

PEMODELAN KARAKTERISTIK TINGKAT PENDIDIKAN ANAK DI PROVINSI JAWA BARAT MENGGUNAKAN *LOG LINEAR*

Resa Septiani Pontoh¹⁾, Neneng Sunengsih²⁾

^{1),2)}Departemen Statistika Universitas Padjadjaran

¹⁾resa.septiani@unpad.ac.id, ²⁾nenks.stat@gmail.com

ABSTRAK. Pendidikan adalah hak warga Negara yang diatur oleh undang-undang. Evaluasi mengenai capaian tingkat pendidikan pun terus dilakukan. Pemodelan yang dapat memberikan gambaran apakah faktor pendidikan orang tua (ayah dan ibu), pemilihan lokasi sekolah, jenis kelamin dan lokasi tempat tinggal mempengaruhi tingkat pendidikan seseorang perlu dilakukan. Jika kelima hal tersebut berdampak positif, maka tentunya faktor-faktor tersebut dapat dijadikan salah satu acuan untuk dapat meningkatkan motivasi warga Negara Indonesia dalam mengikuti pendidikan formal. Penelitian ini akan melihat faktor atau karakteristik apa saja yang mempunyai pengaruh terhadap tingkat pendidikan seorang anak di Provinsi Jawa Barat. Data yang digunakan adalah sebanyak 2010 responden masyarakat di wilayah ini. Jenis data tersebut bersifat kualitatif sehingga analisis regresi sederhana menjadi kurang tepat. Model log linear dapat menjadi salah satu alternatif solusi karena model ini menunjukkan ada tidaknya hubungan antar beberapa variabel pada tabel kontingensi banyak arah. Analisis ini memungkinkan dilakukannya modifikasi ada atau tidaknya interaksi antar dua atau lebih variabel. Hasil dari penelitian ini adalah bahwa tingkat pendidikan orang tua dan lokasi sekolah memberikan kontribusi positif terhadap tingkat pendidikan masyarakat di Provinsi Jawa Barat. Seseorang yang bersekolah di daerah kota, begitu pula orang tua yang berpendidikan lebih tinggi, cenderung memiliki tingkat pendidikan yang lebih tinggi.

Kata Kunci: *log linear model; Tingkat Pendidikan Provinsi Jawa Barat; Data Kategori*

1. PENDAHULUAN

Menurut UNDP tahun 2015 Indeks Pembangunan Manusia Indonesia menempati urutan ke 110 dari 183 negara di dunia. Seperti yang kita ketahui, pendidikan merupakan salah satu faktor yg menentukan Indeks Pembangunan Manusia. Meskipun pendidikan merupakan hak bagi semua warga negara yang dilindungi oleh UU, tetapi kenyataannya belum dapat dinikmati oleh semua warga masyarakat. Hal ini dapat dilihat berdasarkan APS di setiap provinsi di Indonesia, melalui APS dapat diketahui berapa persen dari penduduk yang tidak menikmati pendidikan sesuai dengan jenjang usianya. Provinsi Jawa Barat merupakan provinsi yang memiliki jumlah penduduk terbanyak, penduduknya kebanyakan laki - laki dan 26% dari penduduk Jabar tinggal di daerah penyangga ibu kota. Menurut BPS(2014) penduduk yang tidak bersekolah sebanyak 0.7 % berusia 7 sd 12 tahun, 7.16% berusia 13 sd 15 th, 65.4 % berusia 16 sd 18 th. Berdasarkan data di atas, dilakukan penelitian di propinsi Jawa Barat untuk mengkaji karakteristik apa saja yang mempengaruhi sekolah anak. Variabel yang diteliti dalam penelitian ini adalah, pendidikan orang tua, lokasi sekolah dan gender yang disajikan dalam tabel kontingensi untuk dianalisis variabel mana saja yang signifikan.

Model regresi linear adalah model yang paling umum untuk menggambarkan hubungan antara dua atau lebih variabel. Namun, model ini memiliki beberapa kekurangan yaitu diantaranya adalah adanya asumsi linearitas, adanya asumsi normalitas,

serta adanya asumsi *equidistance* untuk variabel bebas dan tak bebasnya (Agresti, [2]). Namun, ketika data yang digunakan bersifat data kategori, maka model regresi linear menjadi tidak tepat untuk digunakan. Untuk menyelesaikan persoalan ini, banyak metode yang dapat digunakan, salah satunya adalah Log Linear model.

Model log linear adalah bagian dari Generalized Linear Model. Model ini sering digunakan untuk memodelkan frekuensi atau banyaknya kejadian pada beberapa sel dari suatu tabel kontingensi. Komponen sistematis yang ada pada model ini menjelaskan bagaimana keragaman nilai ekspektasi dari frekuensi sel yang ada sebagai akibat dari variabel bebasnya. Perlu diketahui bahwa model log linear tidak membedakan antara variabel bebas dan tak bebas dikarenakan variabel respon atau tak bebas dari model ini adalah frekuensi dari sel yang ada. Hasil dari analisis ini adalah untuk melihat pola hubungan sekelompok variabel (kategori) yang terlihat pada interaksi yang dituangkan dalam model. Oleh karena itu, penelitian ini akan memodelkan karakteristik tingkat pendidikan anak di Jawa Barat dengan menggunakan Log Linear.

2. METODE PENELITIAN

Penelitian ini bersifat kuantitatif, dengan menggunakan data sekunder tingkat pendidikan masyarakat Provinsi Jawa Barat untuk melihat karakteristik tingkat pendidikan anak di wilayah tersebut berdasarkan jenis kelamin, pendidikan ibu, pendidikan ayah dan lokasi sekolah.

Seperti yang telah dijelaskan pada bagian sebelumnya, model log linear akan mengacu pada tabel kontingensi yang dibuat. Terdapat tiga model pengambilan sampel untuk analisis ini yaitu (Agung [2]):

1. Poisson apabila pengambilan sampel berdasarkan pengamatan pada interval waktu tertentu dengan mengasumsikan n_i bersifat random dan independen,
2. Multinomial apabila ukuran sampel telah ditentukan,
3. *Product* multinomial ketika total marjinalnya telah ditentukan.

Penelitian ini menggunakan pengambilan sampel multinomial karena ukuran pada prosesnya ukuran sampel memang telah ditentukan terlebih dahulu kemudian di sebar ke dalam sel-selnya sesuai dengan observasi yang telah dilakukan.

A. Model Log Linear

Berikut ini akan dijelaskan beberapa contoh model log linier yang biasa digunakan. Jika tidak ada variabel yang mempengaruhi model (*zero order*), model log linier (Von Eye, [4]) secara umum adalah sebagai berikut:

$$\text{Log}E(Y) = \mu$$

dimana $E(Y)$ = frekuensi diharapkan setiap sel.

$$\mu = \text{intercept atau constant atau rata-rata umum}$$

Jika semua variabel mempunyai status yang sama, dan hanya *main effect* yang digunakan (*first order*), model log linier (Von Eye, 2002) secara umum adalah sebagai berikut:

$$\text{Log}E(Y_{ij\dots}) = \mu + A_i + B_j + \dots$$

dimana $E(Y_{ijk})$ = frekuensi diharapkan setiap sel.

- μ = *intercept* atau *constant* atau rata-rata umum
 A_i = Efek utama faktor A pada kategori ke-i
 B_j = Efek utama faktor B pada kategori ke-j

Jika variabel yang akan diteliti terjadi interaksi, misal interaksi A dan B dan C dan D maka, model log linier (Von Eye, 2002) yang digunakan adalah sebagai berikut:

$$\text{Log}E(Y_{ijkl}) = \mu + A_i + B_j + AB_{ij} + C_k + D_l + CD_{kl}$$

dimana $E(Y_{ijk})$ = frekuensi diharapkan setiap sel.

- μ = *intercept* atau *constant* atau rata-rata umum
 A_i = Efek utama faktor A pada kategori ke-i
 B_j = Efek utama faktor B pada kategori ke-j
 C_k = Efek utama faktor C pada kategori ke-k
 D_l = Efek utama faktor D pada kategori ke-l
 AB_{ij} = Interaksi faktor A dan B pada kategori ke-i dan ke-j
 CD_{kl} = Interaksi faktor C dan D pada kategori ke-k dan ke-l

Pada penelitian ini akan melihat interaksi dua arah antara tingkat pendidikan anak dengan jenis kelamin, pendidikan ibu, pendidikan ayah, serta lokasi sekolah yaitu sebagai berikut:

$$\text{Log} E(Y_{ijklm}) = \mu + A_i + B_j + C_k + D_l + E_m + AB_{ij} + AC_{ik} + AD_{il} + AE_{im} \quad (2.1)$$

dimana $E(Y_{ijk})$ = frekuensi diharapkan setiap sel.

- μ = *intercept* atau *constant* atau rata-rata umum
 A_i = Efek utama Faktor Tingkat pendidikan anak pada kategori ke-i
 B_j = Efek utama Faktor Jenis kelamin pada kategori ke-j
 C_k = Efek utama Faktor Pendidikan Ibu pada kategori ke-k
 D_l = Efek utama Faktor Pendidikan Ayah pada kategori ke-l
 E_m = Efek utama Faktor Lokasi Sekolah pada kategori ke-m
 AB_{ij} = Interaksi faktor Tingkat pendidikan anak dan Jenis kelamin pada kategori ke-i dan ke-j
 AC_{ik} = Interaksi Faktor Tingkat pendidikan dan Pendidikan Ibu pada kategori ke-i dan ke-k
 AD_{il} = Interaksi Faktor Tingkat pendidikan dan Pendidikan Ayah pada kategori ke-i dan ke-k
 AE_{im} = Interaksi Faktor Tingkat pendidikan dan Pendidikan Ayah pada kategori ke-i dan ke-m

B. Estimasi Frekuensi Ekspektasi

Setelah meletakkan frekuensi observasi pada tabel kontingensi, maka langkah selanjutnya adalah melakukan estimasi terhadap ekspektasi frekuensi konfigurasi ($E(Y_{ijkl})$) digunakan dengan menggunakan metode *Maksimum Likelihood*.

Fungsi dari distribusi multinomial dengan frekuensi sel Y_1, \dots, Y_N dengan peluang tiap sel adalah $\theta_1, \dots, \theta_N$ dengan nilai n yang telah ditentukan sebelumnya (Dobson, [3]) adalah sebagai berikut:

$$f(y; \theta | n) = n! \prod_{i=1}^N \frac{\theta_i^{y_i}}{y_i!}$$

Dimana $n = \sum_{i=1}^N y_i$ dan $\sum_{i=1}^N \theta_i = 1$

Untuk fungsi Log Likelihood dari distribusi multinomial yaitu:

$$L = \log n! + \sum_{i=1}^N (y_i \log \theta_i - \log y_i!) \quad \dots(2.2)$$

yang mana dapat ditulis dalam bentuk:

$$L = \text{constant} + \sum_{i=1}^N y_i \log \theta_i$$

Untuk memperoleh penaksir *Maximum Likelihood* dari parameter θ_i diperoleh dengan memaksimalkan fungsi *Likelihood* dengan *constraint* yang dapat dilakukan dengan menggunakan *Lagrange Multiplier* yang akan meminimalkan ξ dan θ_i dari persamaan berikut:

$$t = \text{constant} + \sum_{i=1}^N y_i \log \theta_i - \xi \left(\sum \theta_i - 1 \right)$$

Dengan $\frac{\partial t}{\partial \xi} = 0$ dan $\frac{\partial t}{\partial \theta_i} = 0$ untuk $i = 1, \dots, N$ sehingga diperoleh:

$$\xi = n$$

Kemudian substitusikan $\xi = n$ ke dalam persamaan sebelumnya, maka diperoleh:

$$t = \text{constant} + \sum_{i=1}^N y_i \log \theta_i - n \left(\sum \theta_i - 1 \right)$$

Dari persamaan diatas diperoleh *Maximum Likelihood Estimator* dari parameter θ_i untuk $E(Y_i)$ yaitu:

$$\frac{\partial t}{\partial \theta_i} = \frac{Y_i}{\theta_i} - n = 0$$

$$n\theta_i = Y_i$$

C. Uji Kecocokan Model

Langkah selanjutnya adalah membandingkan frekuensi observasi dan ekspektasi sehingga terlihat apakah model log linear yang terbentuk sesuai dengan data atau fakta yang ada. Adapun rumusan hipotesisnya adalah sebagai berikut:

H_0 : Model Log Linear yang digunakan cocok dengan keadaan sebenarnya

H_1 : Model Log Linear yang terbentuk tidak cocok dengan keadaan yang sebenarnya

Kemudian dilakukan uji chi-kuadrat dengan statistik uji sebagai berikut:

$$\chi^2 = \sum_i \sum_j \sum_k \sum_l \frac{(n_{ijkl} - e_{ijkl})^2}{e_{ijkl}}$$

Kriteria uji:

Tolak H_0 jika $\chi^2_{\text{hitung}} \geq \chi^2_{\alpha}$ atau $p\text{-value} < \alpha$

Terima H_0 jika $\chi^2_{hitung} < \chi^2_{\alpha}$ atau $p\text{-value} \geq \alpha$

3. HASIL PENELITIAN DAN PEMBAHASAN

Berdasarkan pengolahan data yang telah dilakukan, diperoleh hasil pemodelan main effect seperti pada tabel 3.1.

Tabel 3.1 *Goodness of fit* untuk *Main effect* model

	Value	df	Sig.
Likelihood Ratio	791.050	26	.000
Pearson Chi-Square	1488.853	26	.000

Berdasarkan Tabel 3.1 terlihat bahwa model telah fit dengan data. Kemudian dilakukan uji individual seperti terlihat pada Tabel 3.2.

Tabel 3.2 Estimasi Parameter model *Main Effect*

Parameter	Estimate	Std. Error	Z	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Constant	.787 ^a					
[P_Anak = 1]	.465	.063	7.340	.000	.341	.590
[P_Anak = 2]	0 ^b
[JK = 1]	.149	.062	2.405	.016	.028	.270
[JK = 2]	0 ^b
[P_Ibu = 1]	1.922	.092	20.792	.000	1.741	2.103
[P_Ibu = 2]	0 ^b
[P_Ayah = 1]	1.163	.072	16.053	.000	1.021	1.305
[P_Ayah = 2]	0 ^b
[Lokasi = 1]	.465	.063	7.340	.000	.341	.590
[Lokasi = 2]	0 ^b

Tabel 3.2 menunjukkan semua variabel signifikan terhadap model yang ada. Selanjutnya akan dilakukan uji kecocokan model untuk interaksi dua faktor. Pada proses pengolahan data untuk Persamaan (2.1), interaksi antara jenis kelamin dan pendidikan anak tidak signifikan, sehingga persamaan menjadi sebagai berikut:

$$\text{Log } E(Y_{ijklm}) = \mu + A_i + B_j + C_k + D_l + E_m + AC_{ik} + AD_{il} + AE_{im}.$$

Hasil pengolahan data dijelaskan pada Tabel 3.3.

Tabel 3.3 Uji Kecocokan Model Dua Faktor

	Value	df	Sig.
Likelihood Ratio	377.331	23	.000
Pearson Chi-Square	605.364	23	.000

Tabel 3.3 menjelaskan bahwa model telah cocok dengan data yang ada. Penaksiran parameter untuk model ini tercantum pada Tabel 3.4.

Tabel 3.4 Penaksiran Parameter untuk Interaksi Dua Faktor

Parameter	Estimate	Z	Sig.
Constant	1.237 ^a		
[P_Anak = 1]	-1.533	-5.581	.000
[P_Anak = 2]	0 ^b	.	.
[JK = 1]	.149	2.405	.016
[JK = 2]	0 ^b	.	.
[P_Ibu = 1]	1.037	9.174	.000
[P_Ibu = 2]	0 ^b	.	.
[P_Ayah = 1]	.223	2.231	.026
[P_Ayah = 2]	0 ^b	.	.
[Lokasi = 1]	1.673	12.281	.000
[Lokasi = 2]	0 ^b	.	.
[P_Anak = 1] * [P_Ibu = 1]	2.056	9.184	.000
[P_Anak = 1] * [P_Ibu = 2]	0 ^b	.	.
[P_Anak = 2] * [P_Ibu = 1]	0 ^b	.	.
[P_Anak = 2] * [P_Ibu = 2]	0 ^b	.	.
[P_Anak = 1] * [P_Ayah = 1]	1.883	11.671	.000
[P_Anak = 1] * [P_Ayah = 2]	0 ^b	.	.
[P_Anak = 2] * [P_Ayah = 1]	0 ^b	.	.
[P_Anak = 2] * [P_Ayah = 2]	0 ^b	.	.
[P_Anak = 1] * [Lokasi = 1]	-1.788	-11.358	.000
[P_Anak = 1] * [Lokasi = 2]	0 ^b	.	.
[P_Anak = 2] * [Lokasi = 1]	0 ^b	.	.
[P_Anak = 2] * [Lokasi = 2]	0 ^b	.	.

Dari Tabel 3.4, semua variabel signifikan terhadap model. Dapat dijelaskan berdasarkan nilai odd ratio bahwa:

1. Anak yang tidak tamat SMA dengan kondisi Ibu yang tidak tamat SMA pula mempunyai kecenderungan 7.81 kali untuk terjadi dibandingkan dengan kondisi lainnya.
2. Anak yang tidak tamat SMA dengan kondisi Ibu yang tidak tamat SMA pula mempunyai kecenderungan 6.25 kali untuk terjadi dibandingkan dengan kondisi lainnya.
3. Anak yang tidak tamat SMA dengan kondisi bersekolah di kota mempunyai kecenderungan tidak terjadi sebesar 83,27% dibandingkan dengan kondisi lainnya.

4. SIMPULAN

Berdasarkan analisis yang telah dilakukan, dapat disimpulkan bahwa terdapat tiga variabel yang berinteraksi dengan tingkat pendidikan anak di Propinsi Jawa Barat yaitu Pendidikan Ibu, Pendidikan Ayah, dan Lokasi Sekolah. Terlihat adanya kecenderungan bahwa semakin tinggi tingkat pendidikan ayah dan ibu maka semakin tinggi pula kecenderungan seorang anak berpendidikan tinggi. Selain itu, dapat disimpulkan bahwa lokasi anak bersekolah di kota cenderung berpengaruh terhadap tamatnya seorang anak dari tingkatan SMA.

DAFTAR PUSTAKA

- [1] Agresti, A. 2007. *An Introductory to Categorical Analysis*. John Willey & Sons, Inc. New York.
- [2] Agung, I Gusti Ngurah. 2002. *Analisis Hubungan Kausal Berdasarkan Data Kategori*. PT Raja Grafindo Persada, Jakarta.
- [3] Dobson, A. J. 1983. *Introduction to Statistical Modelling*. Chapman and Hall Ltd: London
- [4] Von Eye, A. 2002. *Configural Frequency Analysis: Methods, Models, and Applications*. Lawrence Erlbaum Associates: London.