

ANALISIS REGRESI LOGISTIK EKSAK PADA PENANGANAN SAMPEL KECIL

Herlina Jusuf

Staf Dosen Jurusan Kesmas FIKK
Universitas Negeri nGorontalo

Abstract

Correlation between a categorical response variable and one or several predictor variables involving larger samples is analyzed using logistic regression method. However, if the sample size is small and the data are sparse, the relevance of conventional (asymptotic) logistic regression method to use in such correlation analysis will be questioned.

This is a statistical study for application of exact logistic regression in the conditional of small sample size and sparse data.

Exact logistic regression analysis was done to cases with sample size of 10, 20, 29 and 55 taken from results of random sampling data. obtained by immunization. The dependent variable was immunization status, and he independent variable was people's exposure to information, education, occupation, living children, knowledge, attitude and participation. The data were analyzed with logxact Turbo program.

Result showed that parameter estimation and hypothesis test using exact test provided better solution compared to conventional (asymptotic) logistic regression test such as likelihood ratio, Wald, and Score test. Exact test also provided correlation type or model with sample size of 10, 20, 29 and 55 and wider confidence interval compared to asymptotic inference type. Probability between those intervals had, therefore, larger parameter of population. For test involving numerous independent variables, exact logistic regression method also provided better solution compared to there conventional (asymptotic) logistic regression test. Exact logistic regression method should be used in statistical test with small sample size and spare data and tests that use table with sparse value in each of its cells

Keywords : *Exact logistic regression, asymptotic*

PENDAHULUAN

Metode regresi merupakan komponen penting dalam data analisis untuk menggambarkan hubungan antara suatu variabel respon (*Response Variabel*) dengan satu atau beberapa variabel predictor (*eksplanatory variables*) (Hosmer, 1989). Pola hubungan antara variabel respon dengan satu atau lebih variabel predictor dapat menggunakan regresi linier berganda dimana variabel prediktornya memiliki skala pengukuran interval atau rasio, tetapi jika variabel respon berskala pengukuran nominal atau berupa data biner (bernilai 0 atau 1) maka model pengguna yang digunakan adalah regresi logistic (Weisberg,1985).

Metode regresi logistic memiliki tehnik dan procedure analisis yang tidak jauh berbeda dengan metode regresi linier berganda. Jika procedure regresi linier dalam mengestimasi nilai parameter sering menggunakan pendekatan dengan metode OLS (*Ordinary Least Squares*) maka dalam regresi logistic dalam mengestimasi nilai parameter menggunakan MLE (*Maximum Likelihood Estimation*).

Model Regresi Logistik

Bentuk model regresi logistic secara umum (Hosmer,1989, Agresti, 1990) :

$$\pi(x) = \frac{\exp(\sum_{j=0}^p \beta_j x_j)}{1 + \exp(\sum_{j=0}^p \beta_j x_j)}, \quad \text{dengan } p = \text{banyaknya predictor}$$

Untuk mempermudah penaksir parameter regresinya, maka $\pi(x)$ di transformasikan dengan transformasi logit (Weisberg,1985, Hosmer 1989) sehingga dapat dituliskan sebagai berikut : $g(x) = \ln \left[\frac{\pi(x)}{1-\pi(x)} \right] = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p = \sum_{j=1}^p \beta_j x_j$

dimana $g(x)$ merupakan fungsi linier dari parameter-parameternya.

Pada regresi logistic, pola distribusi bersyarat variabel responnya adalah

$Y = \pi(x) + \varepsilon$ dimana nilai error ε mempunyai dua kemungkinan nilai yaitu :

$Y = 1$, maka $\varepsilon = 1 - \pi(x)$, dengan peluang $\pi(x)$

$Y = 0$, maka $\varepsilon = -\pi(x)$, dengan peluang $1 - \pi(x)$ dan dapat dinyatakan bahwa ε memiliki $\sum \varepsilon = 0$ dan $\text{var}(\varepsilon) = \pi(x)[1 - \pi(x)]$ yang mengikuti distribusi binomial (Hosmer, 1989) karena banyaknya kejadian/pengamatan yang dilakukan adalah n bukan individual.

Penaksir Parameter

Penaksir parameter pada model regresi logistic yang mempunyai variabel respon dikotom adalah menggunakan metode *Maksimum Likelihood Estimation (MLE)*. Pada dasarnya metode MLE menetapkan asumsi distribusi Bernoulli dan obyek pengamatan yang saling bebas, $\sum \varepsilon_i, \varepsilon_j = 0, i \neq j$, atau memberikan nilai taksiran parameter dengan memaksimumkan fungsi *likelihood*.

Bentuk fungsi *likelihood* adalah : $\zeta(x_i) = \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i}$

Karena setiap pengamatan diasumsikan bersifat independen, maka *likelihood* pengamatan merupakan perkalian dari fungsi *likelihood* masing-masing. Dan untuk memaksimumkan fungsi *likelihood* maka fungsi diubah kedalam bentuk log *likelihood*. Untuk mendapatkan nilai β dari $L(\beta)$ yang maksimum, diperoleh melalui turunan parsial yang hasilnya dinyatakan sama dengan nol diperoleh sebagai berikut :

$$\sum_i y_i x_{ia} - \sum_i n_i x_{ia} \pi_i = 0 ; \quad a = 0, 1, 2, \dots, p \text{ parameter}$$

Uji Kesesuaian Model Regresi Logistik

Beberapa statistik uji yang dapat digunakan untuk menguji kesesuaian model regresi logistic adalah :

- $-2 \text{ Log likelihood} : G^2 = 2 \sum_i \sum_j O_{ij} \log \left[\frac{O_{ij}}{E_{ij}} \right]$
- Goodness of Fit* : $\sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$
- $\chi^2 = -2 \ln \left[\frac{\text{Likelihood tanpa variabel}}{\text{Likelihood dengan variabel}} \right]$

Spesifikasi Model Regresi Logistik Eksak

Secara umum regresi logistic dapat digunakan untuk menganalisis dua model data yaitu regresi logistic untuk *unstratified data* dan *Stratified data*. (Mehta, 1992). Model *unstratified* dan *Stratified* logistic ditulis dalam bentuk :

$$\text{Log} \frac{\pi_{ij}}{1-\pi_{ij}} = \beta X_{ij} + \gamma_i \quad \text{dan} \quad \text{Log} \frac{\pi_{ij}}{1-\pi_{ij}} = \beta X_j + \gamma$$

Dimana γ dan $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ adalah parameter yang tidak diketahui. Pada beberapa literature statistik, symbol γ diartikan sebagai konstanta atau intersep.

Sampel Kecil

Sampel adalah sebagian dari populasi (Kendall, 1971). Sampel merupakan sejumlah cuplikan tertentu yang diambil dari suatu populasi dan diteliti secara rinci atau semacam miniature (mikrokosmos) dari populasinya (santoso, 2001). Sampel selalu mempunyai ukuran yang kecil jika dibandingkan dengan populasi yang merupakan himpunan yang lengkap dari satuan-satuan atau individu-individu yang karakteristiknya ingin diketahui (Small htm, 2002).

Besar atau ukuran sampel memang satu hal penting yang harus diperhatikan dalam suatu penelitian. Cochran (1991) hal yang harus diperhatikan dalam pemilihan ukuran sampel diantaranya adalah harus ada beberapa pernyataan mengenai apa yang diharapkan dari sampel. Pernyataan ini berhubungan dengan batas kesalahan yang diinginkan, atau sehubungan dengan beberapa keputusan yang dibuat atau diambil bila hasil sampel diketahui.

Hakikat dari sampel itu sendiri (misalnya sampel orang-orang yang mempunyai sesuatu penyakit jiwa, atau sampel-sampel kebudayaan) maka mengharuskan peneliti mengumpulkan data

dalam jumlah kecil, dalam hal ini uji nonparametric akan sangat berguna, dimana ukuran sampel kecil tersebut adalah kurang dari 25 (Siegel,1988), Scheafer (1986) Sampel $(n) \geq 30$ sudah dikatakan besar dan memiliki distribusi yang normal. *Small Sampel for Population Mean* (2001) juga menyebutkan bahwa sampel kecil adalah kurang dari 30. Merujuk pada table kontingensi yang sama jarangnya ketika banyak sel mempunyai frekuensi kecil, jika N adalah jumlah sel dan n adalah ukuran sampel total, n/N adalah indeks kejarangan (*sparness*), nilai yang lebih kecil merujuk pada table yang lebih jarang, maka table *sparse* terjadi ketika ukuran sampel kecil. Tabel ini juga muncul ketika ukuran sampel besar tapi jumlah selnya juga banyak dan sering muncul jika ada banyak variabel, atau sebagian klasifikasi memiliki beberapa kategori (Agresti, 1990).

METODE PENELITIAN

Penelitian ini merupakan penelitian terapan (*applied research*), dimana teori statistika diterapkan dalam bidang kesehatan (Zainuddin, 2000) dengan menggunakan data sekunder sebagai bahan analisisnya (*secondary data analysis*). Sebagai penekanan utama dalam penelitian ini adalah melihat penggunaan metode regresi logistic eksak untuk penanganan kasus-kasus sampel kecil dan sampel yang mengandung data missing (*sparse data*) dan penggunaan metode regresi logistic konvensional (asimptotik) atau metode regresi logistic biasa. Adapun data sekunder yang digunakan adalah data dari hasil penelitian mengenai pengetahuan dan sikap terhadap pelaksanaan program imunisasi. Dengan objek penelitiannya adalah ibu yang mempunyai anak balita berumur dibawah dua tahun (*baduta*).

Variabel Penelitian

Dalam penelitian ini sebagai Variabel respon Y dalam hal ini status imunisasi anak baduta dengan kategori :

Y = 0 jika status imunisasi baduta (> 1 tahun) tidak lengkap

Y = 1 jika status imunisasi baduta (> 1 tahun) lengkap.

Adapun Variabel bebas (prediktor) yang digunakan dalam penelitian ini adalah factor pengetahuan dan sikap ibu baduta terhadap status imunisasi antara lain :

1. Keterpaparan terhadap informasi immunisasi (X_1) dengan kategori :
0 = tidak dapat informasi 1 = dapat informasi
2. Pendidikan responden (X_2) jenjang pendidikan terakhir ibu baduta dengan kategori :
0 = Pendidikan dasar 1 = Pendidikan $>$ tamat SLTP
3. Pekerjaan (X_3) adalah pekerjaan ibu baduta dengan kategori :
0 = Tidak bekerja 1 = Bekerja
4. Jumlah anak (X_4) adalah jumlah anak yang hidup yang dimiliki ibu baduta :
0 = Kecil (< 3 orang) 1 = Besar ≥ 3 orang)
5. Tingkat pengetahuan ibu baduta tentang imunisasi (X_5) dengan kategori :
0 = kurang 1 = baik
6. Sikap responden atau persepsi responden terhadap imunisasi (X_5) dengan kategori :
0 = Kurang mendukung 1 = Mendukung
7. Peran serta masyarakat (X_6) dalam program imunisasi dengan kategori :
0 = tidak ada 1 = ada

Tehnik analisis Data

Tehnik analisis data digunakan adalah regresi logistic (asimptotik) dan regresi logistic eksak dengan paket program Logxact-Turbo dengan langkah :

1. Melihat atau menentukan semua kemungkinan yang mungkin muncul dari populasi variabel independen atau predictor dan distribusi eksak dengan *Multivariate shift Algorithm* (MSA).
2. Mengestimasi model regresi logistic dengan penggunaan rumus :

$$g(x) = \ln \left[\frac{\pi(x)}{1-\pi(x)} \right] = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p = \sum_{j=1}^p \beta_j x_j$$

dan regresi logistic eksak : $\text{Log} \frac{\pi_{ij}}{1-\pi_{ij}} = \beta X_k + yek$

3. Pengujian hipotesis dan menginterpretasi model estimasi :

- Dengan nilai p yaitu : $p(t_1/t_0) = \Pr(S \geq s) = \sum_{u \in \Phi} f_0(u | t_0)$ dan
- Dengan nilai tengah dari statistik uji p yaitu : $p(t_1 | t_0) - \frac{1}{2} f_0(t_1 | t_0)$

4. Menganalisis hasil / output

HASIL PENELITIAN

Data penelitian yang dipergunakan memiliki skala pengukuran yang bersifat kategori baik untuk variabel respon (Y) maupun variabel predictor (X), serta akan digunakan dua jenis sampel data, yaitu data sampel kecil (*Small Sampel*) dan data yang *spare* dengan $n = 10, 20, 29$ (kurang dari 30) dan yang mengandung data missing 55. Data tersebut diambil secara random dan memiliki nilai harapan (*expected*) di bawah 5. Sampel berukuran 55 mengandung data missing dengan nilai *expected* kurang dari 18 %, data dengan besar sampel 10, 20 dan 29 memiliki nilai *expected* kurang dari lima dengan persentasi di atas 20 %.

Model Sederhana (Singel variabel)

Model yang dihasilkan dari estimasi parameter adalah model dengan variabel respon (Y) yaitu status imunisasi (SI), sedangkan variabel bebas (X) meliputi 7 (tujuh) variabel. Dimulai dengan model sederhana dengan variabel bebas hanya satu variabel. Dilakukan pembentukan model;

1. Status Imunisasi = *constant* + Informasi ($Y = C + X_1$), untuk model sederhana ini untuk ukuran sampel 10, 20, 29 maupun 55 pada pengujian hipotesis ke 4 uji statistik ke empatnya memberikan kesimpulan menyatakan bahwa keterpaparan terhadap informasi tidak berpengaruh terhadap status imunisasi anak baduta.
2. Status Imunisasi = *constant* + Pendidikan ($Y = C + X_2$), pada pengujian hipotesis ke 4 uji statistik ke empatnya memberikan kesimpulan menyatakan bahwa tingkat pendidikan tidak berpengaruh terhadap status imunisasi anak baduta.
3. Status Imunisasi = *constant* + Pekerjaan ($Y = C + X_3$), pada pengujian hipotesis ke 4 uji statistik ke empatnya memberikan kesimpulan bahwa pekerjaan ibu tidak berpengaruh terhadap status imunisasi anak baduta.
4. Status Imunisasi = *constant* + Jumlah anak ($Y = C + X_4$), pada pengujian hipotesis ke 4 uji statistik ke empatnya memberikan kesimpulan bahwa jumlah anak hidup yang dimiliki tidak berpengaruh terhadap status imunisasi anak baduta.
5. Status Imunisasi = *constant* + Pengetahuan ($Y = C + X_5$), pada pengujian hipotesis nihil pada ukuran sampel 10, 20, 29 pengetahuan tidak berpengaruh terhadap status imunisasi anak baduta, ukuran sampel 55 memberikan kesimpulan berpengaruh terhadap status imunisasi anak baduta dengan nilai $p = 0.0003$ dan $p - \text{mid } 0,0002..$
6. Status Imunisasi = *constant* + Sikap ($Y = C + X_6$), pada pengujian hipotesis nihil bahwa sikap tidak berpengaruh pada status imunisasi anak baduta hanya uji eksak yang memberikan solusi pada ukuran 10. Tidak menolak hipotesis nihil, hanya dengan uji eksak nilai statistik lebih kecil dari nilai p lebih besar. Pada ukuran sampel 20 dan 29 memberikan kesimpulan menolak hipotesis nihil dengan nilai p sebesar 0.0055 dan 0.0013 dengan p-mid sebesar 0.0041 dan 0.0007.
7. Status Imunisasi = *constant* + Peran serta Masyarakat ($Y = C + X_7$), pada pengujian hipotesis nihil bahwa peran serta masyarakat tidak berpengaruh pada status imunisasi anak baduta kecuali pada ukuran sampel 10 hal tersebut berbeda. Keunggulan uji eksak dibandingkan dengan ketiga uji statistik lainnya adalah bisa menghasilkan p-mid.

Model Lengkap (Multi Variabel)

Variabel predictor yang terdiri dari tujuh buah variabel akan dimasukkan kedalam status Imunisasi = Constant + Informasi (X_1) + Pendidikan (X_2) + Pekerjaan (X_3) + Jumlah anak hidup (X_4) + Tingkat Pengetahuan (X_5) + Sikap (X_6) + Peran Serta Masyarakat (X_7). Parameter tersebut diestimasi, dihitung odds rasio dan dilakukan pengujian hipotesis atas semua variabel secara serentak. Dimana dari hasil penelitian bahwa inferensi eksak memberikan solusi pada variabel bebas Pendidikan, Pekerjaan, Jumlah Anak, Sikap dan Pengetahuan dengan galat baku $<< \text{MUE}$ dan interval kepercayaan berpeluang mengandung parameter koefisien regresi parameter tersebut. Sedangkan tipe

inferensi asimptotik tidak memberikan solusi apapun terhadap penaksiran ketujuh koefisien regresi tersebut.

1. Pengujian hipotesis atas parameter model Status Imunisasi = *Constant* + Informasi + Pendidikan + Pekerjaan + Jumlah Anak + Pengetahuan + Sikap + Peran Serta Masyarakat Model $Y = f(X_1 + X_7)$ untuk $n = 10, n = 20, n = 29, n = 55$
Untuk ukuran sampel 10, 20, 29 dan 55 inferensi eksak memberikan solusi pada variabel bebas, Pekerjaan, Jumlah anak, Sikap, Pengetahuan dan Peran Serta Masyarakat dengan galat baku $< < MUE$ dan interval kepercayaan berpeluang mengandung parameter koefisien regresi parameter tersebut. Sedangkan tipe inferensi asimptotik tidak memberikan solusi apapun terhadap penaksiran ketujuh koefisien regresi tersebut. Untuk sampel 29 variabel sikap menunjukkan berada di kedua interval dengan probabilitas meleset sebesar 0.0302. untuk sampel 55 variabel sikap menunjukkan berada di kedua interval dengan probabilitas meleset sebesar 0,0056 dan 0,0723.
2. Pengujian hipotesis atas parameter model Status Imunisasi = *Constant* + Informasi + Pendidikan + Pekerjaan + Jumlah Anak + Pengetahuan + Sikap + Peran Serta Masyarakat ($Y = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7$) untuk $n = 10$.
Dengan uji Rasio Likelihood dan uji Wald tidak memberikan solusi dalam pengujian hipotesis bahwa ketujuh variabel bebas secara bersama-sama berpengaruh terhadap status imunisasi. Hanya uji scores yang memberikan solusi dengan statistik scores = 10,0000 dan $p = 0,1886$ sehingga bisa disimpulkan bahwa ketujuh variabel bebas secara bersama-sama tidak berpengaruh terhadap status imunisasi. Namun uji ini tidak memberikan solusi untuk p-mid. Uji eksak memberikan nilai statistik lebih kecil dari pada uji scores dan memberikan nilai p lebih kecil serta memberikan solusi p-mid.
3. Pengujian hipotesis atas parameter model Status Imunisasi = *Constant* + Informasi + Pendidikan + Pekerjaan + Jumlah Anak + Pengetahuan + Sikap + Peran Serta Masyarakat ($Y = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7$) untuk $n = 20$.
Dengan uji Rasio Likelihood dan uji Wald tidak memberikan solusi dalam pengujian hipotesis bahwa ketujuh variabel bebas secara bersama-sama berpengaruh terhadap status imunisasi. Hanya uji scores yang memberikan solusi dengan statistik scores = 14,9604 dan $p = 0,0365$ sehingga bisa disimpulkan bahwa ketujuh variabel bebas secara bersama-sama berpengaruh terhadap status imunisasi. Namun uji ini tidak memberikan solusi untuk p-mid. Uji eksak memberikan nilai statistik lebih kecil dari pada uji scores (14,2124) dan memberikan nilai p lebih kecil (0.0075)serta memberikan solusi untuk p-mid.
4. Pengujian hipotesis atas parameter model Status Imunisasi = *Constant* + Informasi + Pendidikan + Pekerjaan + Jumlah Anak + Pengetahuan + Sikap + Peran Serta Masyarakat ($Y = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7$) untuk $n = 29$.
Dengan uji Rasio Likelihood dan uji Wald tidak memberikan solusi dalam pengujian hipotesis bahwa ketujuh variabel bebas secara bersama-sama berpengaruh terhadap status imunisasi. Hanya uji scores yang memberikan solusi dengan statistik scores = 15,7868 dan $p = 0,0271$ sehingga bisa disimpulkan bahwa ketujuh variabel bebas secara bersama-sama berpengaruh terhadap status imunisasi. Namun uji ini tidak memberikan solusi untuk p-mid. Uji eksak memberikan nilai statistik lebih kecil dari pada uji scores dan memberikan nilai p lebih kecil (0.0111) serta memberikan solusi untuk p-mid sebesar 0,0111.
5. Pengujian hipotesis atas parameter model Status Imunisasi = *Constant* + Informasi + Pendidikan + Pekerjaan + Jumlah Anak + Pengetahuan + Sikap + Peran Serta Masyarakat ($Y = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7$) untuk $n = 55$.
Dengan uji Rasio Likelihood dan uji Wald tidak memberikan solusi dalam pengujian hipotesis bahwa ketujuh variabel bebas secara bersama-sama berpengaruh terhadap status imunisasi. Hanya uji scores yang memberikan solusi dengan statistik scores = 17,3124 dan $p = 0,0155$ sehingga bisa disimpulkan bahwa ketujuh variabel bebas secara bersama-sama berpengaruh terhadap status imunisasi. Namun uji ini tidak memberikan solusi untuk p-mid. Uji eksak memberikan nilai statistik lebih kecil dari pada uji scores dan memberikan nilai p lebih kecil (0.0093)serta memberikan solusi untuk p-mid sebesar 0,0093.

Odds Rasio Regresi Logistik Eksak Model

1. Status Imunisasi = *Constant* + Informasi + Pendidikan + Pekerjaan + Jumlah Anak + Pengetahuan + Sikap + Peran Serta Masyarakat Model $Y = f (X_1 - X_7)$ untuk $n = 10$.
Odds variabel sikap dan PSM dapat diselesaikan pada ukuran sampel 10 dengan koefisien regresi masing-masing 1,0002 dan 2,0005 masih dalam interval kepercayaan, tipe inferensi asimptotik tidak memberikan solusi sama sekali dalam menghasilkan nilai odds rasio.
2. Status Imunisasi = *Constant* + Informasi + Pendidikan + Pekerjaan + Jumlah Anak + Pengetahuan + Sikap + Peran Serta Masyarakat Model $Y = f (X_1 - X_7)$ untuk $n = 20$.
Odds variabel sikap ternyata memiliki nilai tertinggi dibanding variabel lainnya, yaitu sebesar 6,6422 dengan interval kepercayaan 0,7600 sampai positif tidak terhingga yang artinya peluang terjadinya status imunisasi lengkap untuk ibu yang memiliki anak baduta yang mempunyai sikap mendukung adalah 6,6422 kali jika dibandingkan dengan ibu anak baduta yang mempunyai sikap tidak mendukung terhadap program imunisasi.
3. Status Imunisasi = *Constant* + Informasi + Pendidikan + Pekerjaan + Jumlah Anak + Pengetahuan + Sikap + Peran Serta Masyarakat Model $Y = f (X_1 - X_7)$ untuk $n = 29$.
Odds variabel sikap ternyata memiliki nilai tertinggi dibanding variabel lainnya, yaitu sebesar 9,5461 dengan interval kepercayaan 1,1921 sampai 299,4357 yang artinya peluang terjadinya status imunisasi lengkap untuk ibu yang memiliki anak baduta yang mempunyai sikap mendukung adalah 9.5461 kali jika dibandingkan dengan ibu anak baduta yang mempunyai sikap tidak mendukung terhadap program imunisasi. Sedangkan tipe inferensi asimptotik tidak memberikan solusi sama sekali dalam menghasilkan nilai odds rasio.
4. Status Imunisasi = *Constant* + Informasi + Pendidikan + Pekerjaan + Jumlah Anak + Pengetahuan + Sikap + Peran Serta Masyarakat Model $Y = f (X_1 - X_7)$ untuk $n = 55$.
Odds variabel sikap ternyata memiliki nilai tertinggi dibanding variabel lainnya, yaitu sebesar 7,7616 artinya peluang terjadinya status imunisasi lengkap untuk ibu yang memiliki anak baduta yang mempunyai sikap mendukung adalah 7.7616 kali jika dibandingkan dengan ibu anak baduta yang mempunyai sikap tidak mendukung terhadap program imunisasi. Tapi sebaliknya dengan variabel pengetahuan adalah 0,0691 kali terjadinya peluang tersebut.

Distribusi Bersyarat Eksak

Parameter yang akan ditaksir dari suatu *sufficient statistic* dilakukan inferensi berdasarkan distribusi permutasi eksak sehingga dapat ditentukan seberapa mungkin nilai pengamatan pada variabel respon. Nilai *sufficient statistic* untuk ukuran sampel 10,20,29 dan 30, 55 adalah sebagai berikut :

Tabel 1 : Permutation distribusi dari *sufficient statistic*

n	<i>sufficient statistic</i>							
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	C
10	3	2	3	2	6	5	7	7
20	13	16	18	15	20	20	14	11
29	12	5	5	4	7	16	18	19
55	15	16	19	14	10	19	11	10

Distribusi permutasi dari nilai statistik yang sesuai untuk ukuran sampan $n = 10, n = 20, n = 29, n = 55$ tergantung pada variabel yang diobservasi, misalkan variabel X_6 pada besar sampel $n = 20$ dengan probabilitas $Pr [X_6 | X_2 = 16, X_3 = 18, X_4 = 15, X_1 = 13, X_5 = 20, X_7 = 14, constant = 20]$.

PEMBAHASAN

Pada pemodelan lengkap dimana ketujuh variabel tersebut dimasukkan secara bersama-sama pada ukuran sampel 10, 20, 29 dan 55 berdasarkan pengujian hipotesis model pada analisis regresi logistic konvensional (asimptotik) tidak memberikan solusi, terlihat baik uji wald dan uji likelihood rasio tidak memberikan nilai p yang digunakan hipotesis nihil. dalam kasus sampel kecil ini dengan jumlah variabel 7 buah maka terdapat sudah lebih dari 20 % sel yang mempunyai nilai ekspektasi < 5 . (Koehler dan Larnzt, 1980) menunjukkan bahwa χ^2 valid dengan ukuran sampel yang lebih kecil dan table yang lebih jarang dari pada G^2 atau $-2 \log$ likelihood. jadi disini bahwa distribusi G^2 menjadi kurang baik untuk didekati dengan distribusi χ^2 yang diharapkan kurang dari 5.

1. Bentuk hubungan pada besar sampel 20 : $\text{Log} \frac{\pi_j}{1-\pi_j} = 5.133 + 2,3582 X_6$ (sikap).

Dilihat dari *odds ratio* pada model yang memiliki nilai p dan p – mid sebesar 0.0008 dimana peluang terjadinya status imunisasi lengkap untuk ibu anak baduta yang mempunyai sikap mendukung adalah 10,5724 kali jika dibandingkan dengan ibu anak baduta yang mempunyai sikap tidak mendukung terhadap program imunisasi.

2. Bentuk hubungan pada besar sampel 29 : $\text{Log} \frac{\pi_j}{1-\pi_j} = 2,9047 X_6$ (sikap).

Dilihat dari *odds ratio* pada model yang memiliki nilai p dan p – mid sebesar 0.0007 dimana peluang terjadinya status imunisasi lengkap untuk ibu anak baduta yang mempunyai sikap mendukung adalah 18,2595 kali jika dibandingkan dengan ibu anak baduta yang mempunyai sikap tidak mendukung terhadap program imunisasi.

3. Bentuk hubungan pada besar sampel 55 : $\text{Log} \frac{\pi_j}{1-\pi_j} = -3.2517 X_5 + 1.4889 X_6$ (sikap).

Dilihat dari *odds ratio* pada model yang memiliki nilai p dan p – mid sebesar 0.0001 dimana peluang terjadinya status imunisasi lengkap untuk ibu anak baduta yang mempunyai sikap mendukung adalah 4,4862 kali jika dibandingkan dengan ibu anak baduta yang mempunyai sikap tidak mendukung terhadap program imunisasi dan berbeda dengan peluang 0.0347 kali. Peningkatan pengetahuan justru terjadi kecenderungan kearah tidak lengkapnya imunisasi anak baduta.

Pada regresi logistic eksak untuk menentukan kemungkinan nilai pengamatan pada respon Y_0 yang dinyatakan dalam bentuk 2^n , maka semua kemungkinan yang mungkin terjadi pada respon. Missal dalam penelitian ini dimilkiki contoh 20 buah pengamatan maka semua kemungkinan yang mungkin timbul adalah $2^{20} = 1.048.576$ kemungkinan. Jika menghitung secara manual probabilitas setiap kemungkinan akan sangat menyulitkan karena ada berbagai kemungkinan yang akan terjadi. *Multifvariate Shift Algoritm* (MSA) yang diperkenalkan oleh Hijri, Metha, dan Patel (1987) distribusi bersyarat eksak dapat dibangkitkan dengan cara setiap vector $Y'X_0 = t_0$ dengan menghitung jumlah vector Y untuk setiap pengamatan $Y'X$, pada saat t tertentu dimana nilai pengamatan t dengan mengalikan Y dengan X_{ij} .

Dalam penanganan sampel kecil maupun *sparse data* hal ini juga sesuai dengan situasi kondisional dimana komponen tertentu t adalah constant, kita perlu mengetahui rangkaian nilai *feasible* dari komponen lain pada t hitung yang bersesuaian untuk setiap anggota rangkaian kondisional tersebut dimana kasus sampel kecil pada penelitian ini dalam hitungan distribusinya dengan statistik sufisien dari distribusi tersebut. Sebagai contoh variabel sufisien dengan X_6 (sikap) : $\text{Pr} [X_6 | X_2 = 16, X_3 = 18, X_4 = 15, X_1 = 13, X_5 = 20, X_7 = 14, \text{constant} = 11]$. Nilai statistik X_6 yang diobservasi adalah 20 maka $t_1 = 19, t_2 = 16, t_3 = 18, t_4 = 15, t_5 = 13, t_6 = 20, t_7 = 14, t_8 = 20$, menjadi statistik yang bersesuaian atau sufisien yang berhubungan dengan t_1 (sikap), t_2 (pendidikan), t_3 (pekerjaan), t_4 (jumlah anak), t_5 (informasi), t_6 (pengetahuan), t_7 (Peran Serta Masyarakat) dan t_8 constant term maka perhitungan distribusi tersebut adalah : c ($t_1, t_2 = 16, t_3 = 18, t_4 = 15, t_6 = 20, t_7 = 14, t_8 = 20$) dengan nilai semua kemungkinan t_1 dalam range 15 – 19 dimana $t_1 = 19$ merupakan maksimum dari range. Dari output logxact kita dapatkan nilai p value dari uji satu arah adalah tidak terdefinisi, jika jumlah probabilitas semua kemungkinan dari t_1 berhubungan dengan nilai yang lebih besar dari atau sama dengan *conditional scorers* $X_6 = 5,29412$ sehingga probabilitas yang dihasilkan dari semua kemungkinan = $0,0444444 + 0,0444444 = 0,0888888$ atau 0,0889 seperti dalam output logxact untuk nilai probabilitas X_6 pada pengujian hipotesis model lengkap.

Agar distribusi eksak χ^2 dan rasio likelihood atau G^2 mendekati distribusi χ^2 maka tidak mudah menentukan besar sampelnya. untuk sejumlah sel yang tetap, χ^2 biasanya berkonvergensi lebih cepat dari pada rasio likelihood atau G^2 . Pendekatan ke distribusi χ^2 untuk rasio likelihood atau G^2 menjadi lebih jelek bila frekuensi yang diharapkan (*expected frequency*) kurang dari 5. Bila hal tersebut terjadi maka tipe inferensi eksak amat dianjurkan. (Lamtz, 1978 ; Koehler, 1986 ; dalam Agresti, 1990)

SIMPULAN

1. Hubungan antara variabel respon (*respose variable*) dengan beberapa variabel predictor (*explanatory variables*) yang memiliki sampel kecil (*Sparse*) data dapat dilakukan dengan pendekatan metode regresi logistic eksak.
2. Model atau bentuk hubungan regresi logistic eksak yang paling signifikan adalah antara variabel respon status imunisasi dengan variabel predictor sikap untuk ukuran 20,29 dan pengetahuan serta sikap pada ukuran sampel 55.
3. Metode regresi logistic eksak dalam melakukan inferensi berdasarkan distribusi bersyarat eksak dari suatu statistik sufisien yang sesuai dengan parameter regresi yang ditaksir menggunakan *Multivariate Shift Algorithm* atau MSA.
4. Pengujian hipotesis atas parameter dari model regresi logistic eksak baik sampel kecil maupun *Sparse* data memberikan peluang penolakan hipotesis nihil lebih besar (nilai p lebih kecil) dan memberikan nilai p-mid. Regresi logistic konvensional (asimptotik) tidak dapat memberikan solusi nilai p-mid.
5. Untuk pengujian dengan banyak memasukkan variabel bebas metode regresi logistic eksak memberikan solusi lebih baik dari ketiga uji dalam metode regresi yang asimptotik.

SARAN :

1. Untuk menggunakan Metode Regresi Logistik Eksak bila ukuran sampel kecil, sampel yang mengandung data missing atau data sparse, table yang lebih jarang nilai di dalam selnya atau data yang memiliki nilai harapan (*expected*) kurang dari 5 lebih dari 20 %.
2. Perlu dikembangkan penggunaan data yang tidak seimbang (*Imbalance data*) pada kasus dengan jumlah variabel yang berbeda-beda.

PUSTAKA;

Agresti, Alan, 1990, *Categorical Data Analysis*, New York : John Wiley & Son, Inc

Agung, I Gusti Ngurah, 2001. *Statistika Analisis Hubungan kausal Berdasarkan Data Kategorik*, Jakarta : PT Raja Grafindo Persada.

Dixon WJ, Massey FJ, 1991. *Pengantar Analisa Statistik*, Yogyakarta : Gama University

Gaspersz, Vincent 1995, *Tehnik Analisis Dalam Penelitian Percobaan*, Bandung : Tarsito.

Hirji, Karim F; Cyrus R; Metha and Patel; Nitin R, 1987, *Computing Distributions For Exact Logistitc Regression*, JASA; 82 , 1110 -1117

Hosmer, David H Jr, Stanley Lemeshow, 1989. *Applied Logistic Regression*, Canada : John Wiley & Sons, Inc pp 135 – 171.

Johnsos, Richard A; Dean W. Wicherm, 1992, *Applied Multivariate Statistical Analysis. Third Edition*, New Jersey, Prentice-Hall, pp 1-36, 92 – 125.

Kendall, Sir Maurice G, William R. Buckland, 1976, *A Dictionary of Statistical Terms, Third Edition Revised and Enlarged*, London : Longman group Limited.

Kuntoro, 2000. *Regresi Logistik Metode Eksak*, Surabaya : Forum Ilmu Kesehatan Masyarakat th XIX no 18, hlm 74 – 82.

Mehta, Cyrus and Nitin Patel, 1993, *LogXact-Turbo Logistic Regression Software Featuring Exact Methods User Manua*. Cambbidge : Cytel Software Corporation.

Riono, Pandu Asri C. Adisasmita; Iwan Ariawan 1992, *Aplikasi Regresi Logistik* Jakarta : Fakultas Kesehatan Masyarakat Universitas Indonesia