

Regresi *Conway–Maxwell–Poisson* untuk Jumlah Gol Tuan Rumah pada Sepak Bola *English Premier League* (EPL)

Tasya Pujiati Sa'diyah *, Abdul Kudus

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

tasyapujis250202@gmail.com, abdul.kudus@unisba.ac.id

Abstract. Poisson regression is a commonly used model for count data, but it has limitations because it assumes equidispersion, where the variance equals the mean. In practice, many datasets exhibit overdispersion (variance greater than the mean) or underdispersion (variance smaller than the mean). To address this issue, the Conway-Maxwell-Poisson (COM-Poisson) regression is used, as it can handle both conditions. This study aims to examine the effect of nine odds variables : Pinnacle home win odds (X1), Pinnacle draw odds (X2), Pinnacle away win odds (X3), maximum home win odds (X4), market maximum draw odds (X5), maximum away win odds (X6), market average home win odds (X7), market average draw odds (X8), and market average away win odds (X9) on the number of home team goals (Y) in English Premier League (EPL) matches during the 2023–2024 season. Using secondary data from football-data.co.uk and the Maximum Likelihood Estimation (MLE) method, the analysis revealed the presence of overdispersion, making the COM-Poisson model more appropriate than the classic Poisson model. Three variables, X3, X4, and X7, were found to be significant. The COM-Poisson model provided more stable and accurate parameter estimates in modeling goal count data.

Keywords: *COM-Poisson Regression, Overdispersion, English Premier League.*

Abstrak. Regresi Poisson merupakan model yang umum digunakan untuk memodelkan data cacah, tetapi memiliki keterbatasan karena mengasumsikan *equidispersi*, yaitu varians sama dengan rata-rata. Dalam praktiknya, banyak data mengalami *overdispersi* (variens lebih besar dari rata-rata) atau *underdispersi* (variens lebih kecil dari rata-rata). Untuk mengatasi hal ini, regresi Conway-Maxwell-Poisson (COM-Poisson) digunakan karena mampu menangani kedua kondisi tersebut. Penelitian ini bertujuan menguji pengaruh sembilan variabel *odds* yaitu *Pinnacle home win odds* (X1), *Pinnacle draw odds* (X2), *Pinnacle away win odds* (X3), *maximum home win odds* (X4), *market maximum draw odds* (X5), *maximum away win odds* (X6), *market average home win odds* (X7), *market average draw odds* (X8), dan *market average away win odds* (X9) terhadap jumlah gol tim tuan rumah (Y) pada pertandingan *English Premier League* (EPL) musim 2023–2024. Menggunakan data sekunder dari football-data.co.uk dan metode *Maximum Likelihood Estimation* (MLE), hasil analisis menunjukkan adanya *overdispersi* sehingga model COM-Poisson lebih tepat digunakan dibandingkan Poisson klasik. Tiga variabel, yaitu X3, X4, dan X7, terbukti signifikan. Model COM-Poisson memberikan estimasi parameter yang lebih stabil dan akurat dalam memodelkan data jumlah gol.

Kata Kunci: *Regresi Conway-Maxwell-Poisson, Overdispersi, Sepak Bola English Premier League (EPL).*

A. Pendahuluan

Sepak bola dikenal sebagai salah satu jenis olahraga dengan penggemar global yang sangat besar. Di antara berbagai liga profesional, *English Premier League* (EPL) menempati posisi istimewa sebagai salah satu liga paling populer dan kompetitif di dunia. Dalam setiap musimnya, ribuan data pertandingan tercatat termasuk jumlah gol yang dicetak dan hasil pertandingan yang menyimpan informasi penting untuk dianalisis secara statistik. Jumlah gol yang dicetak oleh tim tuan rumah dalam setiap pertandingan merupakan salah satu aspek menarik yang layak untuk diteliti lebih lanjut.

Data jumlah gol tersebut merupakan jenis data cacah yang pada umumnya dimodelkan menggunakan regresi Poisson, dengan asumsi bahwa nilai rata-rata sama dengan nilai variansnya. Namun, dalam praktiknya, sering kali ditemukan bahwa varian dari data jauh berbeda dari rata-ratanya, baik lebih besar (*overdispersi*) maupun lebih kecil (*underdispersi*). Hal ini menyebabkan model Poisson klasik menjadi kurang tepat karena mengasumsikan sifat *equidispersi*. Oleh karena itu, diperlukan model alternatif yang lebih fleksibel seperti regresi Conway-Maxwell-Poisson (COM-Poisson) yang mampu mengakomodasi berbagai kondisi dispersi pada data cacah.

Studi terdahulu oleh Alrweili (2024) mengembangkan estimasi *hybrid* baru, yaitu Kibrida-Lukman *Hybrid* Estimator dalam model regresi COM-Poisson. Penelitian tersebut berhasil mengindikasikan bahwa penggunaan model COM-Poisson mampu menghasilkan estimasi yang lebih akurat dibandingkan Poisson klasik ketika diterapkan pada data jumlah gol tim tuan rumah dalam pertandingan sepak bola, terutama ketika data menunjukkan adanya *overdispersi* atau *underdispersi*. Dalam studinya, Alrweili menguji sembilan variabel *odds* sebagai variabel bebas, dan menunjukkan bahwa beberapa di antaranya memiliki pengaruh signifikan terhadap jumlah gol tim tuan rumah.

Dengan mempertimbangkan kesesuaian karakteristik data EPL 2023/2024 yang menunjukkan gejala *overdispersi*, serta adanya referensi dari Alrweili, penggunaan regresi COM-Poisson menjadi pilihan tepat untuk dianalisis. Model ini dapat menangani ketidaksesuaian varian dan memberikan estimasi parameter yang lebih stabil.

Lebih lanjut, pemilihan data *odds* yang aktual, seperti *Pinnacle odds* dan *market odds*, memberikan keunggulan tersendiri. Selain sifatnya yang bersifat publik dan mudah diakses, *odds* juga menggambarkan informasi yang sangat kontekstual dan dinamis, yang relevan dalam menjelaskan hasil pertandingan, khususnya produktivitas tim tuan rumah dalam mencetak gol.

Penelitian ini menggunakan data sekunder pertandingan EPL musim 2023–2024 yang diperoleh dari situs [football-data.co.uk](https://www.football-data.co.uk).

Secara keseluruhan, penelitian ini akan memberikan kontribusi baru pada model Regresi COM-Poisson dalam menangani data. Pendekatan ini tidak hanya sesuai untuk analisis data olahraga tetapi juga memiliki potensi untuk diterapkan di berbagai bidang lain yang memerlukan analisis data cacah.

B. Metode

Data yang digunakan dalam penelitian ini merupakan data sekunder pada pertandingan Liga *English Premier League* (EPL) musim 2023–2024 yang diperoleh dari website <https://www.football-data.co.uk/englandm.php>. Data tersebut terdiri dari 1 variabel respon (Y) dengan 9 variabel bebas (X). *Premier League* pertama kali digelar pada tahun 1992, setelah klub-klub besar Inggris memisahkan diri dari Divisi Pertama *Football League* untuk meningkatkan keuntungan dari hak siar televisi [10]. Sistem liga sepak bola Inggris tersusun dalam bentuk piramida yang kompleks, mencakup total 24 tingkatan. *English Premier League* (EPL) menjadi yang paling populer, diikuti oleh Liga *Championship*, Liga *One*, dan Liga *Two*. Setiap liga memiliki mekanisme promosi dan degradasi yang unik, di mana klub-klub bersaing untuk naik ke tingkat yang lebih tinggi atau menghindari penurunan ke tingkat yang lebih rendah [4]. Penelitian ini mempunyai tujuan yaitu mengetahui apakah variabel-variabel bebas memiliki pengaruh signifikan terhadap Jumlah Gol Tim Tuan Rumah (Y) pada Pertandingan Sepak Bola *English Premier League* (EPL) dengan menggunakan model regresi Poisson atau model regresi Conway-Maxwell-Poisson (COM-Poisson). Keterangan mengenai variabel-variabel yang digunakan dalam penelitian ini dapat dilihat pada table berikut:

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Satuan
Y	Jumlah gol tim tuan rumah	Rasio
X1	<i>Pinnacle home win odds</i>	Rasio
X2	<i>Pinnacle draw odds</i>	Rasio
X3	<i>Pinnacle away win odds</i>	Rasio
X4	<i>Maximum home win odds</i>	Rasio
X5	<i>Market maximum draw odds</i>	Rasio
X6	<i>Maximum away win odds</i>	Rasio
X7	<i>Market average home win odds</i>	Rasio
X8	<i>Market average draw odds</i>	Rasio
X9	<i>Market average away win odds</i>	Rasio

Berikut ini merupakan langkah-langkah pengerjaan dari penelitian:

1. Melakukan analisis deskriptif untuk memperoleh gambaran umum karakteristik data yang digunakan
2. Uji kecocokan distribusi Poisson dilakukan menggunakan Chi-Kuadrat yang dituliskan pada persamaan (1) untuk memeriksa apakah data respon sesuai dengan asumsi distribusi Poisson. Berikut uji kecocokan distribusi poisson [9]:

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad \dots (1)$$

Dengan O_i merupakan frekuensi observasi pada kategori ke- i , E_i merupakan frekuensi harapan pada kategori ke- i , k merupakan jumlah kategori (setelah penggabungan jika ada $E_i < 5$).

Dimana: $E_i = P(Y = y_i) \cdot n = \left(\frac{\lambda^y e^{-\lambda}}{y!} \right) \cdot n$

Dengan y_i merupakan nilai ke- i dari variabel Y , n : jumlah total pengamatan, λ : nilai rata-rata dari data observasi (λ diestimasi dari data $\lambda = \frac{\sum y_i \cdot O_i}{n}$). Menggunakan derajat bebas atau $df = k - 1 - m$, dengan $m = 1$ karena satu parameter (λ) diestimasi dari data. Dengan kriteria uji adalah jika $\chi^2_{hitung} < \chi^2_{tabel} \rightarrow$ terima H_0 dan jika $\chi^2_{hitung} \geq \chi^2_{tabel} \rightarrow$ tolak H_0 .

3. Selanjutnya melakukan pengujian multikolinearitas antar variabel bebas dengan menghitung nilai *Variance Inflation Factor* (VIF) yang dituliskan pada persamaan (2).

$$VIF = \frac{1}{1 - R_j^2} \text{ dengan } j = 1, 2, \dots, p \quad \dots (2)$$

Dengan R_j^2 adalah nilai koefisien determinasi yang menggambarkan hubungan antara variabel x_j dengan variabel x lainnya. Nilai $VIF \geq 10$ menunjukkan adanya indikasi multikolinieritas.

4. Equidispersi

Equidispersi adalah sifat distribusi Poisson bahwa nilai rata-rata dan varians data sama besar [8]. Pemeriksaan terhadap gejala *overdispersi* dengan cara membandingkan nilai deviansi terhadap derajat bebas (df) dari model menggunakan persamaan (3).

$$\frac{D}{db} > 1; D = 2 \sum_{i=1}^n \left[y_i \ln \frac{y_i}{\hat{y}_i} - (y_i - \hat{y}_i) \right] \quad \dots (3)$$

Dimana, D adalah nilai deviansi, $db = n - p$ adalah derajat bebas, dengan n jumlah observasi dan p jumlah parameter. Dengan y_i nilai observasi dan $\hat{y}_i$ nilai prediksi model. Jika nilai rasio deviansi terhadap derajat bebasnya > 1 , maka data mengalami *overdispersi*.

5. Regresi Poisson

Regresi Poisson adalah model dalam statistika yang diterapkan ketika variabel respons (Y) merupakan data hitung, seperti jumlah kejadian suatu peristiwa dalam satuan waktu atau ruang tertentu. Model ini termasuk dalam *Generalized Linear Model* (GLM) karena tidak mengasumsikan distribusi normal pada data dan menggunakan fungsi hubung [6]. *Generalized Linear Model* (GLM) adalah suatu struktur statistik yang memperluas konsep dari model regresi linier klasik untuk menangani berbagai tipe

data yang umum muncul dalam bidang asuransi dan aktuarial, seperti data klaim, data frekuensi, dan data biaya [2]. Distribusi Poisson memiliki fungsi massa peluang (*Probability Massa Function*) sebagai berikut:

$$f(y; \lambda) = \frac{\lambda^y e^{-\lambda}}{y!}; y = 0, 1, 2, \dots, \lambda > 0 \quad \dots (4)$$

Dengan $\lambda > 0$: parameter distribusi (juga rata-rata banyaknya kejadian dalam satuan waktu tertentu) dan y merupakan bilangan bulat non-negatif banyaknya kejadian per satuan waktu. Distribusi Poisson memiliki rata-rata sama dengan variansnya $E(Y) = Var(Y) = \lambda$. Dalam regresi Poisson, terdapat beberapa fungsi hubung yang umum digunakan. Salah satu yang paling sederhana adalah fungsi hubung identitas, yang merepresentasikan hubungan langsung antara rata-rata dan prediktor linear:

$$g(\lambda_i) = \lambda_i = X'_i \beta \quad \dots (5)$$

Dengan menggunakan fungsi ini, maka prediksi nilai respons adalah: $E(y_i) = \lambda_i = X'_i \beta$ karena $\lambda_i = g^{-1}(X'_i \beta) = X'_i \beta$. Alternatif lainnya yang populer adalah log hubung, yang didefinisikan sebagai:

$$g(\lambda_i) = \ln(\lambda_i) = X'_i \beta \quad \dots (6)$$

Dengan menggunakan log hubung, rata-rata dari variabel respons diperoleh melalui eksponensiasi dari prediktor linear:

$$\lambda_i = g^{-1}(X'_i \beta) = \exp(X'_i \beta) \quad \dots (7)$$

Fungsi log hubung ini memiliki keunggulan utama, yaitu memastikan bahwa semua nilai prediksi untuk respons tetap dalam rentang non-negatif, yang sesuai dengan karakteristik distribusi Poisson. Dalam regresi Poisson, proses estimasi parameter dilakukan dengan metode *Maximum Likelihood Estimation* (MLE), yang pada dasarnya sejalan dengan prosedur dalam regresi logistik. Misalkan terdapat data pengamatan acak sebanyak n dengan respons y_i dan prediktor x_i , maka fungsi *likelihood* dapat dituliskan sebagai:

$$L(\beta|y) = \prod_{i=1}^n f(y_i) = \prod_{i=1}^n \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!} = \frac{\prod_{i=1}^n \lambda_i^{y_i} \exp(-\sum_{i=1}^n \mu_i)}{\prod_{i=1}^n y_i!} \quad \dots (8)$$

Dengan $\lambda_i = \exp(X'_i \beta)$. Untuk mempermudah perhitungan, biasanya dilakukan transformasi ke dalam bentuk log-likelihood:

$$\ln L(y, \beta) = \sum_{i=1}^n y_i \ln(\mu_i) - \sum_{i=1}^n \mu_i - \sum_{i=1}^n \ln(y_i!) \quad \dots (9)$$

Prosedur numerik seperti *Iteratively Reweighted Least Squares* (IRLS) dapat digunakan untuk memperoleh estimasi parameter maksimum likelihood tersebut. Setelah diperoleh estimasi parameter $\hat{\beta}$, maka model regresi Poisson hasil estimasi dituliskan sebagai:

$$\hat{\lambda}_i = g^{-1}(X'_i \hat{\beta}) \quad \dots (10)$$

misalnya, jika log hubung dipilih, maka:

$$\hat{\lambda}_i = g^{-1}(X'_i \hat{\beta}) = \exp(X'_i \hat{\beta}) \quad \dots (11)$$

6. Regresi Conway-Maxwell-Poisson

Distribusi Conway-Maxwell-Poisson merupakan distribusi yang sangat fleksibel dalam menangani data cacah yang memiliki masalah dispersi. Fleksibilitas ini diperoleh melalui adanya parameter tambahan, yaitu ν , yang memungkinkan distribusi ini untuk menyesuaikan terhadap kondisi *overdispersi* ($\nu < 1$) maupun *underdispersi* ($\nu > 1$) [1]. Distribusi COM-Poisson ditentukan oleh dua parameter utama, yaitu λ sebagai parameter rata-rata dan ν sebagai parameter dispersi. Secara formal, apabila variabel acak y mengikuti distribusi COM-Poisson dengan parameter λ dan ν , maka bentuk fungsi probabilitas (*probability mass function*) dapat dinyatakan sebagai berikut:

$$P(Y = y; \lambda, \nu) = \frac{\lambda^y}{(y!)^\nu Z(\lambda, \nu)}; y = 0, 1, 2, \dots; \lambda > 0, \nu \geq 0 \quad \dots (12)$$

Parameter ν ($\nu \geq 0$) berfungsi mengendalikan tingkat dispersi dalam distribusi [7], sementara λ ($\lambda > 0$) merupakan parameter yang mewakili nilai tengah atau rata-rata. Dimana $Z(\lambda, \nu)$ merupakan konstanta normalisasi yang tidak memiliki bentuk tertutup dan didefinisikan sebagai deret tak hingga berikut:

$$Z(\lambda, \nu) = \sum_{i=0}^{\infty} \frac{\lambda^i}{(i!)^{\nu}} \quad \dots (13)$$

Kemudian fungsi $Z(\lambda, \nu)$ didekati dengan

$$Z(\lambda, \nu) \approx \frac{\exp\left(\nu \lambda^{\frac{1}{\nu}}\right)}{\lambda^{\frac{\nu-1}{2\nu}} (2\pi)^{(\nu-1)/2} \sqrt{\nu}} \quad \dots (14)$$

Untuk mendekati nilai rata-rata dan variansi dari distribusi COM-Poisson, digunakan pendekatan asimtotik terhadap konstanta $Z(\lambda, \nu)$. Berdasarkan pendekatan ini, rata-rata dan variansi dapat diaproksimasi sebagai:

$$E(Y) \approx \lambda^{\frac{1}{\nu}} + \frac{1}{2\nu} - \frac{1}{2} ; V(Y) \approx \frac{1}{\nu} \lambda^{\frac{1}{\nu}}$$

Selanjutnya, dengan melakukan reparametrisasi terhadap distribusi COM-Poisson, maka $\mu = \lambda^{\frac{1}{\nu}}$ dalam Persamaan (12). sehingga fungsi probabilitas berubah menjadi:

$$P(Y = y; \mu, \nu) = \frac{1}{S(\mu, \nu)} \left(\frac{\mu^y}{y!} \right)^{\nu} ; y = 0, 1, 2, \dots, \quad \dots (15)$$

Dengan konstanta $S(\mu, \nu) = \sum_{i=0}^{\infty} \left(\frac{\mu^i}{i!} \right)^{\nu}$. Pada bentuk ini rata-rata dan variansi dari persamaan (16) diturunkan sebagai berikut $E(X) \approx \mu + \frac{1}{2\nu} - \frac{1}{2}$ dan $V(X) \approx \frac{\mu}{\nu}$. Formulasi baru dari distribusi COM-Poisson memungkinkan pembentukan kerangka *Generalized Linear Model* (GLM) yang memudahkan interpretasi hasil melalui penggunaan fungsi hubung [3]. *Log-likelihood* dari fungsi probabilitas COM-Poisson yang telah direparametrisasi dan jika digunakan fungsi hubung log, yaitu $\eta_i = \log(\mu_i) = x_i^t \beta$, maka fungsi *log-likelihood* dapat ditulis sebagai berikut:

$$l(y_i; \beta, \nu) = \nu \sum_{i=1}^n y_i \eta_i - \nu \sum_{i=1}^n \log(y_i!) - \sum_{i=1}^n \log(S(\eta_i, \nu)) \quad \dots (16)$$

Untuk mengestimasi parameter yang tidak diketahui (β dan ν) melalui metode *Maximum Likelihood Estimation* (MLE), dilakukan pengambilan turunan parsial dari fungsi *log-likelihood* dari Persamaan (17) terhadap β dan ν dan disamakan dengan 0.

$$\frac{\partial l(y_i; \beta, \nu)}{\partial \beta_j} = \sum_{i=1}^n \left(y_i \nu - \frac{\partial}{\partial \eta_i} \log S(\eta_i, \nu) \right) x_{ij} = 0 ; j = 0, 1, 2, \dots, p+1, x_{i0} = 1 \dots (17)$$

$$\frac{\partial l(y_i; \beta, \nu)}{\partial \nu} = \sum_{i=1}^n \left(-\log(y_i!) - \frac{\partial}{\partial \nu} \log S(\eta_i, \nu) \right) = 0 \quad \dots (18)$$

Dalam proses estimasi parameter β , nilai ν terlebih dahulu harus diasumsikan tetap. Dalam kerangka model regresi COM-Poisson, parameter rata-rata dan dispersi dapat dimodelkan hubungannya dengan set kovariat [3], sebagai berikut:

$$\log(\hat{\mu}_i) = \hat{\beta}_0 + \sum_{j=1}^p \hat{\beta}_j x_{ij} = x_i^t \hat{\beta} \quad \dots (19)$$

Dimana $\nu_i = \exp(x_i^t \nu)$. Untuk memperoleh estimasi parameter, dapat digunakan metode *Iteratively Reweighted Least Squares* (IRLS) untuk menyelesaikan Persamaan (17) dan (18). *Estimasi Maximum Likelihood* dari parameter β untuk model Conway-Maxwell-Poisson untuk vektor β

dirumuskan sebagai berikut:

$$\hat{\beta}_{CMPMLE} = (X^T \hat{U} X)^{-1} X^T \hat{U} \hat{c} \quad \dots (20)$$

di

mana $c = \log(\hat{\mu}) + \frac{y - \hat{\mu}}{\hat{\mu}^2}$ adalah vektor variabel respons yang disesuaikan, dan $\hat{U}$ adalah matriks diagonal bobot, yaitu, $\hat{U} = \text{diag}(\hat{u}_i)$, di mana, nilai $\hat{u}_i$ ditentukan berdasarkan:

$$\hat{u}_i = \frac{\hat{t}_i}{\hat{v}} + \frac{\hat{v}^2 - 1}{24\hat{v}^3} \hat{t}_i^{-1} + \frac{\hat{v}^2 - 1}{12\hat{v}^4} \hat{t}_i^{-2} + \frac{\hat{v}^2 - 1}{6\hat{v}^5} \hat{t}_i^{-3} \quad \dots (21)$$

Dengan $\hat{t}_i = \frac{\hat{\mu}_i}{\hat{v}}$.

7. Pengujian Parameter Model secara Simultan dilakukan menggunakan uji *likelihood ratio* (uji G) pada persamaan (23), untuk mengetahui apakah variabel-variabel bebas secara bersama-sama memengaruhi variabel respon.

$$G = -2 \ln \left[\frac{L(\beta_0, v)}{L(\beta, v)} \right] \quad \dots (22)$$

Dimana $L(\beta_0, v)$ merupakan fungsi *likelihood* dari model tanpa variabel bebas, sedangkan $L(\beta, v)$ adalah fungsi *likelihood* dari model yang mencakup seluruh variabel bebas. Statistik uji yang diterapkan adalah uji G, yang mengikuti distribusi chi-kuadrat dengan derajat bebas sebesar p , yang menyatakan jumlah variabel bebas dalam model. Jika nilai $G > \chi^2_{(1-\alpha, p)}$, dan dengan menggunakan *p-value* $< \alpha$ maka tolak H_0 , artinya terdapat cukup bukti untuk menyimpulkan bahwa setidaknya terdapat satu variabel bebas yang memiliki pengaruh signifikan pada variabel respon.

8. Pengujian Parameter Model secara Parsial untuk mengetahui pengaruh masing-masing variabel bebas secara individual, dilakukan menggunakan uji Wald pada persamaan (24).

$$Z_{hitung} = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad \dots (23)$$

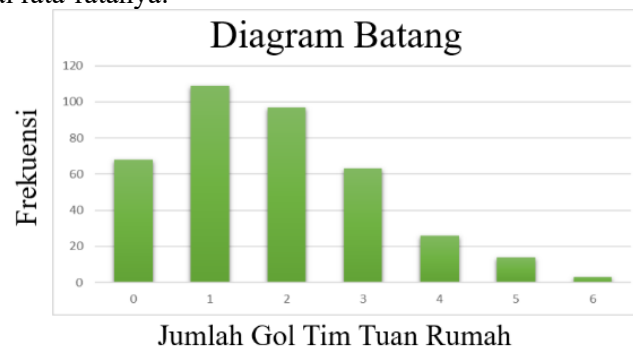
$\hat{\beta}_j$ merupakan penduga dari koefisien regresi untuk variabel bebas ke- j dan $SE(\hat{\beta}_j)$ sebagai *standard error* atau penduga galat baku $\hat{\beta}_j$. Kriteria ujinya tolak H_0 jika $|z| > z_{\alpha/2}$ dan terima H_0 jika $|z| \leq z_{\alpha/2}$. Pengujian juga dapat dilakukan dengan menggunakan nilai *P-value*, yaitu tolak H_0 jika *P-value* $< \alpha$ dan terima H_0 jika *P-value* $\geq \alpha$.

C. Hasil Penelitian dan Pembahasan Analisis Deskriptif

Tabel 2. Deskriptif Statistik setiap Variabel

Variabel	Minimum	Maksimum	Rata-rata	Simpangan Baku	Varsians
Y	0	6	1.80	1.366	1.865
X1	3	18	12.38	6.813	46.411
X2	4	15	7.43	4.227	17.871
X3	2	29	11.85	10.949	119.890
X4	1	22	4.26	4.979	24.795
X5	4	17	5.69	3.280	10.758
X6	3	37	16.70	14.949	223.483
X7	2	18	6.34	6.226	38.768
X8	4	13	5.60	2.657	7.058
Variabel	Minimum	Maksimum	Rata-rata	Simpangan Baku	Varsians
X9	3	29	16.52	11.623	135.084

Tabel 2 menunjukkan statistik deskriptif untuk seluruh variabel dalam penelitian. Variabel Y sebagai variabel respon memiliki nilai rata-rata sebesar 1,80, dengan nilai minimum 0 dan maksimum 6. Sementara itu, variabel bebas seperti X1 hingga X9 menunjukkan variansi yang berbeda-beda, dengan Variabel X₃ dan X₆ memiliki simpangan baku yang relatif tinggi, yang menunjukkan bahwa nilai-nilainya menyebar secara luas. Variabilitas yang besar ini dapat menjadi indikasi bahwa kedua variabel tersebut memiliki potensi menjelaskan variasi pada variabel respon Y secara lebih efektif, sehingga berpeluang menunjukkan pengaruh yang signifikan dalam model regresi COM-Poisson. Tabel 2 juga menunjukkan bahwa Jumlah gol tim tuan rumah (Y) memiliki rata-rata sebesar 1,80, dengan nilai variansi yang lebih besar, yaitu 1,865. Kondisi ini mengindikasikan adanya *overdispersi*, karena variansi pada variabel respon tidak sama dengan nilai rata-ratanya.



Gambar 1. Diagram Batang Jumlah Gol Tim Tuan Rumah

Dalam histogram pada gambar tersebut, tim tuan rumah paling sering mencetak satu atau dua gol perpertandingan. Semakin besar jumlah gol, semakin jarang terjadi. Karena lebih banyak nilai kecil (0,1,2 gol) dan hanya sedikit nilai besar (5 atau 6 gol), maka bentuknya disebut *positively skewed* atau miring ke kanan. Pola ini sesuai dengan distribusi Poisson, yang umum dalam data jumlah kejadian seperti gol dalam pertandingan.

1. Uji Kecocokan Distribusi Poisson

Dengan menggunakan uji Chi-Kuadrat, diperoleh nilai χ^2_{hitung} sebesar 2,272 dan χ^2_{tabel} sebesar 9,488. Dengan menggunakan taraf signifikansi $\alpha = 5\%$ maka terima H_0 karena $\chi^2_{hitung} < \chi^2_{tabel}$, sehingga dapat disimpulkan bahwa variabel Jumlah gol tim tuan rumah (Y) mengikuti distribusi Poisson.

2. Uji Multikolinieritas

Tabel 3. Tabel Multikolinieritas

Variabel Bebas	Nilai VIF
X1	9,524
X2	1,671
X3	3,622
X4	2,388
X5	2,270
X6	4,992
X7	1,661
X8	2,602
X9	9,756

Sesuai dengan Tabel 4.2, semua nilai VIF untuk variabel bebas berada di bawah batas maksimum 10. Oleh karena itu, dapat disimpulkan bahwa tidak terdapat indikasi multikolinearitas yang kuat di antara variabel bebas dalam model. Hasil ini menunjukkan bahwa seluruh variabel bebas dapat digunakan secara bersamaan dalam model regresi Conway-Maxwell-Poisson tanpa adanya kekhawatiran terhadap multikolinearitas yang berlebihan.

3. Pemeriksaan *Overdispersi*

Untuk memeriksa keberadaan *overdispersi*, dapat dideteksi dengan pemeriksaan rasio deviansi terhadap derajat bebasnya [5], jika $\frac{D}{db} > 1$, artinya terdapat *overdispersi*.

$$\frac{D}{db} = \frac{409,6852}{380 - 10} = \frac{409,6852}{370} = 1,107$$

Karena didapatkan nilai rasio deviansi terhadap deajat bebas $\frac{D}{db} = 1,107 > 1$, maka berarti terjadi *overdispersi*.

4. Pemodelan Regresi COM-Poisson

Karena saat diasumsika bahwa data mengalami masalah *overdispersi*, yaitu kondisi di mana varians data lebih besar dari rata-ratanya. Oleh karena itu, dilakukan pemodelan alternatif yang dapat mengakomodasi *overdispersi*, yaitu dengan menggunakan regresi Conway-Maxwell-Poisson (COM-Poisson). Berdasarkan hasil estimasi, diperoleh model regresi COM-Poisson sebagai berikut:

Tabel 4. Regresi COM-Poisson

Variabel	Koefisien	Std.Error	z-value	p-value
Intercept	1.2307	0.3945	3.1198	0.00181
X1	-0.0164	0.0178	-0.9228	0.3561
X2	-0.0128	0.0127	-1.0072	0.3138
X3	-0.0157	0.0077	-2.0268	0.04268
X4	0.0284	0.0109	2.6091	0.009079
X5	-0.0092	0.0190	-0.4848	0.6278
X6	-0.0061	0.0062	-0.9774	0.3284
X7	0.0271	0.0078	3.4908	0.0004816
X8	0.0032	0.0243	0.1305	0.8962
X9	-0.0146	0.0108	-1.3418	0.1797

$$\hat{\mu}_i = \exp (1,2307 - 0.0164X_1 - 0.0128X_2 - 0.0157X_3 + 0.0284X_4 - 0.0092X_5 - 0.0061X_6 + 0.0271X_7 + 0.0032X_8 - 0.0146X_9)$$

Dengan nilai dispersi (v) = 1,1229 yang mengindikasikan bahwa model COM-Poisson mendeteksi adanya underdispersi ringan dalam data, karena nilai v lebih besar dari 1. Ini berarti bahwa variansi (penyebaran) Jumlah gol tim tuan rumah (Y) cenderung lebih kecil dari pada rata-ratanya. Artinya, dibandingkan dengan model Poisson standar, data dalam kasus ini tidak terlalu tersebar dan memiliki kecenderungan variansi yang lebih terkendali.

5. Pemodelan Regresi Poisson

Tabel 5. Tabel Regresi Poisson

Variabel	Koefisien	Std.Error	z-value	p-value
Intercept	0,9592	0.355459	2.698	0.006969
X1	0,0097	0.017138	-0.563	0.573299
X2	0,0108	0.012173	-0.886	0.375471
X3	0,0148	0.007396	-1.999	0.045569
X4	0,0243	0.010301	2.359	0.018329
X5	0,0089	0.018238	-0.487	0.626063
X6	0,0052	0.005990	-0.874	0.381886
X7	0,0241	0.007179	3.362	0.000773
X8	0,0041	0.023331	0.177	0.859660
X9	0,0105	0.010335	-1.015	0.310249

$$\hat{\mu}_i = \exp (0,9592 - 0,0097X_1 - 0,0108X_2 - 0,0148X_3 + 0,0243X_4 - 0,0089X_5 - 0,0052X_6 + 0,0241X_7 + 0,0041X_8 - 0,0105X_9)$$

Koefisien negatif pada sebagian besar variabel yaitu variabel X₁, X₂, X₃, X₅, X₆, dan X₉ menunjukkan bahwa peningkatan nilai pada variabel-variabel tersebut cenderung menurunkan rata-rata Jumlah Gol Tim Tuan Rumah (Y). Sebaliknya, variabel X₄, X₇, dan X₈ memiliki koefisien positif, yang mengindikasikan bahwa semakin besar nilai dari variabel-variabel tersebut, maka rata-rata Jumlah Gol

Tim Tuan Rumah (Y) cenderung meningkat.

6. Pengujian Signifikansi Parameter Dispersi

Ingin menguji apakah parameter dispersi (v) signifikan dengan menggunakan uji rasio kemungkinan, dengan hipotesis sebagai berikut:

$$2(l_{COM-Poisson} - l_{Poisson})$$

Dengan menggunakan nilai kritis: $\chi^2_{\alpha=0,05;db=1} = 3,841$

$$2(-606,674 - (-631,226)) = 49,104$$

Karena $49,104 > 3,841$ maka tolak H_0 yang artinya equidisersi tidak terpenuhi dan kesimpulan yang diambil adalah model COM-Poisson.

7. Pengujian Parameter Model secara Simultan

Didapatkan nilai statistik uji $G = 49,1041$ lebih besar dari nilai $\chi^2_{(0,95,9)} = 16,919$ dan nilai p -value sebesar $1,586867 \times 10^{-7}$ lebih kecil dari taraf 0,05, maka H_0 ditolak. Artinya, terdapat cukup bukti statistik untuk menyimpulkan bahwa variabel-variabel X1 sampai X9 secara bersama-sama berpengaruh signifikan terhadap jumlah gol tim tuan rumah pada pertandingan sepak bola *English Premier League* (EPL).

8. Pengujian Parameter Model secara Parsial

Tabel 6. Pengujian Parameter Model secara Parsial

Variabel	Koefisien	p -value	Keterangan
X1	-0.0164	0.3561	Tidak Signifikan
X2	-0.0128	0.3138	Tidak Signifikan
X3	-0.0157	0.04268	Signifikan
X4	0.0284	0.009079	Signifikan
X5	-0.0092	0.6278	Tidak Signifikan
X6	-0.0061	0.3284	Tidak Signifikan
X7	0.0271	0.0004816	Signifikan
X8	0.0032	0.8962	Tidak Signifikan
X9	-0.0146	0.1797	Tidak Signifikan

Model menunjukkan bahwa X3, X4, dan X7 berpengaruh signifikan terhadap jumlah gol tim tuan rumah (Y) di pertandingan *English Premier League* (EPL) pada taraf signifikansi 5%, berikut interpretasinya:

Nilai koefisien regresi untuk variabel X3 adalah $\beta_3 = -0,0157$, sehingga $\exp(-0,0157) = 0,985$. Ini mengindikasikan bahwa setiap peningkatan 1 satuan pada X3 berpotensi meningkatkan Y sebesar $(0,985-1)100\% = -1,5\%$ atau penurunan sebesar 1,5%. Ini signifikan, artinya X3 yang tinggi cenderung menurunkan peluang tim tuan rumah mencetak gol. Koefisien regresi untuk variabel X4 adalah $\beta_4 = 0,0284$, sehingga $\exp(0,0284) = 1,029$. Ini mengindikasikan bahwa setiap peningkatan 1 satuan pada X4 berpotensi meningkatkan jumlah gol tim tuan rumah sebesar $(1,029-1)100\% = 2,9\%$. Artinya, semakin tinggi X4, maka Y cenderung meningkat. Nilai X4 yang tinggi dapat menunjukkan persepsi bahwa tim tuan rumah punya peluang besar mencetak gol. Koefisien regresi untuk variabel X7 adalah $\beta_7 = 0,0271$, sehingga $\exp(0,0271) = 1,0275$. Ini mengindikasikan bahwa setiap peningkatan 1 satuan pada X7 berpotensi meningkatkan jumlah gol tim tuan rumah sebesar $(1,0275-1)100\% = 2,7\%$. Artinya, semakin tinggi X7, maka Y cenderung meningkat. Nilai X7 yang tinggi dapat menunjukkan persepsi bahwa tim tuan rumah punya peluang besar mencetak gol.

D. Kesimpulan

Berdasarkan hasil analisis data pertandingan *English Premier League* (EPL) musim 2023–2024, kesimpulan utama yaitu model regresi Conway-Maxwell-Poisson (COM-Poisson) mampu menangani permasalahan tidak terpenuhinya asumsi equidisersi. Model ini menunjukkan bahwa tiga variabel *Pinnacle Away Win Odds* (X3), *Maximum Home Win Odds* (X4), dan *Market Average Home Win Odds* (X7) berpengaruh signifikan terhadap jumlah gol tim tuan rumah (Y).

Ucapan Terimakasih

Segala puji dan syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat dan karunia-Nya, sehingga penelitian ini dapat diselesaikan dengan baik. Ucapan terima kasih juga penulis sampaikan kepada Bapak Abdul Kudus, S.Si., M.Si., Ph.D., selaku dosen pembimbing, atas segala arahan, saran, dan bimbingan yang telah diberikan selama proses penelitian ini berlangsung.

Daftar Pustaka

- Herliani, F. D., & Kudus, A. (2023). Penanganan Data Missing dengan Algoritma Multivariate Imputation By Chained Equations (MICE). *DataMath: Journal of Statistics and Mathematics*, 1(1), 35–42. <https://doi.org/10.29313/datamath.v1i1.25>
- Luhung Mustika Budiharti, & Sunendiari, S. (2021). Pemodelan dan Pemetaan Jumlah Penderita Kusta di Jawa Barat dengan Regresi Binomial Negatif dan Flexibly Shaped Spatial Scan Statistic. *Jurnal Riset Statistika*, 1(2), 99–106. <https://doi.org/10.29313/jrs.v1i2.409>
- Alrweili, H. (2024). Kibrida-Lukman Hybrid Estimator for the Conway-Maxwell- Poisson Regression Model. *Electronic Journal of Applied Statistical Analysis*, 17(2), 436-471.
- De Jong, P., & Heller, G. Z. (2008). *Generalized Linear Models for Insurance Data*. Cambridge: Cambridge University Press.
- Francis, R. A., Geedipally, S. R., Guikema, S. D., Dhavala, S. S., Lord, D., and LaRocca, S. (2012). Characterizing the performance of the conway-maxwell poisson generalized linear model. *Risk Analysis: An International Journal*, 32(1):167–183.
- Koran Jakarta. (2022). Mengenal sistem sepak bola Liga Inggris: Sejarah, sistem, dan level dalam kompetensinya. Diakses pada 2 Januari 2025, dari <https://koran-jakarta.com/mengenal-sistem-sepak-bola-liga-inggris-sejarah-sistem-dan-level-dalam-kompetensinya>
- McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models*. Chapman and Hall. London, UK.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis* (5th ed.). Hoboken, NJ: John Wiley & Sons.
- Sellers, K. F., & Shmueli, G. (2010). A Flexible Regression Model for Count Data. *The Annals of Applied Statistics*, 4(2), 943–961.
- Sellers, K. F., Borle, S., & Shmueli, G. (2012). The COM-Poisson model for count data: A survey of methods and applications. *Applied Stochastic Models in Business and Industry*, 28(2), 104–116. <https://doi.org/10.1002/asmb.918>.
- Sudjana. (2014). *Metoda Statistika* (Edisi Ketujuh). Bandung: Tarsito.
- Tempo.co. (2023). Kilas sejarah awal kompetisi Premier League Inggris. Diakses pada 2 Januari 2025, dari <https://www.tempo.co/sepakbola/kilas-sejarah-awal-kompetisi-premier-league-inggris-216763>.
- Ratih Nurfitri, & Teti Sofia Yanti. (2023). Pemodelan Umur Harapan Hidup di Jabar Tahun 2021 Menggunakan Spatial Durbin Model. *Jurnal Riset Statistika*, 137–146. <https://doi.org/10.29313/jrs.v3i2.3023>