

**ANALISIS REGRESI POISSON UNTUK MENDUGA HUBUNGAN
KELIMPAHAN MAKROBENTHOS DENGAN PARAMETER PERAIRAN
(STUDI KASUS SUNGAI BANJIR KANAL BARAT SEMARANG)**

Dwi Haryo Ismunarti¹, Ria Azizah TN², Rochdi Wasono³

^{1),2)} Staf Pengajar Jur. Ilmu Kelautan UNDIP

³ Staf Pengajar PS Statistika UNIMUS

dwharyois@gmail.com

Abstrak

Fungsi peluang poisson merupakan standar model untuk variabel cacah dengan fungsi peluang poisson. Sebagai contoh kelimpahan makrobenthos dan faktor-faktor oseanografi yang mempengaruhinya dapat didekati dengan model regresi poisson. Model regresi poisson dapat diduga menggunakan fungsi GLM (generalized linear models) dari program S-PLUS. Faktor-faktor oseanografi yang berpengaruh terhadap kelimpahan makrobenthos adalah salinitas, kecerahan, kecepatan arus, DO dan suhu.

Kata Kunci : Peluang poisson, Variabel cacah

1. Pendahuluan

Pengalaman membangun sumberdaya pesisir dan laut selama kurun PJP I selain menghasilkan keberhasilan juga menimbulkan permasalahan ekologis dan sosial ekonomis. Secara ekologis banyak kawasan pesisir dan laut, antara lain Pantai Utara Jawa terancam kapasitas keberlanjutannya karena pencemaran, degradasi fisik habitat, over eksploitasi sumberdaya alam dan konflik penggunaan lahan (Bengen,2002).

Timbulnya pencemaran pada suatu perairan dapat berakibat menurunnya kualitas air yaitu berubahnya sifat fisika dan kimia air. Kualitas perairan sungai dapat diketahui dengan metode kuantitatif menggunakan kelimpahan spesies sebagai indikatornya. Salah satu pendekatan dapat digunakan makrobenthos melalui monitoring jumlah populasi, komposisi komunitas maupun fungsi ekosistemnya. Organisme benthos merupakan jenis hewan yang hidup melekat atau relatif tidak bergerak dengan pola penyebaran yang khas. Struktur komunitasnya akan berubah karena perubahan lingkungan yang antara lain disebabkan oleh pencemaran perairan.

Keberadaan suatu individu E di luasan tertentu dapat dipandang sebagai kejadian ‘ada’ dan ‘tidak ada’. Peluang E ‘ada’ adalah $p(E)=p$ sedangkan peluang E ‘tidak ada’ adalah komplement E’ yaitu $q=p(E')=1-p$. Kelimpahan adalah jumlah individu E ada di setiap sampling unit (n_e). Jika keberadaan individu-individu di suatu luasan tertentu dapat dianggap independen dan peluang keberadaan individu $p(E)$ kecil maka fungsi peluang dari kelimpahan $p(x = n_e)$ yang merupakan variabel cacah akan mengikuti fungsi peluang kejadian Poisson (Casella & Berger, 1990, Bain & Engelhardt, 1992). Model regresi yang tepat untuk jenis data ini adalah model regresi poisson.

2. Materi dan Metode Penelitian

Jika peubah y adalah jumlah makrobenthos pada pengamatan ke- i pada luasan tertentu A maka μ_i merupakan nilai tengah dari y_i . Kemudian μ_i dimodelkan sebagai fungsi dari k peubah bebas sehingga diperoleh $\mu_i = A\mu(x_i; \beta); i = 1, 2, \dots, k$ dengan i menyatakan pengamatan ke- i .

Fungsi $\mu(x_i; \beta)$ dapat dipilih fungsi yang tidak negatif misalkan $e^{x_i\beta} > 0$.

Model regresi poisson dengan $e^{x_i\beta} > 0$ adalah $\hat{\mu} = Ae^{x_i\beta}$.

Nilai β diduga dengan menggunakan metode kemungkinan maksimum sebagai berikut :

$$\begin{aligned} L(y, \beta) &= \prod_{i=1}^k p(y_i; \beta) \\ &= \prod_{i=1}^k \frac{[A\mu(x_i, \beta)]^{y_i} e^{-[A\mu(x_i, \beta)]}}{y_i!} \\ &= \frac{\prod_{i=1}^k [A\mu(x_i, \beta)]^{y_i} e^{-\sum_{i=1}^k A\mu(x_i, \beta)}}{\prod_{i=1}^k y_i!} \end{aligned}$$

$$\text{dan } \ln L(y, \beta) = \sum_{i=1}^k y_i \ln A\mu(x_i, \beta) - \sum_{i=1}^k A\mu(x_i, \beta) - \sum_{i=1}^k \ln y_i!$$

kemudian $\frac{\partial \ln L(y, \beta)}{\partial \beta} = 0$ sehingga diperoleh hasil

$$\sum_{i=1}^k \left[\frac{y_i}{\mu(x_i, \hat{\beta})} - A \right] \left[\frac{\partial \mu(x_i, \hat{\beta})}{\partial \hat{\beta}} \right] = 0$$

Nilai penduga β diperoleh dari persamaan di atas dengan menggunakan metode *Poisson Iteratively Reweighted Least Squares (PIRLS)* (Peterson, 1997).

Pengujian terhadap penduga β digunakan statistik χ^2 dengan hipotesis

$$H_0 : \beta_j = 0 \text{ dan } H_1 : \beta_j \neq 0; \quad \chi^2_{hitung} = \frac{b_j^2}{c_{jj}} \quad j = 0, 1, \dots, p$$

dengan p jumlah peubah bebas dan b_j penduga β untuk peubah bebas ke- j

$$c_{jj} = \frac{-\partial^2 \ln L(\hat{\beta})}{\partial \beta_j^2}, \quad j = 0, 1, \dots, p$$

Model regresi poisson diturunkan dengan *fungsi GLM (Generalize Linear Models)* dalam program S-PLUS 2000 dengan mendefinisikan formula modelnya, fungsi peluang dan fungsi linknya.

Pengambilan keputusan dalam pemodelan statistika bergantung pada kevalidan dari model. Kehati-hatian dalam mengevaluasi data dengan model merupakan bagian terpenting dalam pemodelan data statistika. Empat hal yang memungkinkan ketidaktepatan suatu pemodelan adalah: (1) tidak adanya fungsi peluang yang spesifik dari variabel tak bebas y , (2) tidak ada *fungsi link* yang spesifik, (3) tidak ada bagian sistematis yang spesifik dari model dan (4) adanya penyimpangan data observasi.

Evaluasi terhadap model digunakan sisaan (residual). Sisaan didefinisikan sebagai selisih amatan (data) y_i dan nilai ramalan (dugaan) yang diperoleh dari model

\hat{y}_i , $e_i = y_i - \hat{y}_i$, $i = 1, 2, \dots, n$. Pemeriksaan sisaan bermanfaat dan berlaku pada keadaan yang melibatkan pendugaan model seperti regresi linear maupun nonlinear dan model analisis ragam (Draper & Smith, 1992). Pemeriksaan terhadap sisaan melalui grafik dengan lebih seksama kemungkinan akan membawa ke arah model yang lebih baik atau metode pendugaan yang lebih sesuai. Keefektifan analisis data tergantung pada kegigihan dalam mendapatkan model yang relevan dan menyingkap hal-hal yang masih tersembunyi dalam sisaan.

Sisaan Pearson dapat digunakan untuk menguji kesesuaian model regresi poisson. Sisaan ini mendekati simpangan baku respon dengan nilai tengah nol dan varian mendekati satu (McCullagh & Nelder, 1989). Sisaan Pearson adalah : $r_p = \frac{y - \mu}{\sqrt{V(\mu)}}$

Jumlah kuadrat dari sisaan pearson adalah statistik Chi-squared yaitu

$$\chi^2 = \sum_{i=1}^n \frac{\left(y_i - \hat{\mu}_i \right)^2}{V\left(\hat{\mu}_i \right)}$$

Dalam analisis statistika untuk mengevaluasi adanya penyimpangan data pengamatan ke-i sehingga berpengaruh terhadap pendugaan parameter digunakan beda antara pendugaan dengan melibatkan pengamatan ke-i dan pendugaan dengan tanpa melibatkan pengamatan ke-i. Statistik tersebut adalah

Cook's Distance yaitu $D_i = \frac{h_i r_{pi}^2}{d(1-h_i)}$, h_i adalah diagonal matriks hat, r_{pi} sisaan

pearson terbakukan dan d banyaknya variabel penjelas. Selanjutnya nilai D_i dibandingkan dengan nilai tabel $F(d, n-d, 1-\alpha)$ untuk α yang ditentukan. D_i yang besar menandakan pengamatan ke-i sangat berpengaruh (Draper & Smith, 1992). Plot D_i dengan observasi akan memudahkan dalam analisis.

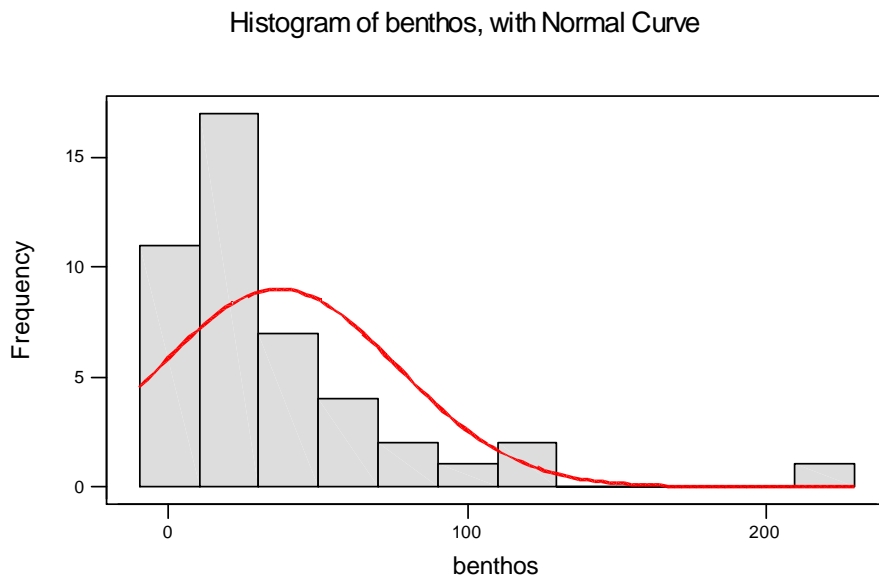
Materi yang digunakan adalah hasil penelitian terhadap parameter kualitas perairan di Sungai Banjir Kanal Barat Semarang pada bulan Juli 2008. Ada 15 stasiun titik sampling yang menyebar dari muara sampai ke badan sungai. Waktu pengamatan dilakukan pengulangan tiga kali. Variabel yang dipergunakan adalah variabel terikat merupakan parameter biologi yaitu kelimpahan makrobenthos (individu/m²) yang

ditemukan per stasiun per waktu pengamatan. Variabel bebas/penjelas merupakan parameter kimia dan fisika perairan : nitrat, DO suhu, salinitas, kecerahan, kedalaman dan kecepatan arus.

3. Hasil dan Pembahasan

Hasil pengamatan menunjukkan kelimpahan makrobenthos di Banjir Kanal Barat rendah dengan kisaran nilai 0 sampai 145 individu/ m². Kelimpahan makrobenthos agak tinggi (28-108 individu/m²) di sekitar muara kemudian menurun dan terendah (0-14 individu/m²) di daerah sekitar jembatan Jalan Arteri. Fenomena yang menarik adalah populasi makrobenthos agak tinggi (17-118 individu/m²) di sekitar daerah kawasan industri (Kelurahan Ngemplak Simongan).

Histogram merupakan salah satu metode untuk menduga pola sebaran. Gambar 1 adalah histogram kelimpahan makrobenthos yang ditemukan pada setiap pengamatan menunjukkan ciri khas dari variabel cacah dengan kemencengan negatif dan *tailed* yang panjang. Pola sebaran ini merupakan typical dari fungsi peluang Poisson



(Anonim,1993).

Gambar 1. Histogram Kelimpahan Makrobenthos

Hasil analisis regresi poisson diperoleh penduga parameter model dan pengujian hipotesis terhadap masing-masing parameter. Hipotesis yang diujikan adalah : $H_0 : \beta_i = 0, i = 1, 2 \dots 8$ dengan hipotesis alternatif $H_1 : \beta_i \neq 0$.

Tabel 1. Hasil dari Fungsi Summary(benthos.fit)

Koefisien	b_i	SE (b_i)	t value
(Intercept)	-18.828	1.643	-11.4544132*
suhu	0.645	0.056	11.5133958*
salinitas	0.039	0.006	6.5248340*
kecerahan	0.042	0.005	8.4678319*
kedalaman	-0.003	0.001	-2.6683791
arus	-0.288	0.027	-10.5006960*
nitrat	0.044	0.103	0.4297283
DO	0.658	0.123	5.3283245*

Tabel 1 menunjukkan bahwa variabel kedalaman dan nitrat tidak berperan besar dalam menentukan keberadaan makrobenthos di perairan. Faktor-faktor yang berpengaruh nyata terhadap peluang keberadaannya berturut-turut adalah salinitas, kecerahan, kecepatan arus, DO dan suhu. Faktor-faktor tersebut berpengaruh secara bersama-sama. Selanjutnya akan diturunkan model baru dengan tidak menyertakan kedua variabel tersebut. Untuk memperoleh model dugaan yang terbaik selanjutnya akan dilakukan pengujian parsial secara beruntun dengan menambahkan satu per satu

variabel-variabel suhu, salinitas, kecerahan, arus dan DO ke dalam model. Pengujian dilakukan dengan menggunakan analisis varian

Tabel 2. Perbandingan Model Menggunakan ANOVA

Response: benthos

Terms added sequentially (first to last)	Dev.	Resid.df	Resid. Dev	Pr(Chi)
NULL		44	1475.600	
suhu	1 31.9523	43	1443.648	1.580e ⁻⁰⁰⁸
suhu+salinitas	1 294.2798	42	1149.368	0.000e ⁺⁰⁰⁰
suhu+salinitas+kecerahan	1 81.5210	41	1067.847	0.000e ⁺⁰⁰⁰
suhu+salinitas+kecerahan+arus	1 128.5854	40	939.261	0.000e ⁺⁰⁰⁰
suhu+salinitas+kecerahan+arus+ DO	1 59.5564	39	879.705	1.000e⁻⁰¹⁴

Nilai-nilai statistik chi-square dari Tabel 2 yang hampir mendekati nol menunjukkan penambahan variabel suhu sangat signifikan menurunkan sebesar 31.9523 keragaman variabel benthos. Hal ini berarti variabel suhu sangat diperlukan dalam model. Meskipun demikian model dengan hanya melibatkan satu variabel suhu belum merupakan model yang terbaik. Penambahan variabel salinitas, kecerahan, arus dan DO secara berturut-turut ke dalam model ternyata sangat signifikan dalam menurunkan simpangan/sisaan dengan demikian variabel-variabel tersebut sangat diperlukan dalam model dan variabel tersebut secara bersama-sama menentukan peluang keberadaan makrobenthos di perairan.

Tabel 3. Pendugaan Model 2

Coefficients	bi	SE(bi)	t value
(Intercept)	-19.16105927	1.607190693	-11.922082
suhu	0.62967088	0.053500497	11.769440
salinitas	0.04308502	0.005784056	7.448929
kecerahan	0.04252950	0.004721578	9.007475
arus	-0.31780545	0.024948448	-12.738486
DO	0.78955590	0.103374746	7.637803

(Dispersion Parameter for Poisson family taken to be 1)

Null Deviance: 1475.6 on 44 degrees of freedom

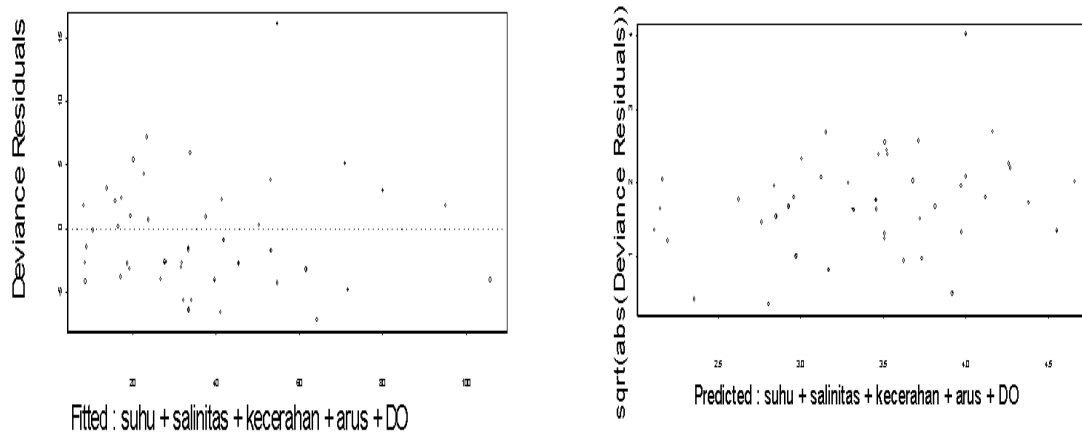
Residual Deviance: 879.705 on 39 degrees of freedom

Number of Fisher Scoring Iterations: 4

Tabel 3 menunjukkan penduga koefisien regresi dari model 2. Tanda negatif pada koefisien arus berarti semakin tinggi kecepatan arus semakin kecil peluang keberadaan makrobenthos ditemukan. Sedangkan tanda positif dari variabel suhu, salinitas, kecerahan dan DO menunjukkan semakin tinggi variabel-variabel tersebut semakin besar peluang keberadaan makrobenthos ditemukan pada kondisi perairan yang seperti ini.

Pemeriksaan Sisaan

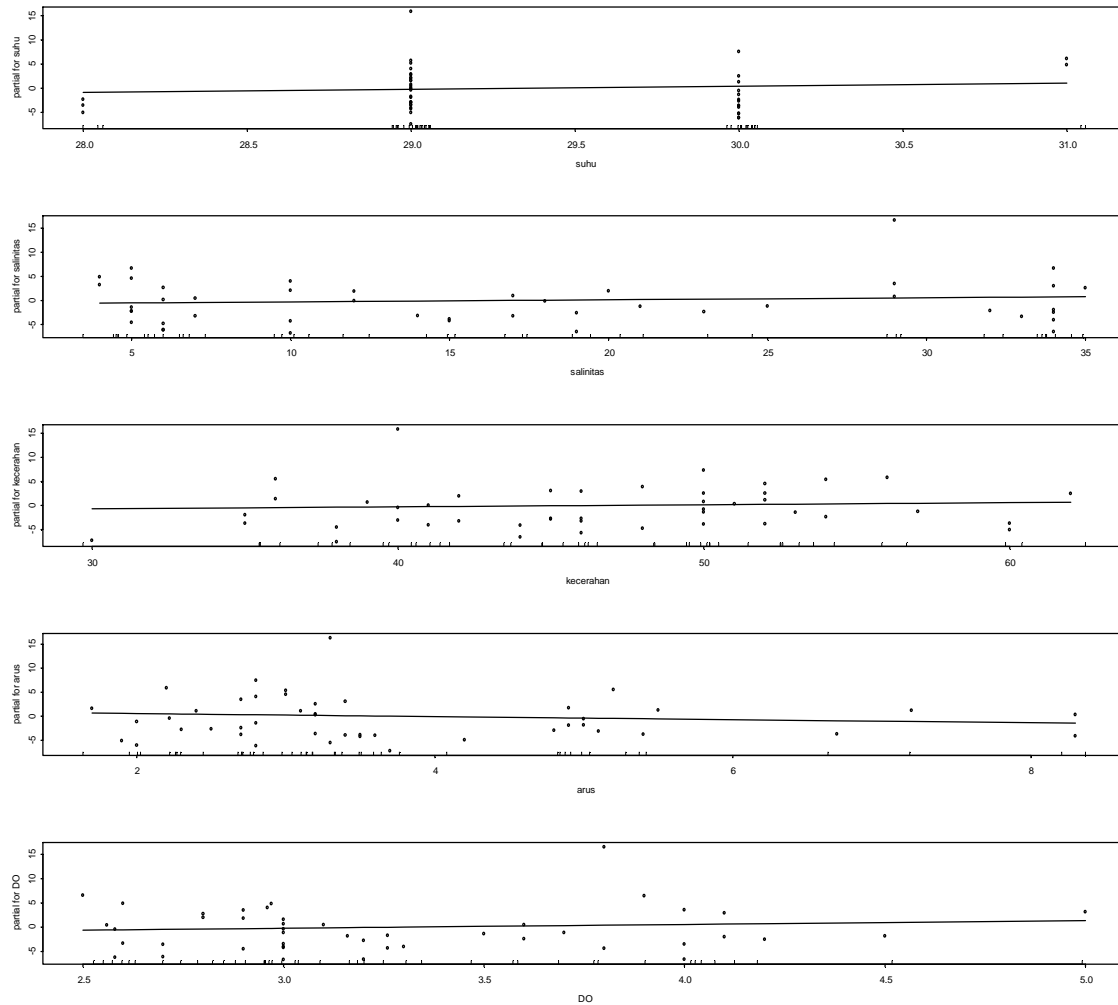
Pemeriksaan terhadap sisaan (residual) sangat bermanfaat untuk mengevaluasi kebaikan model penduga. Sisaan yang akan digunakan dalam model regresi poisson adalah sisaan Pearson. Sisaan ini mendekati simpangan baku respon dengan nilai tengah nol dan varian mendekati satu (McCulagh & Nelder, 1989). Jumlah kuadrat dari sisaan pearson adalah statistik Chi-squared.



Gambar 2 a. Plot Sisaan vs Penduga **b.** Plot Simpangan Baku Sisaan vs Penduga

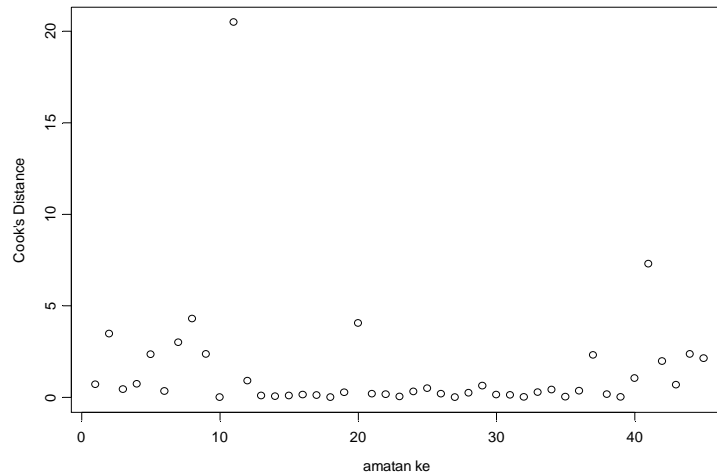
Gambar 2a plot sisaan dengan penduga dan **2b** simpangan baku sisaan dengan penduga tidak menampakkan adanya pola kenaikan keragaman dengan demikian tidak perlu dilakukan transformasi terhadap variabel.

Plot sisaan terhadap masing-masing variabel penjelas secara parsial jika diperoleh mendekati pola garis lurus maka tidak perlu adanya transformasi terhadap variabel tersebut.



Gambar 3. Plot Sisaan Terhadap Masing-Masing Variabel

Gambar 2 dan 3 di atas tidak menunjukkan perlu adanya transformasi terhadap variabel. Dari gambar tampak ada satu nilai pengamatan yang selalu jauh dari pengamatan lainnya pengamatan tersebut memiliki sisaan sebesar 21.237. Pengamatan tersebut memiliki pengaruh besar terhadap pendugaan parameter model, untuk itu perlu dilakukan evaluasi apakah pengamatan tersebut perlu dikeluarkan ataukah tidak sehingga diperoleh model penduga yang terbaik.



Gambar 4. Plot Statistik Cook's Distance

Plot Gambar 4 menunjukkan terdapat 2 pengamatan yang ganjil yang jauh dari kumpulan data lainnya. Pengamatan tersebut dengan nilai $D_{11} = 20.4916$ dan nilai $D_{41} = 7.293822$. Nilai-nilai tersebut lebih besar dari $F(5, 40, 95\%) = 4.46$. Tidak adanya hal yang khusus dari kedua pengamatan untuk selanjutnya akan dilakukan pemodelan kembali dengan tidak menyertakan kedua pengamatan tersebut.

Hasil pada Tabel 4 menunjukkan dengan mengeluarkan dua pengamatan akan diperoleh model yang lebih baik dengan criteria penurunan sisaan Deviance dari $879.705/39 = 22.559$ (Tabel 3) menjadi $527.1748/37 = 14.287$ (Tabel 4).

Tabel 4. Analisis Model 3

Coefficients	Value	Std. Error	t value
(Intercept)	-14.32695443	1.909122850	-7.504470
Suhu	0.46470393	0.064021168	7.258598

salinitas	0.03763404	0.006194789	6.075112
kecerahan	0.05291657	0.005170446	10.234431
kec.arus	-0.21472136	0.026694470	-8.043664
DO	0.51997455	0.106964778	4.861175

(Dispersion Parameter for Poisson family taken to be 1)

Null Deviance: 926.2559 on 42 degrees of freedom

Residual Deviance: 527.1748 on 37 degrees of freedom

Number of Fisher Scoring Iterations: 4

4. Kesimpulan dan Saran

Hasil pengamatan menunjukkan kelimpahan makrobenthos di Banjir Kanal Barat rendah dengan keragaman tinggi. Faktor-faktor yang berpengaruh nyata terhadap peluang keberadaannya berturut-turut adalah salinitas, kecerahan, kecepatan arus, DO dan suhu. Faktor-faktor tersebut berpengaruh secara bersama-sama. Model terbaik diperoleh : $\log(\mu) = -14.33 + 0.46 \text{ Suhu} + 0.038 \text{ salinitas} + 0.053 \text{ kecerahan} - 0.215 \text{ kec.arus} + 0.52 \text{ DO}$

Hal yang paling mendasar dalam pendugaan model adalah menelusuri fungsi peluang yang spesifik dari variabel tak bebas y . Tidak terpenuhinya asumsi-asumsi pada metode pendugaan optimum model akan mengakibatkan ketidaktepatan pendugaan dan model yang diperoleh tidak dapat diandalkan. Selain itu keefektifan analisis data juga tergantung pada kegigihan dalam mendapatkan model yang relevan dan menyingkap hal-hal yang masih tersembunyi dalam sisaan. Penelusuran pola sisaan dan cook's distance bermanfaat untuk mendapatkan model terbaik.

Ucapan Terima kasih

Ucapan terima kasih disampaikan kepada mahasiswa Jurusan Ilmu Kelautan FPIK UNDIP sdr. Widodo, Rudiono, Rahmat dan Edy Yusuf Hamzah atas kerja samanya.

Daftar Pustaka

- Aitkin M., Anderson D., Francis B. and Hinde J. 1990. *Statistical Modelling in GLM*. Oxford Science Publishing, New York
- Anonim. 1993. *S-PLUS Guide to Statistical and Mathematical Analysis*. Stat Sci Adivision Of MathSoft.Inc. Washington.
- Aunuddin. 1989. *Analisis Data*. PAU Ilmu Hayat Institut Pertanian Bogor, Bogor.
- Bain, L.J. and M.Engelhardt. 1992. *Introduction to Probability and Mathematical Statistics*. Duxbury Press Belmont, Calofornia.
- Bengen, D.G. 2002. *Konsep Pengelolaan Wilayah Pesisir Secara Terpadu dan Berkelanjutan*. Makalah Pelatihan Pengelolaan Sumberdaya Wilayah Pesisir dan Lautan Semarang, 4-8 Maret 2002
- Casella, G and R.L.Berger. 1990. *Statistical Inference*. Wadsworth & Brooks/Cole Publishing Company, California
- Draper N. and Smith H.1992. *Analisis Regresi Terapan*. Ed. Kedua Terjemahan Bambang Sumantri. PT Gramedia, Jakarta.
- Everitt B. 1994. *A Handbook of Statistical Analyses using S-PLUS*. Champman & Hall , London.
- McCullagh,P. and J.A.Nelder. 1989. *Generalized Linear Model*. Chapman & Hall. London
- Myers,R.H. 1990. *Clasical and Modern Regression with Application*. PWS-Kent Publishing Comp. Boston
- Nasution,A.H. dan A.Rambe. 1984. *Teori Statistika*. Ed. Ke-2. Bhratara, Jakarta.

Peterson, L.E. 1997. *PIRLS: Poisson Iteratively Reweighted Least Squares Computer Program for Additive, Multiplicative, Power, and Non-Linear Models*. Center for Cancer Control Research Baylor College of Medicine Houston, Texas.

Rosenberg, D.M. and Resh V.H. 1993. *Fresh Water Biomonitoring and Benthic Macroinvertebrates*. Chapman & Hall Inc. New York.