

# Pemodelan Kawasan Panen Kelapa Sawit Malaysia Menggunakan Model Logistik

**M. Marizal<sup>1</sup>, Noraini Binti Nazam<sup>2</sup>, Rado Yendra<sup>3</sup>, Ari Pani Desvina<sup>4</sup>**

<sup>1,3</sup>Fakultas Sains dan Teknologi, Jurusan Matematika, Universitas Islam Negeri Sultan Syarif Kasim Riau  
28293 (Pekanbaru-Riau)

<sup>2</sup>Jurusan Matematika, Fakultas Sains dan Teknologi, Universiti Kebangsaan Malaysia,  
43600 Bagi-Selangor, Malaysia  
e-mail: muhammadmarizal@gmail.com

## Abstrak

Minyak kelapa sawit dan produk-produk yang berkaitan merupakan ekspor kedua terbesar Malaysia. Pengeluaran minyak kelapa sawit negara pada tahun 2010 diproyeksi mencapai lebih kurang 64,282,738 ton dalam waktu satu tahun. Namun, hasil ini masih belum mencapai tahap maksimum jika dilihat dari kawasan tanaman sawit yang ada di Semenanjung Malaysia, Sabah dan Sarawak seluas 4,691,160 hektar. Dari luas tersebut, sekitar 2.807 juta hektar kelapa sawit ditanam oleh pihak perkebunan swasta (59.8%) sementara 20.2% dikelola oleh Pemerintah (FELDA 675,167 ha, RISDA 80,262 dan FELCRA 160,832) , 321,947 hektar (7,1%) oleh pemerintah Daerah dan sisanya seluas 540,194 hektar (12.9%) yang sedang garap. Penelitian ini membicarakan seluruh kawasan kelapa sawit yang ditanam berbanding kawasan yang panen, dengan faktor-faktor jumlah pekerja, hasil panen dan harga. Model yang digunakan dalam Penelitian ini ialah metode logistik, dengan model probit, model logit dan pelengkap fungsi log-log (complementary log-log) terhadap 39 lokasi pengamatan kelapa sawit yang diambil dari Jabatan Perangkaan Malaysia. Model yang dipilih berdasarkan ujian statistik yang melibatkan nilai AIC (Akaike information criterion) dari metode logistic proc, nilai devians dari metode Genmod Proc serta nilai p, menjadi acuan yang diambil untuk menentukan model terbaik dalam penelitian ini. Setelah dilakukan perbandingan dalam ujian statistik yang dipilih, menunjukkan bahwa model Pelengkap fungsi Log-log adalah model yang lebih baik dibanding model probit dan model logit.

**Kata kunci:** AIC , devians, genmod proc, kelapa sawit, logistik proc

## Abstract

*Palm oil and related products is Malaysia's second largest export. Palm oil state expenditures in 2010 are projected to reach approximately 64,282,738 tons within a year. However, these results still have not reached the stage of maximum when seen from an oil palm plantation areas in Peninsular Malaysia, Sabah and Sarawak covering an area of 4,691,160 hectares. Of the area, approximately 2,807 million hectares of oil palm plantations planted by the private sector (59.8%) while 20.2% is managed by the Government (FELDA 675.167 ha, 80.262 Risda and Felcra 160.832), 321.947 ha (7.1%) by the local government and the rest covering an area of 540.194 ha (12.9%) are currently working on. This study discusses the entire area of oil palm planted area compared to the harvest, the number of workers factors, yields and prices. The model used in this research is the method of logistics, with the probit model, logit model and complementary log-log function (complementary log-log) to 39 locations palm observations drawn from Jabatan Perangkaan Malaysia. Models are selected based on statistical test that involves the value of AIC (Akaike information criterion) of proc logistic method, the value of the method Genmod Proc devians as well as the value of p, the reference is taken to determine the best model in this study. After comparison of the statistical test is selected, indicating that the model Complement function logs is a better model than the models probit and logit models.*

**Keywords:** AIC, devians, genmod proc, logistic proc, palm oil

## 1. Pendahuluan

Dalam kajian ini, peneliti menetapkan jumlah tanaman sawit yang dipanen sebagai prediktor terhadap luasan penanaman kelapa sawit dengan menetapkan variabel-variabel yang lain seperti seperti jumlah pekerja, pengeluaran buah sawit sehektar (ton) dan harga rata-rata sawit (RM). Secara logika, kawasan yang dipanen mempunyai kaitan dengan semua variabel yang digunakan. Jika diteliti dengan jelas, penambahan ataupun kekurangan jumlah pekerja dalam sektor sawit, turut mempengaruhi hasil panen sawit di kawasan penanaman sawit

seperti yang pernah dilaporkan oleh (Malaysian Palm Oil Association – MPOA). Apabila hal ini terjadi, ia akan menjejaskan pengeluaran buah sawit sehektar (ton). Hal ini sekaligus akan mengganggu harga pokok sawit (RM) tahunan negara. [1]

Metode penelitian yang digunakan dalam penelitian ini melibatkan metode logistik yang terdiri dari tiga penghubung logit, probit dan CLL (*complementary log-log*). Penelitian ini pernah dilakukan oleh peneliti terdahulu untuk menganggar hasil minyak sawit yang turut menimbangkan model terbaik melalui metode taklinear, logistik, logistik log, Morgan-Mercer-Flodin dan model pertumbuhan Chapman-Richard mempunyai kemampuan untuk mengukur satu fenomena pertumbuhan yang mempamerkan corak sigmoid sepanjang masa. Berdasarkan ujian statistik dan kebaikan penyuaiian, model yang terbaik ialah model Logistik dan diikuti oleh model taklinear Morgan-Mercer-Flodin, Chapman-Richard (peringkat awal) dan model pertumbuhan logistik Log. [2]. Walau bagaimanapun, masih terdapat para peneliti yang masih mencoba kesahihan model kelapa sawit menggunakan kaedah regresi linear [2] dan regresi linear berganda sedang digunakan secara meluas dalam penyelidikan untuk menentukan hubungan linear antara faktor [3]; [4].

Dalam membuat keputusan pemilihan model terbaik turut mempertimbangkan model devians (penyelewengan) dengan melihat kepada pembahagian varians untuk mendapatkan model yang terbaik dalam ujian hipotesis statistik [5] dan nilai  $p$  yang turut menjadi perbandingan yang baik dalam menentukan faktor yang terbaik dalam model kerana nilai  $p$  yang kecil dari  $\alpha$  yang ditetapkan bagus di masukkan kedalam model.

Penilaian untuk melihat selang keyakinan turut memainkan peranan bagi pembuktian model yang terbaik dipilih dalam kajian ini [6]. Selang keyakinan yang rendah antara model perlu diambil kira kerana terdapat sedikit perbezaan antara data pemerhatian dengan data ramalan jadi reja yang sedikit bagus dalam sesuatu pemerhatian. Malah, Berdasarkan ujian statistik yang melibatkan ukuran kebaikan relatif patut model statistik (*goodness of fit of a statistical model*) untuk BIC (*Bayesian Information Criteria*) dan AIC (*Akaike information criterion*) diambil kira bagi menentukan model yang mana patut dipilih menjadi model penganggar terbaik dalam kajian ini. Kriteria maklumat Akaike adalah ukuran kebaikan relatif patut model statistik (*goodness of fit of a statistical model*). AIC adalah berasaskan konsep entropi maklumat, kesan menawarkan ukuran relatif maklumat hilang apabila model yang diberikan digunakan untuk menggambarkan realiti data dalam erti kata yang mutlak. Jika semua calon model patut buruk, AIC tidak akan memberi apa-apa amaran itu.

Tambahan itu, kajian ini memperlihatkan graf reja melawan ramalan (*fitted vs predicted*) turut memainkan peranan dalam memberi pembuktian dan kesahihan model yang terbaik seperti dalam analisis statistik seperti yang dibincangkan oleh banyak penyelidik-penyelidik terdahulu [4].

## 2. Metode Penelitian

Pada penelitian ini, data dianalisis menggunakan dua metode yaitu logistic proc dan genmod porc ialah *generalized linear model*. Dalam penelitian model logistik yang digunakan menggunakan tiga penghubung (link) iaitu logit, probit dan log-logistik (CLL). Regresi logistik juga sering disebut model logistik atau model logit. Dalam istilah statistik digunakan untuk ramalan probabilitas suatu kejadian atau peristiwa dengan menyamakan data pada fungsi logit kurva logistik. Metode ini merupakan model linear umum yang digunakan untuk regresi binomial. Seperti analisis regresi pada umumnya, metode ini menggunakan beberapa variabel prediktor, baik berangka maupun kategori. Misalnya, peluang orang yang menderita serangan jantung pada waktu tertentu dapat diramal dari informasi usia, jenis kelamin, dan berat tubuh. Data yang di yang diolah dalam penelitian ini ialah data hasil panen kelapa sawit yang diperoleh dari 39 lokasi area perkebunan kelapa sawit diseluruh Malaysia yang diperoleh dari Jabatan Perangkaan Malaysia [7].

Penjelasan regresi logistik berawal dari fungsi logistik, yang sentiasa mengambil pada nilai antara nol dan satu [8].

$$P(x_i) = \frac{e^{(\beta_0 + \beta_1 x)}}{1 + e^{(\beta_0 + \beta_1 x)}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

**Logit:**

Model logit telah diperkenalkan oleh Joseph Berkson pada tahun 1944. Misalkan  $p$  berada pada nilai 0 and 1 dalam rumus:

$$\log it(E[Y_i|X_i]) = \log it(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta \cdot x_i$$

Di mana:  $\ln\left(\frac{p_i}{1-p_i}\right)$  disebut fungsi logit. Andaikakn  $\beta_0$  adalah pekali bagi model dan

$\beta_1 x$  pekali regresi yang dikali dengan nilai  $x$  peramal.

**Probit:**

Model probit diperkenalkan oleh Bliss Chester pada tahun 1934 [9]. Model probit adalah sejenis regresi di mana variabel terikat hanya boleh mengambil dua nilai, contohnya bekerja atau tidak bekerja. Asal nama probit ialah unit+probabilitas. Model probit adalah data spesifikasi yang populer untuk data ordinal atau model *binary response variable* yang menggunakan penghubung fungsi probit. Ia menyelesaikan masalah sama seperti regresi logistik dengan menggunakan teknik yang sama. Model probit adalah yang paling sering diestimasi menggunakan metode maksimum likelihood yang standard, seperti parameter yang dipanggil regresi probit. Dia merupakan metode yang cepat untuk menentukan parameter maksimum likelihood

$$Y = \begin{cases} 1 & \text{if } Y^* > 0 \text{ i. e. } -\varepsilon < X'\beta \\ 0 & \text{selain} \end{cases}$$

$$\begin{aligned} \Pr(Y = 1|X) &= \Pr(Y^* > 0) = \Pr(X'\beta + \varepsilon > 0) \\ &= \Pr(\varepsilon > -X'\beta) \\ &= \Pr(\varepsilon < X'\beta) \quad (\text{by symmetry of the normal distribution}) \\ &= \Phi(X'\beta) \end{aligned}$$

**Pelengkap fungsi log-log (complementary log-log):**

Model ini kelihatan sama seperti fungsi bagi model logit dan probit untuk nilai  $\pi$  berhampiran dengan 0.5 tetapi perbedaannya apabila  $\pi$  menghampiri 0 dan 1. Rumus yang digunakan dalam pelengkap fungsi log-log ialah :

$$\log[-\log(1-\pi)] = \beta_1 + \beta_2 x + \beta_3 x_2$$

**3. Hasil dan Pembahasan**

Untuk menguji model yang terbaik, kita lakukan analisis dengan menggunakan dua metode yaitu metode Logistic Proc dan Genmod Proc. Berikut ini ialah hasil dari software SAS untuk kedua metode tersebut:

Tabel 1. Metode Logistic Proc

Model	Intercept	logx1	logx2	logx3	AIC With Intercept and Covariantes
Logit	-5.2694	-0.2500	0.5176	0.1531	61,961,003
Probit	-2.9480	-0.1546	0.3009	0.0922	61,952,208
CLL	-2.8205	-0.1476	0.2685	0.0847	61,939,954

Tabel 2. Metode Genmod Proc

Model	Intercept	logx1	logx2	logx3	Deviance	DF
Logit	-9.0533	0.2529	3.2141	0.0888	82.0521	35
Probit	-5.2312	0.1296	1.8726	0.0564	79.5857	35
CLL	-4.8519	0.0989	1.6489	0.0519	76.3072	35

Dari Tabel 1 dan Tabel 2, kita dapati bahwa metode Logistic proc, complementary log log (CLL) mempunyai AIC=61,939,954 yang paling kecil berbanding model logit dan probit. Berdasarkan hasil dari Metode Genmod proc, CLL juga mempunyai nilai Deviance= 76.3072 yang mana lebih kecil daripada model probit dan logit. Jadi model yang terbaik untuk data kelapa sawit ini ialah dengan menggunakan model CLL.

Dalam perolehan nilai devians kelihatan bahwa model ini mengalami kelebihan sebaran (*overdispersion*) kerana nilai min devians melebihi satu. Ketidakbagusan ini disebabkan ada beberapa data pencilan atau mungkin struktur hubungan prediktor kurang tepat. Untuk mengujinya, kita perlu lakukan penambahbaikan (*adjustment*).

Tabel 3. Model CLL yang dihasilkan oleh Genmod proc dengan analisis maksimum likelihood

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.8519	0.9052	-6.6260	-3.0778	28.73	<.0001
logx1	1	0.0989	0.0493	0.0024	0.1954	4.03	0.0447
logx2	1	1.6489	0.2682	1.1232	2.1746	37.79	<.0001
logx3	1	0.0519	0.0617	-0.0690	0.1728	0.71	0.4001
Scale	0	1.0000	0.0000	1.0000	1.0000		

Uji signifikan

$$\text{Adjustment} = \sqrt{2.1802} = 1.48$$

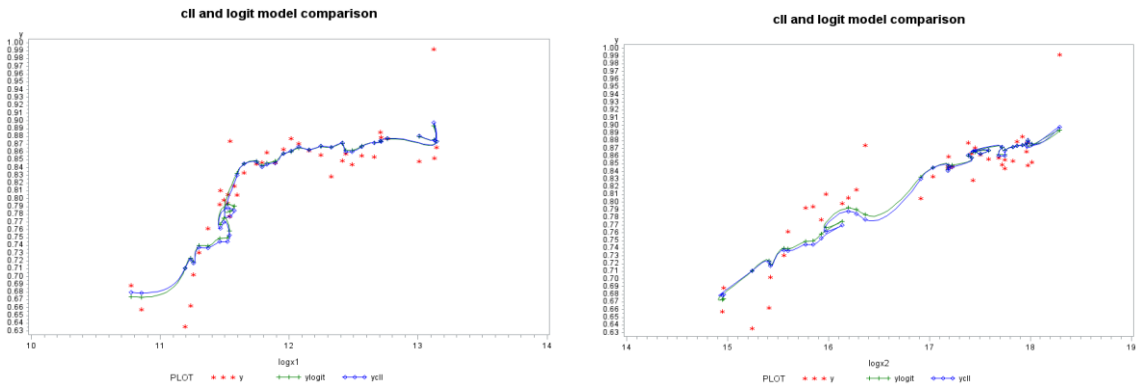
$$\log x1 : F_{3,35} = \left[ \frac{0.0989}{0.0493 \times 1.48} \right]^2 = 1.3555^2 = 1.8373$$

$$\log x2 : F_{3,35} = \left[ \frac{1.6489}{0.2682 \times 1.48} \right]^2 = 4.151^2 = 17.23$$

$$\log x3 : F_{3,35} = \left[ \frac{0.0519}{0.0493 \times 1.48} \right]^2 = 0.3230$$

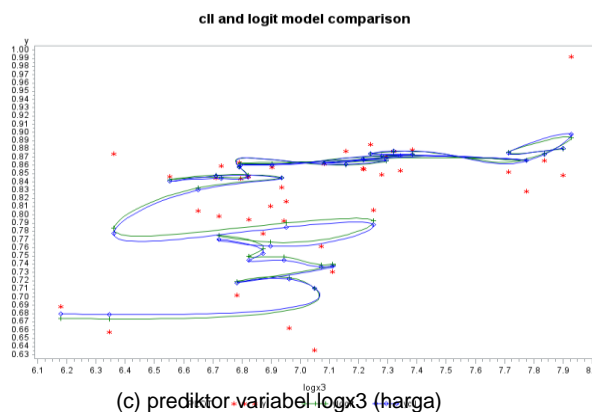
Nilai  $F_{0.05,3,35} = 2.88$

Maka, berdasarkan hasil pengubahbaikan (*adjustment*) hanya terdapat variabel logx2 yang signifikan, dan dapat digunakan di dalam model. Sekiranya kita mempertimbangkan nilai  $p$  terlihatkan logx1 turut signifikan dan dapat digunakan dalam model kerana  $p = 0.0447 < \alpha = 0.05$ . Selanjutnya kita mencoba menganalisis variabel dari metode logistic proc. Dapat kita lihat dari gambar 1 berikut:



(a) prediktor variabel logx1 (pekerja)

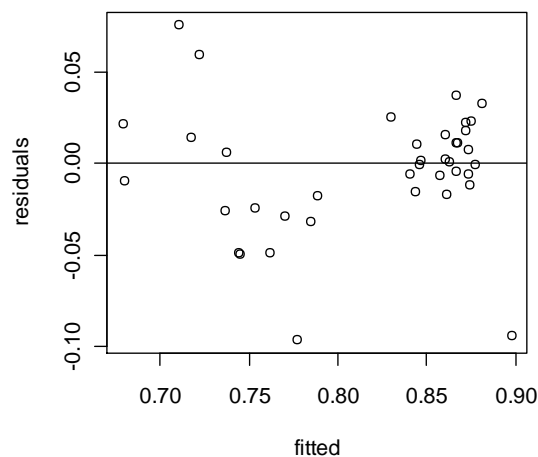
(b) prediktor variabel logx2 (hasil buah)



(c) prediktor variabel logx3 (harga)

Gambar 1. Predictor masing-masing variabel

Prodicator yang mengikuti pola alur variabel hanya pada logx1 (pekerja) dan logx2 (hasil buah), sedangkan pada variabel logx3 (harga) terlihat alur predictor yang tidak beraturan. Hal ini menambah keyakinan kita untuk mengatakan bahwa variabel logx3 memang tidak berpengaruh terhadap model.



Gambar 2. Grafik eror untuk model CLL

#### 4. Kesimpulan

Setelah menganalisis ketiga- model yaitu model logit, probit dan CLL, dengan metode logistic proc dan gonmod proc dapat disimpulkan bahwa model CLL lebih baik dibandingkan model logit dan model probit. Sementara untuk variabel yang berpengaruh pada model ialah x1 (pekerja) dan x2 (hasil buah) sedangkan untuk variabel x3 (harga) tidak begitu mempengaruhi terhadap model. Hal ini sesuai dengan realita sebenarnya bahwa pekerja yang banyak akan mampu manggarap (memanen) kawasan tanaman sawit dengan maksimal dan buah yang dipanen juga turut meningkat, sedangkan harga buah tidak mempengaruhi luas kawasan yang dipanen ataupun tidak saling mempengaruhi terhadap variabel yang lain. Sebaliknya, bertolak belakang dengan variabel pekerja dan hasil buah. Sebab, semakin banyak buah yang sawit maka harga pasar untuk buah sawit justru menurun. Oleh itu, metode logistik mampu menjelaskan realita tentang kawasan sawit dan variabel yang berkenaan sesuai dengan fakta sebenarnya secara statistik.

#### Referensi

- [1] Fauziah Sulaiman. 2010. A Perspective of Oil Palm and Its Wastes. *Journal of Physical Science*, Vol. 21(1), 67–77.
- [2] Azme Khamis, dkk. 2006. Modeling Oil Palm Yield Using Multiple Linear Regression and Robust M-regression. *Journal of Agronomy*, 5: 32-36.
- [3] Draper, N.R. and H.Smith, 1981. *Applied Regression Analysis Edn 2<sup>nd</sup>*. New York. John Wiley and Sons.
- [4] Birkes, D. & Y. Dodge, 1993. *Alterbative Methods of Regression*. N.Y. John Wiley and Sons
- [5] McCullagh, Peter; Nelder, John. 1989. *Generalized Linear Models, Second Edition*. USA. Chapman & Hall/CRC.
- [6] Cyrus R. Mehta & Nitin R. Patel. 1995. Exact Logistic regression : Theory and examples. *Journal Statistics in Medicine*, Vol 14, 2143-2160.
- [7] MPOB(2010). *Overview of the Malaysian Oil Palm Industry 2009*. MPOB, Bangi.
- [8] Hosmer, David W. Lemeshow, Stanley. 2000. *Applied Logistic Regression (2nd ed.)*. UK. Wiley.
- [9] Bliss, C.I. (1935). The calculation of the dosage-mortality curve. *Annals of Applied Biology* (22)134–167.