports data by using bivariate Poisson models*. Journal of the Royal Statistical Society: Series D (The Statistician), 52 (3) : 381-393.
- [4] Kawamura, K., (1984). *Direct Calculation Of Maximum Likelihood Estimator For The Bivariate Poisson Distribution*, *Kôdai Math. Sem. Rep.*, 25, No. 2, (211-221).
- [5] Mutaqin, A.K., Adri, Y.P. (2019). *Prediksi Klasemen Akhir Kompetisi Sepakbola Indonesia Menggunakan Metode Perluasan Ekspektasi Phytagoras*. Journal of Data Analysis, Vol. 2, No. 1, 22-30.
- [6] Nurhayati, Siti, Suliadi. (2022). *Diagram Kendali Multivariat Poisson untuk Pengendalian Kualitas Proses Produksi di PT. X*. Jurnal Riset Statistika 2(1). 58-66.