

Penerapan Model Regresi *Zero-One Inflated Beta* (BEINF) pada Proporsi Pemain Sepak Bola yang Dimainkan dalam Satu Musim

Rida Faridah*, Nusar Hajarisman

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

*ridafaridah109@gmail.com, nusarhajarisman@unisba.ac.id

Abstract. near regression is usually used to model the relationship between a dependent variable and one or more independent variables where the response is continuous and normally distributed. However, if the assumption of normality of the response is not met, and the estimation for the linear regression model is not accurate, then zero-one inflated beta (BEINF) regression is an alternative method to model the relationship between the response variable and predictors where the response is a mixture of beta distribution and bernouli distribution. The results of this research are that there are several factors that affect the response variable, namely the BMI variable (X1), Forty Yard Run (X2), Vertical Jump (X3), Wide Jump (X4) and for predicting the chances of soccer players it can be said that, in the model $0 < Y < 1$ the chance that a soccer player will be chosen to play in the match is equal to $\hat{\pi}=0.4514$ times more than before assuming the value of the other predictor variables is fixed. In the $Y = 1$ model, the chance that a soccer player will be chosen to play in the match increases by $f(y=1)= 0.2086$ times compared to before assuming the value of the other predictor variables remains constant. In the $Y = 0$ model, the chance that a soccer player will not be chosen to play in the match increases by $f(y=0)=0,7576$ times compared to before, assuming the values of the other predictor variables remain constant.

Keywords: *Beta Distribution, Maximum Likelihood Estimation (MLE), Zero-One Beta.*

Abstrak. Regresi linier biasanya digunakan untuk memodelkan hubungan antara variabel dependen dengan satu atau lebih variabel independen yang dimana respon bersifat kontinu dan berdistribusi normal. Namun apabila asumsi normalitas respon tidak terpenuhi, dan estimasi untuk model regresi linier tidak akurat, maka regresi zero-one inflated beta (BEINF) adalah metode alternatif untuk memodelkan hubungan antara variabel respon dan prediktor dimana respon berdistribusi campuran antara distribusi beta dan distribusi bernouli. Hasil dari penelitian ini adalah terdapat beberapa faktor yang mempengaruhi variabel respon yaitu variabel BMI (X1), Lari Empat Puluh Yard (X2), Lompat Vertikal (X3), Lompat Lebar (X4) dan untuk prediksi peluang pemain sepak bola dapat dikatakan bahwa, pada model $0 < Y < 1$ peluang pemain sepak bola akan dipilih untuk bermain di pertandingan adalah sebesar $\hat{\pi}=0.4514$ kali lipat dibanding sebelumnya dengan asumsi nilai variabel prediktor lain tetap. Pada model $Y = 1$ peluang pemain sepak bola akan dipilih untuk bermain di pertandingan meningkat sebesar $f(y=1)= 0.2086$ kali lipat dibanding sebelumnya dengan asumsi nilai variabel prediktor lain tetap. Pada model $Y = 0$ peluang pemain sepak bola tidak akan dipilih untuk bermain di pertandingan meningkat sebesar $f(y=0)=0,7576$ kali lipat dibanding sebelumnya dengan asumsi nilai variabel prediktor lain tetap.

Kata Kunci: *Distribusi Beta, Maximum Likelihood Estimation (MLE), Zero-One Inflated Beta.*

A. Pendahuluan

Sepak bola merupakan olahraga yang populer di kalangan masyarakat dari segala usia, termasuk pelajar, di mana olahraga ini dimainkan oleh dua tim. Setiap tim sepak bola memiliki pelatih yang bertanggung jawab untuk mengembangkan strategi, taktik permainan dan memberikan para pemain keterampilan teknis dan fisik yang mereka butuhkan untuk bermain sepak bola. Pemantauan secara berkala juga dilakukan untuk mengukur kualitas kebugaran tim sepak bola. Tes fisik seperti BMI (*Body Mass Index*), lari sejauh 40 yard (sekitar 36,6 meter), lompat vertikal, lompat lebar, dan *bench press* (angkat beban) digunakan untuk mengevaluasi kemajuan pemain atau tim secara keseluruhan dan data-data ini digunakan untuk memprediksi persentase pemain yang akan dimainkan pada musim selanjutnya. Dengan menggunakan proporsi pemain sebagai variabel respon, dan beberapa kategori pemain yang menjalani pelatihan tetapi tidak pernah bermain di pertandingan besar (0), adapun beberapa pemain yang menjalani pelatihan dan bermain dipertandingan selama satu musim (1) dan beberapa pemain yang menjalani pelatihan dan bermain di pertandingan besar namun tidak sepenuhnya bermain dalam satu musim (0-1).

Regresi *zero-one inflated beta* (BEINF) adalah suatu jenis model regresi yang digunakan untuk mengatasi masalah data yang terdiri dari variabel respon nan berdistribusi antara 0 dan 1 dengan tingkat inflasi nol. Pemodelan regresi BEINF menggunakan campuran distribusi beta dan distribusi bernoulli (Ospina dan Ferari, 2007). Sebagai model regresi yang lebih kompleks, regresi BEINF sangat memungkinkan untuk memodelkan hubungan antara variabel prediktor dengan kemungkinan terjadinya nol, serta hubungan antara variabel prediktor dengan distribusi beta pada nilai positif antara 0 dan 1. Dengan begitu, model regresi BEINF dapat memberikan estimasi parameter yang terkait dengan perubahan rata-rata dan deviasi standar dari variabel respons, dan menarik kesimpulan darinya.

Berdasarkan penjelasan pada latar belakang, untuk rumusan masalah dalam penelitian ini adalah “Faktor-faktor apa saja yang mempengaruhi peluang pemain sepak bola untuk bisa dimainkan dalam suatu permainan dan Bagaimana prediksi peluang pemain sepak bola yang akan bermain selama musim selanjutnya”. Adapun, tujuan dari penelitian ini adalah untuk mengetahui Faktor-faktor apa saja yang mempengaruhi peluang pemain sepak bola untuk bisa dimainkan dalam suatu permainan dan untuk memprediksi peluang pemain sepak bola yang akan bermain pada musim selanjutnya.

B. Metodologi Penelitian

Dalam Penelitian ini penulis menggunakan jenis data sekunder. Data yang digunakan oleh penulis adalah data Tim Sepak Bola Perguruan Tinggi yang diperoleh dari buku *Advanced Regression Models With SAS and R* oleh Olga Korosteleva (2019). Yang berisi satu variabel respon dan empat variabel prediktor.

Metode Analisis Data

Distribusi Bernoulli

Merupakan sebuah distribusi probabilitas diskrit yang digunakan untuk menghitung peluang suatu peristiwa biner. Distribusi ini menggambarkan hasil dari eksperimen yang hanya memiliki dua kemungkinan hasil, yaitu sukses atau kegagalan. Distribusi bernoulli memiliki parameter tunggal, yaitu probabilitas sukses p . Probabilitas kegagalan, yang dinyatakan sebagai q , adalah $1 - p$. Oleh karena itu, distribusi bernoulli hanya bergantung pada satu parameter. Berikut fungsi peluang distribusi bernoulli:

$$p(x) = P(X = x) = p^x(1 - p)^{1-x}; x = 0, 1 \quad (1)$$

Distribusi Beta

Distribusi beta didefinisikan pada rentang nilai dari 0 hingga 1, dan digunakan untuk memodelkan variabel acak yang memiliki rentang nilai terbatas seperti data proporsi atau persentase. Distribusi beta dengan parameter μ dan ϕ ($0 < \mu < 1$ dan $\phi > 0$), dilambangkan dengan $B(\mu, \phi)$, memiliki fungsi densitas menurut Ferrari dan Cribari-Neto. 2004 sebagai berikut:

$$f(y, \mu, \varphi) = \frac{\Gamma(\varphi)}{\Gamma(\mu\varphi)\Gamma((1-\mu)\varphi)} y^{\mu\varphi-1}(1-y)^{(1-\mu)\varphi-1}, \quad 0 < y < 1 \quad (2)$$

Dimana $\Gamma(\cdot)$ adalah fungsi gamma. Jika $y \sim B(\mu, \varphi)$, lalu $E(y) = \mu$ dan $\text{var}(y) = V(\mu)/(\varphi+1)$, dimana $V(\mu) = \mu(1-\mu)$ menunjukkan fungsi varians. Parameter φ memainkan peran parameter presisi dalam arti bahwa, untuk diperbaiki μ , semakin besar nilainya φ , semakin kecil varian y nya. Nilai parameter yang berbeda menghasilkan bentuk densitas beta yang berbeda.

Excess Zero

Merupakan salah satu permasalahan pada regresi *zero-one inflated beta* yaitu terdapat nilai nol yang berlebih. Pada Variabel respon mungkin akan ditemukan data kosong atau bernilai nol. Akan tetapi dalam banyak kasus, nilai nol memiliki arti penting pada penelitian yang bersangkutan. Jika nilai nol memiliki arti penting dalam data diskrit maka data tersebut tetap dimasukkan dalam analisis.

Regresi Zero One Infalted Beta

Model regresi *Zero-and-One-Inflated beta* atau bisa disebut dengan BEINF adalah distribusi probabilitas yang umum digunakan dalam analisis data biner atau kategori dimana beberapa nilai mengandung 0, 1. Untuk situasi ini Ferrari (2004) mengusulkan campuran antara distribusi beta dan distribusi bernoulli. Dengan model sebagai berikut:

$$f_{\text{BEINF}}(y; \alpha, \gamma, \mu, \varphi) = \begin{cases} \alpha(1-\gamma), & \text{jika } y = 0, \\ \alpha\gamma, & \text{jika } y = 1, \\ (1-\alpha)f(y; \mu, \varphi), & \text{jika } y \in (0, 1), \end{cases} \quad (3)$$

Dari model distribusi BEINF tersebut, maka diperoleh nilai harapan dan varians sebagai berikut:

$$\begin{aligned} E(y^r) &= \alpha\gamma + (1-\alpha)\mu_r, \quad r = 1, 2, \dots, \\ \text{var}(y) &= \alpha V_1 + (1-\alpha)V_2 + \alpha(1-\alpha)(\gamma - \mu)^2 \end{aligned} \quad (4)$$

Dengan estimasi parameter dalam model dapat ditulis sebagai berikut:

$$\hat{\pi} = \frac{\exp\{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_m x_m\}}{1 + \exp\{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_m x_m\}}, \quad (5)$$

$$\alpha(\widehat{1-\gamma}) = \frac{\hat{t}}{1 + \hat{v} + \hat{t}} \quad (6)$$

$$\widehat{\alpha\gamma} = \frac{\hat{v}}{1 + \hat{v} + \hat{t}} \quad (7)$$

Dimana

$$\begin{aligned} \hat{v} &= \exp\{\hat{\gamma}_0 + \hat{\gamma}_1 x_m + \dots + \hat{\gamma}_{l-m} x_1\} \\ \hat{t} &= \exp\{\hat{\zeta}_0 + \hat{\zeta}_1 x_1 + \dots + \hat{\zeta}_{k-l} x_k\}, \end{aligned}$$

Dengan

$\hat{\pi}$ = Untuk model jika $y \in (0, 1)$

$\alpha(\widehat{1-\gamma})$ = Untuk model jika $y = 0$

$\widehat{\alpha\gamma}$ = Untuk model jika $y = 1$

Maximum Likelihood Estimation

Misalkan $(y_1, \dots, y_n)$ menjadi sampel acak dari sebuah distribusi BEINF. Karena distribusi *Zero-One Inflated Beta* yang diberikan pada persamaan (2) merupakan distribusi keluarga eksponensial dengan empat parameter dengan pangkat penuh maka dapat diimplementasikan sebagai berikut:

$$\sum_{t=1}^n T(y_t) = (T_1, T_2, T_3, T_4) \quad (8)$$

Dengan

$$T_1 = \sum_{t=1}^n I_{\{0,1\}}(y_t), T_2 = \sum_{t=1}^n y_t I_{\{0,1\}}(y_t), T_3 = \sum_{t:y_t \in (0,1)} \log(y_t) \text{ dan } T_4 = \sum_{t:y_t \in (0,1)} \log(1 - y_t)$$

Yang dimana $\sum_{t=1}^n T(y_t)$ Merupakan statistik yang cukup lengkap, untuk melengkapi fungsi Likelihood dengan $\theta = (\alpha, \gamma, \mu, \varphi)$ yang diberikan pada sampel $(y_1, \dots, y_n)$ sebagai berikut:

$$L(\theta) = \prod_{t=1}^n beinf(y_t; \alpha, \gamma, \mu, \varphi) = L_1(\alpha) L_2(\gamma) L_3(\mu, \varphi), \quad (9)$$

Dengan

$$L_1(\alpha) = \prod_{t=1}^n \alpha^{I_{\{0,1\}}(y_t)} (1 - \alpha)^{1 - I_{\{0,1\}}(y_t)} = \alpha^{T_1} (1 - \alpha)^{(n - T_1)} \quad (10)$$

$$L_2(\gamma) = \prod_{t:y_t \in \{0,1\}} \gamma^{y_t} (1 - \gamma)^{(1 - y_t)} = \gamma^{T_2} (1 - \gamma)^{(T_1 - T_2)} \quad (11)$$

$$L_3(\mu, \varphi) = \prod_{t:y_t \in \{0,1\}} f(y_t; \mu, \varphi). \quad (12)$$

Fungsi *Likelihood* $L(\theta)$ memfaktorkan dalam tiga suku, yaitu L_1, L_2 dan L_3 , untuk L_1 hanya bergantung pada α , L_2 hanya bergantung pada γ dan L_3 hanya bergantung pada μ dan φ . Oleh karena itu α, γ , dan (μ, φ) adalah parameter yang dapat dipisahkan dan inferensi MLE untuk α, γ , dan (μ, φ) dapat dilakukan secara independen. Fungsi log *Likelihood* dapat ditulis sebagai berikut:

$$\ell(\theta) = \ell_1(\alpha) + \ell_2(\gamma) + \ell_3(\mu, \varphi) \quad (13)$$

Dimana

$$\ell_1(\alpha) = \log \alpha^{T_1} + \log (1 - \alpha)^{(n - T_1)} \quad (14)$$

$$\ell_2(\gamma) = T_2 \log \gamma + (T_1 - T_2) \log (1 - \gamma), \quad (15)$$

$$\begin{aligned} \ell_3(\mu, \varphi) = & (n - T_1) \log \left\{ \frac{\Gamma(\varphi)}{\Gamma(\mu\varphi)\Gamma((1 - \mu)\varphi)} \right\} + T_3(\mu\varphi - 1) \\ & + T_4((1 - \mu)\varphi - 1) \end{aligned} \quad (16)$$

Dengan membedakan $\ell_1(\alpha)$ terhadap α , $\ell_2(\gamma)$ terhadap γ dan $\ell_3(\mu, \varphi)$ terhadap μ dan φ memperoleh vektor skor turunan pertama dari fungsi log-likelihood. Sebagai berikut:

$$U(\theta) = (U_\alpha(\alpha), U_\gamma(\gamma), U_\mu(\mu, \varphi), U_\varphi(\mu, \varphi)), \quad (17)$$

Dimana

$$\begin{aligned} U_\alpha(\alpha) &= T_1/\alpha - (n - T_1)/(1 - \alpha) \\ U_\gamma(\gamma) &= T_2/\gamma - (T_1 - T_2)/(1 - \gamma) \\ U_\mu(\mu, \varphi) &= \varphi \{ (n - T_1) [\psi((1 - \mu)\varphi) - \Psi(\mu\varphi)] + T_3 - T_4 \} \end{aligned} \quad (18)$$

$$U_\varphi(\mu, \varphi) = (n - T_1) [\psi(\varphi) - \mu\psi(\mu\varphi) - (1 - \mu)\psi((1 - \mu)\varphi)] + \mu T_3 - (1 - \mu) T_4$$

Dalam praktiknya, MLE dapat diperoleh melalui maksimalisasi numerik dari log-*Likelihood* $\ell_3(\mu, \varphi)$ menggunakan algoritma optimasi nonlinier. Estimator bentuk tertutup untuk μ dan φ , $\hat{\mu}$ dan $\hat{\varphi}$ dapat diperoleh dengan menggunakan momen bersyarat dari y , mengingat bahwa $y \in (0, 1)$, yang tidak bergantung pada α tidak juga pada γ . Demikian juga, estimator bentuk tertutup dari $E(y)$ dan $\text{var}(y)$ dapat diperoleh dengan menggantikan α, γ, μ dan φ menjadi $\hat{\alpha}, \hat{\gamma}, \hat{\mu}$ dan $\hat{\varphi}$ dalam persamaan (2). Matriks informasi fisher untuk parameter distribusi BEINF dapat ditulis sebagai berikut:

$$K(\theta) = \begin{pmatrix} k_{\alpha\alpha} & 0 & 0 & 0 \\ 0 & k_{\gamma\gamma} & 0 & 0 \\ 0 & 0 & k_{\mu\mu} & k_{\mu\varphi} \\ 0 & 0 & k_{\varphi\mu} & k_{\varphi\varphi} \end{pmatrix} \quad (19)$$

Dimana

$$\begin{aligned} k_{\alpha\alpha} &= 1/\{\alpha(1-\alpha)\}, \\ k_{\gamma\gamma} &= \alpha/\{\gamma(1-\gamma)\}, \\ k_{\mu\mu} &= (1-\alpha)\varphi^2\{\psi(\mu\varphi) + \psi((1-\mu)\varphi)\}, \\ k_{\mu\varphi} &= k_{\varphi\mu} = (1-\alpha)\varphi\{\psi(\mu\varphi)\mu - \psi((1-\mu)\varphi)(1-\mu)\} \\ k_{\varphi\varphi} &= (1-\alpha)\{\mu^2\psi(\mu\varphi) + (1-\mu)^2\psi((1-\mu)\varphi) - \psi(\varphi)\}. \end{aligned} \quad (20)$$

Disini, α, γ dan (μ, φ) adalah parameter ortogonal dan karenanya, masing-masing komponen vektor skor tidak berkorelasi. Karena distribusi *zero-one inflated beta* pada persamaan (2) milik keluarga ekponensial dengan peringkat penuh, maka $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N_4(0, K(\theta)^{-1})$, dengan $k(\theta)$ diberikan dalam persamaan (24) dan $\hat{\alpha}, \hat{\gamma}$, dan $(\hat{\mu}, \hat{\varphi})$ merupakan hubungan antara dua atau lebih variabel yang menjadi semakin independen satu sama lain ketika ukuran sampel atau jumlah pengamatan mereka mendekati tak hingga (asymptotically independent) (Rosa & Karyana, 2023).

Tahapan Selanjutnya masuk pada tahap memaksimalkan α, γ, μ dengan menghitung $\alpha^{(k+1)}, \gamma^{(m+1)}$ dan $\mu^{(m+1)}$ menggunakan *fisher's scoring*. Iterasi dilakukan sampai diperoleh penduga parameter yang konvergen, dikatakan konvergen ketika perbedaan relatif antara estimasi parameter pada dua iterasi berurutan dianggap kecil, yaitu, kurang dari 1% atau 0.1%.

Uji Signifikansi Parameter Secara Simultan

Untuk mengetahui signifikansi parameter regresi secara simultan dapat menggunakan Statistik Uji G menurut Agresti, (2007) dengan hipotesis sebagai berikut:

- Hipotesis
 $H_0 : \beta_0 = \beta_3 = \beta_5 = \gamma_0 = \gamma_4 = \alpha_0 = \alpha_1 = \alpha_2 = 0$
 $H_1 : \text{setidaknya terdapat satu } \beta_j \neq 0, \gamma_j \neq 0 \text{ dan } \alpha_j \neq 0$
 Dengan $j = 1, 2, 3, 4, 5$
 Statistik Uji G

$$G_{hitung} = -2 \ln[L_0 - L_1] \sim \chi^2_{p(\alpha)} \quad \dots(21)$$

Dimana:

L_0 = *Likelihood* model tanpa semua peubah penjelas

L_1 = *Likelihood* model dengan semua peubah penjelas

- Kriteria Uji: Jika Statistika uji $G > \chi^2_{p(\alpha)}$, maka hipotesis nol ditolak. Sehingga dapat disimpulkan bahwa parameter regresi signifikan.

Uji Signifikansi Parameter Secara Parsial

Untuk mengetahui signifikansi parameter secara parsial menggunakan Uji t dengan hipotesis sebagai berikut:

3. Parameter β_j

Hipotesis

$H_0: \beta_j = 0$; Variabel prediktor tidak memiliki pengaruh yang signifikan terhadap variabel respon.

$H_1: \beta_j \neq 0$; Variabel prediktor memiliki pengaruh yang signifikan terhadap variabel respon.

Dengan $j = 1, 2, 3, \dots, p$

Statistik Uji

$$t_{hitung} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \quad (22)$$

dengan $se(\hat{\beta}_j) = \sqrt{var(\hat{\beta}_j)}$

Kriteria Uji: Jika $t > t_\alpha$ atau p-value $< \alpha$ dengan α adalah taraf nyata yang diinginkan, maka hipotesis nol ditolak, dan dapat disimpulkan bahwa parameter signifikan.

4. Parameter γ_j

Hipotesis

$H_0: \gamma_j = 0$; Variabel prediktor tidak memiliki pengaruh yang signifikan terhadap variabel respon.

$H_1: \gamma_j \neq 0$; Variabel prediktor memiliki pengaruh yang signifikan terhadap variabel respon.

Dengan $j = 1, 2, 3, \dots, p$

Statistik Uji

$$t_{hitung} = \frac{\hat{\gamma}_j}{se(\hat{\gamma}_j)} \quad (23)$$

dengan $se(\hat{\gamma}_j) = \sqrt{var(\hat{\gamma}_j)}$

Kriteria Uji: Jika $t > t_\alpha$ atau p-value $< \alpha$ dengan α adalah taraf nyata yang diinginkan, maka hipotesis nol ditolak, dan dapat disimpulkan bahwa parameter signifikan.

5. Parameter α_j

Hipotesis

$H_0: \alpha_j = 0$; Variabel prediktor tidak memiliki pengaruh yang signifikan terhadap variabel respon.

$H_1: \alpha_j \neq 0$; Variabel prediktor memiliki pengaruh yang signifikan terhadap variabel respon.

Dengan $j = 1, 2, 3, \dots, p$

Statistik Uji

$$t_{hitung} = \frac{\alpha_j}{se(\alpha_j)} \quad (24)$$

dengan $se(\alpha_j) = \sqrt{var(\alpha_j)}$

Kriteria Uji: Jika $t > t_\alpha$ atau p-value $< \alpha$ dengan α adalah taraf nyata yang diinginkan, maka hipotesis nol ditolak, dan dapat disimpulkan bahwa parameter signifikan.

C. Hasil Penelitian dan Pembahasan

Deskripsi Data

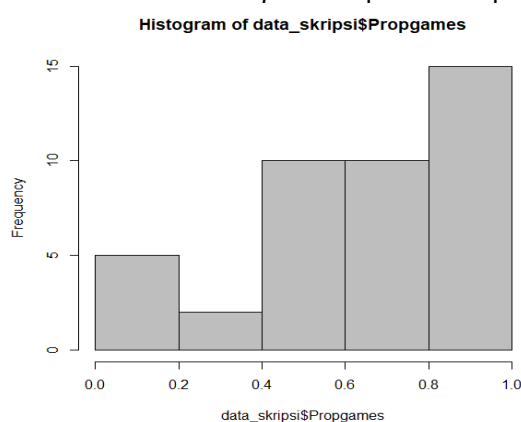
Data yang digunakan pada penelitian ini berjumlah 42 data pemain sepak bola yang dimana

terdapat data bernilai nol (pemain yang menjalani pelatihan tetapi tidak pernah bermain di pertandingan besar) sebanyak 5 data, kemudian terdapat data bernilai satu (pemain yang menjalani pelatihan dan bermain di pertandingan besar) sebanyak 8 data dan terdapat data antara 0-1 (pemain yang menjalani pelatihan dan bermain di pertandingan besar namun tidak bermain di semua) sebanyak 29. Untuk melihat kecenderungan *zero-one inflated beta* pada data dapat dilakukan analisis deskripsi sebagai berikut:

Tabel 1. Hasil Analisis Deskripsi Data tim Sepak Bola

Variabel	Minimum	Maksimum	Rata-Rata	Varsians
Y	0	1	27.28	0.0980
X1	27.6	32.9	29.89047619	1.7516
X2	4.37	4.8	4.557380952	0.0111
X3	30.5	41	34.52380952	5.0847
X4	105	130	116.9285714	26.5069
X5	14	29	19.9047619	14.4297

Berdasarkan Tabel 1 dapat dilihat bahwa rata-rata pada variabel Y menunjukkan bahwa rata-rata nilai proporsi pemain sepak bola sebesar 27.28 dalam satu musim, dengan nilai terendah 0 yang menunjukkan bahwa pemain sepak bola yang tidak pernah bermain selama satu musim dengan nilai terbesar 1 yang menunjukkan bahwa pemain sepak bola bermain selama satu musim. Pada variabel X1 menunjukkan bahwa rata-rata *Body Mass Index* seorang pemain pada satu musim sebesar 29.89047619, dengan dengan nilai terendah sebesar 27.6 menunjukkan bahwa pemain sepak bola kelebihan berat badan dan nilai terbesarnya yaitu 32.9 menunjukkan bahwa pemain sepak bola obesitas. Pada variabel X2 atau bisa disebut lari sejauh 40 yard (36,6 meter) menunjukkan rata-rata kecepatan dan kekuatan berlari pemain sepak bola sebesar 4.557380952 detik dengan nilai terendah atau tercepat sebesar 4.37 detik dan nilai tertinggi atau terlambat sebesar 4.8 detik. Pada variabel X3 menunjukkan bahwa rata-rata kemampuan lompat vertikal setiap pemain sebesar 34.52380952 inci dengan nilai terendah atau lompatan terpendek sebesar 30.5 inci dan nilai tertinggi atau lompatan tertinggi sebesar 41 inci. Pada variabel X4 menunjukkan bahwa rata-rata kemampuan lompat lebar seorang pemain sepak bola sebesar 116.9285714 inci dengan nilai terendah atau lompatan terdekat sebesar 105 inci dan nilai tertinggi atau lompatan terjauh sebesar 130 inci. Kemudian pada variabel X5 menunjukkan rata-rata kemampuan angkat beban seorang pemain sepak bola secara berulang sebesar 19.9047619 dengan nilai terendah sebesar 14 dan nilai tertinggi sebesar 29. Adapun untuk melihat kecenderungan atau karakteristik *Zero One Inflated* dapat dilihat pada gambar sebagai berikut:



Gambar 1. Histogram Pada Variabel Respon Propgames

Pada Gambar 1 ini menggambarkan tiga kondisi data yaitu terdapat proporsi data 0, proporsi data 1, dan proporsi data 0-1 yang dimana dengan karakteristik data ini tidak bisa

dilakukan analisis regresi biasa karena terdapat keterbatasan data pada rentang (0, 1), maka dilakukanlah analisis Regresi *Zero One Inflated Beta*.

Hasil Pemodelan Regresi *Zero-and-One-Inflated Beta*

Untuk mendapatkan interpretasi yang lebih tepat untuk mengetahui bagaimana setiap variabel prediktor (X) berkontribusi terhadap variabel respon (Y) yaitu dengan membedakan parameter disetiap variabelnya maka diperoleh hasil pengujian data proporsi pemain sepak bola dengan menggunakan *software SAS* didapat hasil sebagai berikut:

Tabel 2. Hasil Estimasi Parameter

Variabel	Parameter	Estimasi	SE	T-Value	P-Value	Keterangan
Intersept	β_0	-6.3365	2.4421	-2.59	0.013	Untuk model $0 < Y < 1$
X3	β_3	0.2141	0.06943	3.08	0.0036*	
X5	β_5	-0.01399	0.03066	-0.46	0.6504	
	φ	8.6232	2.1696	3.97	0.0003	
Intersept	γ_0	45.5487	19.0286	2.39	0.0212	Untuk model $Y = 1$
X4	γ_4	-0.4147	0.1691	-2.45	0.0184*	
Intersept	α_0	-27.0077	26.3559	-1.02	0.3114	Untuk model $Y = 0$
X1	α_1	-1.0664	0.4812	-2.22	0.0322*	
X2	α_2	12.4225	5.7433	2.16	0.0363*	

*) variabel signifikan

Berdasarkan Tabel 2 dapat dilihat bahwa terdapat empat variabel prediktor yang berpengaruh signifikan terhadap variabel responnya yaitu variabel BMI (Body Mass Index) sebagai variabel (X1), Kemampuan Lari Empat Puluh Yard (Fortyyd) sebagai variabel (X2), Lompat Vertikal (Vertical) sebagai variabel (X3), Lompat Lebar (Broad) sebagai variabel (X4). Karena variabel tersebut memiliki nilai p-value yang lebih kecil dari α sebesar 0.05. Dengan demikian, diperoleh empat variabel prediktor yang berpengaruh dalam model dan variabel lainnya dapat dikeluarkan dari model.

Hasil Uji Simultan

Pengujian signifikansi parameter model regresi secara simultan dilakukan untuk mengetahui apakah parameter yang dihasilkan memiliki pengaruh terhadap peubah respon. Pengujian parameter secara simultan data proporsi tim sepak bola dengan menggunakan Uji G seperti pada persamaan (21) sebagai berikut:

Dengan hipotesis:

$$H_0: \beta_0 = \beta_3 = \beta_5 = \gamma_0 = \gamma_4 = \alpha_0 = \alpha_1 = \alpha_2 = 0$$

$$H_1: \text{setidaknya terdapat satu } \beta_j \neq 0, \gamma_j \neq 0 \text{ dan } \alpha_j \neq 0$$

Dengan $j = 1, 2, 3, 4, 5$

Dari data diperoleh nilai uji G sebagai berikut:

$$G_{hitung} = -2(47.3 - 13.6) = 33.7$$

Kriteria uji dengan menggunakan taraf nyata 5%, dari tabel *chi-square* diperoleh $X_{0.05;4}^2 = 9.488$. Maka tampak bahwa $33.7 > 9.488$ sehingga dapat diputuskan bahwa H_0 ditolak, yang artinya terdapat setidaknya satu variabel prediktor yang berpengaruh signifikan terhadap proporsi pemain tim sepak bola.

Hasil Uji Parsial

Pengujian parameter secara parsial pada data tim sepak bola menggunakan uji-t seperti pada persamaan (22)-(24) sebagai berikut:

1. Parameter β_j : Berdasarkan hasil pengujian pada tabel 4.2 dapat dilihat bahwa variabel Lompatan Vertikal sebagai variabel (X3) signifikan, artinya variabel X3 memiliki

pengaruh signifikan terhadap variabel respon pada taraf nyata 5%. Sedangkan pada variabel angkat beban (*Bench*) sebagai variabel (X5) tidak signifikan, artinya variabel X5 tidak memiliki pengaruh signifikan terhadap variabel respon pada taraf nyata 5%.

2. Parameter γ_j : Berdasarkan hasil pengujian pada tabel 4.2 dapat dilihat Lompatan Lebar (*Broad*) sebagai variabel (X4) signifikan, artinya variabel X4 memiliki pengaruh signifikan terhadap variabel respon pada taraf nyata 5%.
3. Parameter α_j : Berdasarkan hasil pengujian pada tabel 4.2 dapat dilihat variabel BMI (*Body Mass Index*) sebagai variabel (X1) dan variabel Lari Empat Puluh Yard (*Fortyyd*) sebagai variabel (X2) signifikan, artinya variabel X1 dan X2 memiliki pengaruh signifikan terhadap variabel respon pada taraf nyata 5%.

Interpretasi Model

Pada model $0 < Y < 1$ menjelaskan bahwa peluang seorang pemain sepak bola dipilih untuk bermain di pertandingan namun tidak sepenuhnya dimainkan pada satu musim dengan rentang nilai proporsi (0, 1) dipengaruhi oleh variabel independen yang berpengaruh signifikan terhadap proporsi pemain sepak bola(Y). Dengan model sebagai berikut:

$$\hat{\pi} = \frac{\exp \{-6.3365 + 0.2141x_3\}}{1 + \exp \{-6.3365 + 0.2141x_3\}},$$

Adapun faktor yang berpengaruh pada peluang seorang pemain sepak bola dipilih untuk bermain di pertandingan namun tidak sepenuhnya dimainkan pada satu musim yaitu lompat vertikal. Misalkan jika lompatan vertikal(X_3) seorang pemain sepak bola mencapai 30 inci maka peluang pemain sepak bola akan dipilih untuk bermain di pertandingan adalah sebesar $\hat{\pi} = 0.5216$ kali lipat dibanding sebelumnya dengan asumsi nilai variabel prediktor lain tetap.

Pada model $Y = 1$ menjelaskan bahwa peluang seorang pemain sepak bola dipilih untuk bermain di pertandingan dengan nilai proporsi (1) dipengaruhi oleh variabel independen yang berpengaruh signifikan terhadap proporsi pemain sepak bola(Y). Dengan model sebagai berikut:

$$f(y = 1) = \frac{\exp \{45.5487 - 0.4147x_4\}}{1 + \exp \{45.5487 - 0.4147x_4\} + \exp \{-27.0077 - 1.0664x_1 + 12.4225x_2\}}$$

Adapun faktor yang berpengaruh pada peluang seorang pemain sepak bola dipilih untuk bermain di pertandingan yaitu lompat lebar. Misalkan jika lompatan lebar(X_4) seorang pemain sepak bola mencapai 100 inci dengan BMI(X_1) mencapai indeks 30 dan kemampuan Lari Empat Puluh Yard(X_2) secepat 5 detik maka peluang pemain sepak bola akan dipilih untuk bermain di pertandingan meningkat sebesar $f(y = 1) = 0.2086$ kali lipat dibanding sebelumnya dengan asumsi nilai variabel prediktor lain tetap.

Pada model $Y = 0$ menjelaskan bahwa peluang seorang pemain sepak bola tidak dipilih untuk bermain di pertandingan dengan nilai proporsi (0) dipengaruhi oleh variabel independen yang berpengaruh signifikan terhadap proporsi pemain sepak bola(Y). Dengan model sebagai berikut:

$$f(y = 0) = \frac{\exp \{-27.0077 - 1.0664x_1 + 12.4225x_2\}}{1 + \exp \{45.5487 - 0.4147x_4\} + \exp \{-27.0077 - 1.0664x_1 + 12.4225x_2\}}$$

Adapun faktor yang berpengaruh pada peluang seorang pemain sepak bola tidak dipilih untuk bermain di pertandingan yaitu BMI dan Lari Empat Puluh Yard. Misalkan jika BMI(X_1) mencapai indeks 30 dan kemampuan Lari Empat Puluh Yard(X_2) mencapai kecepatan 5 detik dengan lompatan lebar(X_4) seorang pemain sepak bola mencapai 100 inci maka peluang pemain sepak bola tidak akan dipilih untuk bermain di pertandingan meningkat sebesar $f(y = 0) = 0,7576$ kali lipat dibanding sebelumnya dengan asumsi nilai variabel prediktor lain tetap.

D. Kesimpulan

Berdasarkan hasil dan pembahasan pada data proporsi pemain tim sepak bola yang telah diperoleh, maka dapat disimpulkan sebagai berikut:

1. Pada penelitian ini terdapat beberapa faktor yang mempengaruhi variabel respon yaitu

variabel BMI (X1), variabel Lari Empat Puluh Yard (X2), variabel Lompat Vertikal (X3), variabel Lompat Lebar (X4).

2. Untuk prediksi peluang pemain sepak bola dapat dikatakan bahwa, pada model $0 < Y < 1$ peluang pemain sepak bola akan dipilih untuk bermain di pertandingan adalah sebesar $\hat{\pi}=0.4514$ kali lipat dibanding sebelumnya dengan asumsi nilai variabel prediktor lain tetap. Pada model $Y = 1$ peluang pemain sepak bola akan dipilih untuk bermain di pertandingan meningkat sebesar $f(y = 1) = 0.2086$ kali lipat dibanding sebelumnya dengan asumsi nilai variabel prediktor lain tetap. Pada model $Y = 0$ peluang pemain sepak bola tidak akan dipilih untuk bermain di pertandingan meningkat sebesar $f(y = 0) = 0.7576$ kali lipat dibanding sebelumnya dengan asumsi nilai variabel prediktor lain tetap.

Acknowledge

Puji dan syukur penulis panjatkan kehadiran Allah SWT karena berkat rahmat dan hidayah-Nya penulis dapat menyelesaikan penelitian ini. Terimakasih kepada kedua orang tua dan keluarga yang selalu mendo'akan dan memberi dukungan baik moral maupun materi kepada penulis. Terimakasih kepada Bapak Dr. Nusar Hajarisman, M.Si yang telah memberi bimbingan kepada penulis hingga penelitian ini selesai. Dosen-dosen Program Studi Statistika Universitas Islam Bandung yang telah banyak memberikan ilmu pengetahuan. Sahabat dan teman-teman serta semua pihak yang telah hingga penelitian ini selesai.

Daftar Pustaka

- [1] Ospina, R.L. & Ferrari, S.L.P (2010). Inflated Beta Distributions. *Stat Papers*, 51, 116-120
- [2] Olga K. (2019). *Advanced Regression Models With SAS and R*. California State University: Taylor & Francis Group, LLC.
- [3] Ferrari S.L.P, & F. Cribari-Neto (2004) Beta Regression For Modelling Rates and Proportions. *Journa of Applied Statistics* 31:799–815.
- [4] Cribari-Neto, F., & Zeileis, A. (2010). Beta Regression in R. *Journal of Statistical Software*, 34(2), 1–24.
- [5] Christopher J.S, Maria S.Melguizo, & Zoran Bursac (2012). Inflated Beta Regression: Zero, One, and Everything in Between. University of Arkansas for Medical Sciences, Little Rock, AR.
- [6] Carolina L. & Maria S. (2020). Estimation of Loss Given Default Distributions for Non-Performing Loans Using Zero-and-One Inflated Beta Regression Type Models. *Degree Project in Mathematic Second Cycle*, Stockholm, Sweden.
- [7] F. Baione, D. Bianca, & P. De Angelis (2021). An Application Of zero-one inflated beta Regression Models For Predicting Health Insurance Reimbursement. *Mathematical And Statistical Methods For Actuarial Sciences And Finance*.
- [8] Bain, L, & Engelhardt. (1992). Introduction to Probability and Mathematical Statistics (2). *California: Wadsworth Publishing Company*.
- [9] Rosa, I. D., & Karyana, Y. (2023). Perkiraan Angka Migrasi Neto Perkelompok Umur Provinsi Jawa Barat Tahun 2020 dan 2021. *Jurnal Riset Statistika*, 35–42. <https://doi.org/10.29313/jrs.v3i1.1773>