Buletin Oseanografi Marina Oktober 2012. vol. 1 47 - 52

ISSN 2089-3507

Regresi Logistik Binomial, Model untuk Toksisitas Logam Berat Timbal Pb terhadap Larva Udang Vannamae

Dwi Haryo Ismunarti

Jurusan Ilmu Kelautan Fakultas Perikanan dan Ilmu Kelautan UNDIP

email : dwiharyois@gmail.com

Abstrak

Model regresi logistik binomial diturunkan untuk mengetahui pengaruh toksisitas logam berat timbal Pb terhadap peluang *(p)* kelulushidupan *Udang Vannamae*. Pendugaan parameter model β dengan metode memaksimumkan *Likelihood* fungsi peluang binom menggunakan metode numerik yaitu dengan iterasi Newton Ralphson. Keberadaan *outlier* dalam pembentukan model akan menurunkan ketepatan model sehingga diperlukan identifikasi keberadaannya. Kemudian model diturunkan tanpa menyertakan *outlier*.

Berdasarkan model diperoleh bahwa hasil pada stadia yang sama setiap penambahan 1 mg/l Pb pada media akan menurunkan peluang untuk lulus hidup sebesar 23%. Pada konsentrasi Pb yang sama peluang untuk lulus hidup udang pada stadia PL 10 lebih besar 156 % dibanding pada stadia PL 5.

Kata kunci : regresi logistik binomial, odd rasio

Abstrack

Binary logistic regression model was derived to determine the toxicity of heavy metal Pb against survival of *Vannamae* larvae. Estimation of the model parameter β by maximizing the likelihood function binomial distribution using the numeric method with Newton Ralphson iteration. The presence of outliers in the modeling will reduce accuracy of the model so that identification of outliers was necessary included. Later models derived without including outliers. Results obtained from the model show that at the same stadia stage each additional 1 mg / l Pb in the media will decreased the survival rate of the larvae by 23%. While the survival rate on stadia PL 10 was 156% greater than on stadia PL 5.

Keywords: binomial logistic regression, odds ratios

Pendahuluan

Peubah respon dalam suatu penelitian tidak selalu merupakan hasil pengukuran yang bersifat kontinyu seperti halnya pengukuran panjang dan berat. Data respon adakalanya bersifat diskret. Variabel acak mengikuti fungsi peluang diskret jika merupakan bilangan cacah dan berupa bilangan bulat.

Data respon dalam penelitian bioassay hanya ada dua kemungkinan yaitu individu yang lulus hidup dan individu yang mati. Kelulushidupan individu tergantung pada faktor-faktor abiotik (fisika dan kimia) dan faktor-faktor biotik. Pengaruh faktor-faktor terhadap kelulushidupan tidak dapat diketahui dengan pasti, akan tetapi pengaruh dapat dimodelkan dengan fungsi peluang *(p)*. Setiap individu akan lulus hidup atau mati tidak diketahui dengan pasti, tetapi merupakan kuantitas dan dinamakan variable random (Y). Individu yang lulus hidup dilambangkan Y=1 dengan peluang *p* atau p*(Y=1)=p*, sedangkan individu yang mati dilambangkan Y=0 dengan peluang *1-p* dilambangkan p*(Y=0)=1-p*. Peluang kejadian setiap individu adalah $P(Y=y)=p^y(1-p)^{1-y}$ yaitu y=1 untuk lulus hidup dan y=0 untuk mati. Fungsi distribusi peluang dari kejadian ini dinamakan *distribusi peluang Bernoulli (*Casella dan Berger, 1990 dan Bain dan Engelhardt,1992)

Jika ada *n* individu yang diujikan, sedangkan R_j variabel acak dari individu ke j (j=1,2, .., n) dengan R_j masing –masing berdistribusi peluang Bernoulli. Jika Y adalah

*) Corresponding author
laboska_undip@yahoo.com

http ://ejournal.undip.ac.id/index.php/buloma

Diterima/Received : 16-07-2012
Disetujui/Accepted : 12-08-2012

ISSN 2089-3507

banyaknya individu yang lulus hidup diantara *n* yang diujikan maka Y merupakan variabel acak yang merupakan jumlah dari kejadian Bernoulli dan diasumsikan setiap kejadian independent satu dengan yang lain. Peluang kejadian *Y* individu dimana ($y \le n$) yang lulus hidup dari *n* individu yang diujikan adalah $P(Y=y) = \binom{n}{y} p^y (1-p)^{n-y}$ untuk $y = 0, 1, .., n$. Variabel random Y dikatakan memiliki fungsi *peluang binomial (*Casella dan Berger, 1990 dan Bain dan Engelhardt,1992) .

Tujuan utama analisis pemodelan data biner adalah membuat model matematika yang menjelaskan hubungan antara respon variabel berupa data biner dan satu atau lebih variabel penjelas. Hal yang utama adalah mentransformasi data biner dengan skala (0,1) menjadi skala (-∞,∞). Salah satu adalah transformasi logistik dari peluang *p* yaitu $\log\left(\frac{p}{1-p}\right)$= logit (*p*). Model linear dari hasil transformasi logit dinamakan regresi logistik binomial (Collett, 1991).

Misalkan y adalah individu yang lulus hidup dari *n* individu dalam observasi dengan *p* adalah peluang untuk lulus hidup. Sedangkan *k* banyaknya variable penjelas atau faktor yang diamati x_1 x_2 .. x_k. Model regresi logistik binomial dari variabel dependen *k* adalah

$$\pi_i = P(Y = 1|X) = \frac{e^{\beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2} + \dots + \beta_p x_{jp}}}{1 + e^{\beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2} + \dots + \beta_p x_{jp}}} \quad \dots\dots (1)$$

transformasi logit terhadap π_i diperoleh

$$logit(\pi_i) = log\left(\frac{\pi_j}{1-\pi_j}\right) = \beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2} + \dots + \beta_p x_{jp} \quad \dots\dots(2)$$

(Collett, 1991).

Peluang π_i dan variabel x berhubungan secara sigmoid sedangkan $logit(\pi_i)$ dan variabel x berhubungan secara linier.

Model regresi logistik binomial merupakan model linier umum (*generalized linear model GLM*). GLM lebih fleksibel dari pada regresi linier biasa karena memungkinkan untuk variabel respon yang memiliki distribusi selain normal. GLM pertama kali diperkenalkan oleh Nelder dan Wedderburn tahun 1972 (Collett, 1991).

Materi dan Metode

Materi penelitian adalah data kelulushidupan larva Udang *Vannamae* dengan variabel penjelas konsentrasi logam berat timbal Pb dan stadia pasca larva. Penelitian dilakukan untuk mengetahui toksisitas logam berat timbal Pb dengan konsentrasi 0 mg/l, 2.63 mg/l, 6.6 mg/l, 16 mg/l, 20.1 mg/l dan 50.48 mg/l. Sub stadia larva yang diujikan PL5 dan PL 10 setiap kombinasi perlakuan dilakukan pengulangan sebanyak tiga kali.

Pendugaan parameter model β dengan metode memaksimumkan *Likelihood*. *Likelihood* fungsi peluang binom yaitu

$$L(\beta) = \prod_{i=1}^{n} \binom{n_i}{y_i} \pi_i^{y_i} (1-\pi_i)^{n_i - y_i} \quad \dots. (3)$$

Berdasarkan persamaan (1) *Likelihood* merupakan fungsi π_i yang tidak diketahui dan bergantung pada nilai parameter β maka untuk memaksimumkan fungsi $L(\beta)$ ekivalen dengan memaksimumkan log L(β) (Collett, 1991) dimana

$$Log\ L(\beta) = \sum_i \left\{ log \binom{n_i}{y_i} + y_i log\ \pi_i + (n_i - y_i) \log(1-\pi_i) \right\} \quad \dots\dots (4)$$

Maksimum fungsi diperoleh dengan mencari turunan parsial fungsi log *Likelihood* log L(β) terhadap parameter β_j dimana j =0,1,2,.. k yaitu $\frac{\partial \log L(\beta)}{\partial(\beta_j)} = 0$ .. .... (5)

Persamaan 5 merupakan persamaan tidak linear dalam β_j sehingga penduga β hanya bisa diperoleh dengan menggunakan metode numerik yaitu dengan metode iterasi Newton Ralphson yaitu dengan mencari turunan parsial ke dua terhadap β dari fungsi log L(β) yaitu $\frac{\partial^2 \log L(\beta)}{\partial\beta_i \partial\beta_j} = 0$.

Iterasi Newton Raphson diperoleh dengan program *generalized linear model* dari MINITAB.

Pengujian signifikansi model dan parameter

1. Uji seluruh model (uji G)

ISSN 2089-3507

H0 : $\beta_1 = \beta_2 = \ldots = \beta_p = 0$ (semua koefisien sama dengan nol)

H1 : paling sedikit ada satu koefisien j dengan $\beta_j \neq 0$

Statistik uji : $G = -2 \ln \frac{likelihood\ tanpa\ variabel\ bebas}{likelihood\ dari\ seluruh\ model}$

Statistik G berdistribusi chi kuadrat dengan derajad bebas p. Kriteria uji ditolak H0 jika $G > \chi^2_{p;\alpha}$

2. Uji siknifikansi masing-masing parameter (uji wald)

   H0 : $\beta_j = 0$ untuk setiap j = 0,1,2, … p

   H1 : $\beta_j \neq 0$

   Statistic uji $W = \frac{\beta_j}{SE(\beta_j)}$ dengan j = 0,1,2, … p

   W berdistribusi Normal baku. Kriteria penolakan H0 jika W < - $Z_{\alpha/2}$ atau W > $Z_{\alpha/2}$

Interpretasi koefisien β_j

Menginterpretasikan koefisien β_j akan lebih mudah menggunakan Ood yaitu rasio peluang individu lulus hidup terhadap peluang individu mati $ood = \frac{\pi_j}{1-\pi_j}$. Jika terdapat dua gugus data (misalkan dari dua perlakuan) untuk membandingkan keduanya maka ukuran relatif dari kedua gugus data yaitu rasio dari ood satu gugus terhadap ood gugus data yang lain dan dinamakan ood rasio $\left(ood\ rasio = \frac{ood1}{ood2} = \frac{\left(\pi_1/(1-\pi_1)\right)}{\left(\pi_2/(1-\pi_2)\right)}\right)$. Jika peluang untuk lulus hidup dari dua perlakuan sama maka nilai ood rasio sama dengan satu (Collett,1991). Berdasarkan persamaan 2 maka ood adalah $eks\ (\beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2} + \ldots + \beta_p x_{jp})$. Jika perlakuan X berupa variabel kategorik maka nilai ood rasio diinterpretasikan sebagai resiko atau kecenderungan lulus hidup dari individu pada kategori X_k adalah sebesar exp(β_k) kali resiko atau kecenderungan lulus hidup dari individu pada kategori pembanding. Sedangkan jika perlakuan X berupa variabel kontinyu maka koefisien pada model regresi diinterpretasikan sebagai setiap kenaikan b satuan unit dari perlakuan Xk akan mengakibatkan resiko atau kecenderungan lulus hidup dari individu sebesar exp(b. β_k) kali.

$$Ood = \frac{\pi_j}{1-\pi_j} = exp(\beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2} + \ldots + \beta_p x_{jp})$$

Ood rasio untuk faktor ke k adalah $Ood\ rasio = exp(\beta_k)$

Statistik untuk menguji kesesuaian model secara keseluruhan digunakan statistik Pearson yaitu

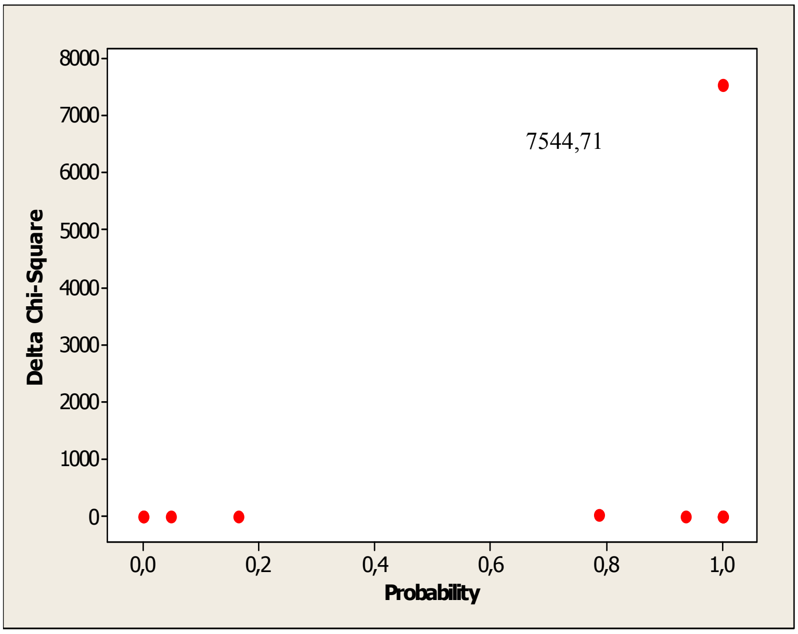
$$Pearson = \sum_j r_j^2\ dengan\ r_j = \frac{y_i - n_i\hat{\pi}_i}{\sqrt{n_i\hat{\pi}_i(1-\hat{\pi}_i)}}.$$

Hasil dan Pembahasan

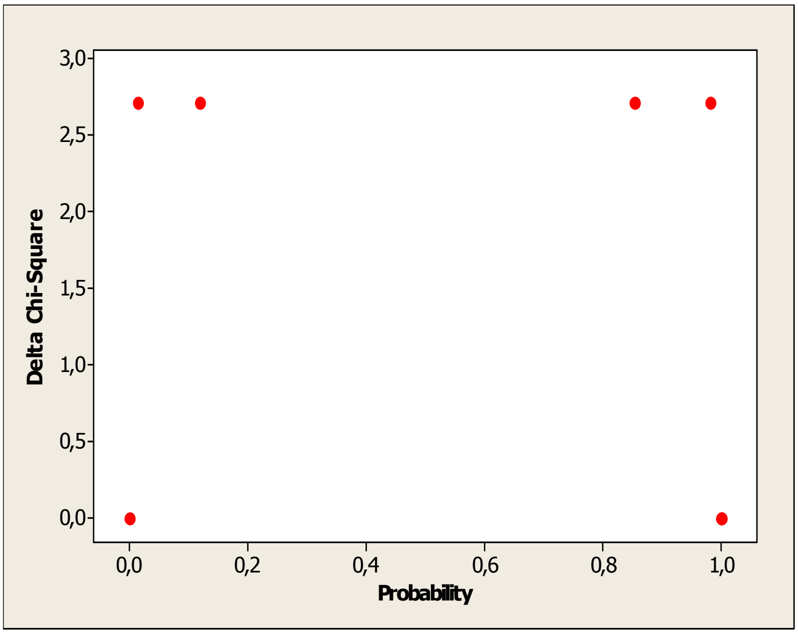
Penelitian dilakukan untuk mengetahui pengaruh pemberian logam berat timbal Pb terhadap kelulus hidupan larva Udang *Vannamae* windu pada sub stadia PL 5 dan PL 10 dengan waktu dedah selama 96 jam. Variabel respon (y) berupa jumlah individu yang lulus hidup dari 30 individu hewan uji dengan demikian variabel y mengikuti fungsi peluang binom. Perlakuan konsentrasi Pb merupakan variabel kontinu sedangkan stadia merupakan variabel kategorik. Model untuk menentukan pengaruh perlakuan terhadap peluang untuk lulus hidup digunakan analisi regresi *logistic binomial.*

Outlier adalah data respon yang menyimpang jauh dari pengamatan lain. Data respon dikatakan menyimpang jika nilai delta χ^2 lebih besar dari 3,84 (Hosmer dan Lemeshow, 2000). Identifikasi keberadaan *Outlier* diperlukan untuk pembentukan model. Identifikasi digunakan plot delta χ^2 terhadap penduga peluang.

ISSN 2089-3507



a. Model dengan satu outlier



b. Model dengan satu outlier

Gambar 1. Plot antara probability dan delta χ^2

Gambar 1 (a) menunjukkan adanya satu *outlier* dengan Delta χ^2=7544.71, dan ada 6 pengamatan lain dengan Delta χ^2 lebih besar dari nol. Keberadaan *outlier* akan mengganggu pendugaan model, sebaiknya model diturunkan dengan tidak menyertakan pengamatan tersebut (Collett, 1991). Gambar 1 (b) menunjukkan tidak ada lagi *outlier* dengan Delta χ^2 lebih besar dari 3,84. Dengan tidak menyertakan *outlier* dalam model juga menurunkan sisaan *Pearson* dari 7545,80 menjadi 2,71247. Sisaan *Pearson* digunakan untuk menguji kesesuain model. Model yang sesuai jika memiliki nilai sisaan *Pearson* mendekati nol.

ISSN 2089-3507

Tabel 1. Nilai maksimum log *Likelihood* pada setiap itersi

Itersi ke	nilai maksimum Log *Likelihood*
0	-622,070
1	-305,520
2	-238,040
3	-183,287
4	-141,676
5	-101,618
6	-88,132
7	-85,983
8	-85,755
9	-85,749
10	-85,749
11	-85,749

Tabel 1 menunjukkan nilai maksimum log *Likelihood*, pada itersi ke 9, 10 dan 11 memiliki nilai yang sama artinya penyelesaian numerik dari iterasi *Newton Ralphson* akan terhenti pada iterasi ke 10.

Tabel 2. Regresi logistik binomial

Variable X	koef (β)	SE koef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	22,8135	2,12949	10,71	0,000			
Pb	-1,45439	0,152159	-9,56	0,000	0,23	0,17	0,31
stadia	0,441521	0,124462	3,55	0,000	1,56	1,22	1,98

Log-Likelihood = -85,749
Test that all slopes are zero: G = 1072,640, DF = 2, P-Value = 0,000

Tabel 2 menunjukkan penduga koefisien, standard error dari koefisien, z hitung dan nilai kesalahan α untuk pengujian koefisien, rasio odds dan interval kepercayaan 95% untuk rasio odds. Penduga koefisien untuk Pb (z = -9.56, p = 0) dan stadia (z = 3.55, p = 0) nilai p kurang dari 0,05, menunjukkan bahwa ada bukti yang cukup bahwa koefisien tidak nol pada level α = 0,05. Penduga koefisien untuk Pb diperoleh -1,145 merupakan perubahan rasio $\frac{\log p(lulus\ hidup)}{\log p(mati)}$ pada setiap penambahan 1 mg/l Pb pada stadia yang sama. Penduga koefisien untuk stadi diperoleh 0.44 merupakan perubahan rasio $\frac{\log p(lulus\ hidup)}{\log p(mati)}$ dari PL 5 dan PL 10 dengan konsentrasi Pb yang sama.

Penduga koefisien untuk Pb bernilai negative dengan Odd rasio diperoleh 0.23 menunjukkan bahwa setiap kenaikan 1mg/l Pb menurunkan peluang untuk lulus hidup sebesar 0.23 kali atau hampir 25 % individu akan mati dengan meningkatnya 1 mg/l Pb di media. Untuk koefisien stadia diperoleh 0.44 dengan odd rasio 1.56 menunjukkan bahwa pada konsentrasi Pb yang sama kelulus hidupan stadia PL 10 lebih besar 156 % dibanding stadia PL 5.

ISSN 2089-3507

Statistik G digunakan untuk menguji hipotesis nol bahwa semua koefisien yang terkait dengan variabel bebas sama dengan nol dibandingkan hipotesis alternatif ada koefisien tidak sama dengan nol. Tabel 2 diperoleh G = 1072,640 dengan nilai p= 0, menunjukkan bahwa ada bukti yang cukup bahwa setidaknya salah satu dari koefisien berbeda dari nol.

Tabel 3. Ukuran Asosiasi antara pengamatan dan penduga peluang dari model

Pairs	cacah	Persentase	Statistik	
Concordant	203274	98,8	Somers' D	0,99
Discordant	414	0,2	Goodman-Kruskal Gamma	1,00
Ties	2133	1,0	Kendall's Tau-a	0,47
Total	205821	100,0		

Tabel 3 menunjukkan ukuran asosiasi antara data pengamatan/respon dan penduga peluang dari model . Concordan adalah persentase pasangan yang sesuai antara pengamatan dan dugaan model yaitu jika stadia PL 10 memiliki probabilitas yang lebih tinggi untuk lulus hidup dari pada PL 5 , dari Tabel 3 terdapat 98.8% pasangan sesuai antara pengamatan dan model. Discordan adalah pasangan yang tidak sesuaiyaitu ketika stadia PL 10 memiliki probabilitas yang lebih kecil untuk lulus hidup dari pada PL 5 diperoleh 0.2 % saja. Ties adalah pasangan dengan peluang yang sama dalam hal ini ada 10 %. Somers 'D, Goodman-Kruskal Gamma, dan Kendall Tau-a adalah ringkasan dari tabel pasangan concordan dan discordan nilainya antara 0 dan 1. Nilai-nilai yang mendekati satu menunjukkan bahwa model memiliki kemampuan prediksi yang lebih baik.

Kesimpulan

Model regresi logistic binomial diturunkan dengan menggunakan pendekatan metode numerik iterasi Newton Ralpson. Iterasi ke 10 berhenti setelah diperoleh nilai maksimum *Likelihood* yang sama pada dua iterasi ke 10 dan ke 11. Hasil menunjukkan adanya satu *outlier* dengan Delta χ^2=7544.7, keberadaan *outlier* akan mengganggu ketepatan model sehingga perlu diturunkan model kembali tanpa *outlier*. Dari model diperoleh hasil pada stadia yang sama setiap penambahan 1 mg/l Pb pada media akan menurunkan peluang untuk lulus hidup sebesar 23%. Peluang untuk lulus hidup udang pada stadia PL 10 lebih besar 156 % dibanding pada stadia PL 5.

Terimakasih

Ucapan terima kasih Penulis sampaikan kepada Saudara Dhira Kurniawan S, MSi, Dr.Ir. Bambang Yulianto, DEA dan Drs. Ali Ridho, MSi. Para pereview yang membuat tulisan ini menjadi lebih baik.

Daftar Pustaka

Aitkin M., Anderson D., Francis B. and Hinde J. 1990. Statistical Modelling in GLM. Oxford Science Publishing, New York

Bain, L.J. and M. Engelhardt. 1992. *Introduction to Probability and Mathematical Statistics*. Duxbury Press Belmont, Calofornia

Casella, G and R.L. Berger. 1990. Statistical Inference. Wadsworth & Brooks/Cole Publishing, California

Collett, D. 1991. Modelling Binary Data. Chapman & Hall, New York

Hosmer,D.W. and S. Lemeshow. 2000. Applied Logistic Regression. 2nd ed. John Wiley & Sons, New York