Jurnal UJMC, Volume 2, Nomor 1, Hal. 50 - 55
ISSN : 2460 - 3333

PEMODELAN REGRESI COX DAN REGRESI WEIBULL WAKTU SEMBUH DIARE PADA BALITA

Siti Alfiatur Rohmaniah[1] dan Danardono[2]

[1] Universitas Islam Darul 'Ulum Lamongan, nia0304@gmail.com,
[2] Universitas Gadjah Mada, danardono@ugm.ac.id

Abstract. There are some consequences that occur because of diarrhea such as low levels of hemoglobin, weight loss, and dehydration. Severe dehydration can lead to death. This study aims to model the time cured of diarrhea on infants used the Cox regression method and Weibull regression to determine the factors that significantly influence the long diarrhea. The method used to analyze data on infants under diarrhea includes sex, duration of diarrhea from beginning to heal, age of the children, the value of nutrition in infants and toddlers hemoglobin levels. Further, it continued by modeling these factors using cox regression and regression weibull. The results obtained the nutritional value on infants significantly effect on diarrhea in infants.

Keywords: *Cox regression, Weibull regression, AFT models*

Abstrak. Ada beberapa akibat yang terjadi karena diare seperti rendahnya kadar Hb, berat badan menurun, dan dehidrasi. Dehidrasi yang berat dapat menyebabkan kematian. Penelitian ini bertujuan untuk memodelkan waktu sembuh diare pada balita menggunakan metode regresi Cox dan regresi Weibull untuk mengetahui faktor yang berpengaruh signifikan terhadap lama diare. Metode yang digunakan adalah menganalisis data mengenai balita yang mengalami diare meliputi jenis kelamin, lama waktu diare dari awal sampai sembuh, usia balita, nilai gizi pada balita, dan kadar hemoglobin balita. Selanjutnya memodelkan faktor-faktor tersebut menggunakan regresi cox dan regresi weibull. Hasil yang diperoleh nilai gizi pada balita berpengaruh signifikan terhadap lama diare pada balita.

Kata Kunci: *regresi Cox, regresi Weibull, model AFT.*

1 Pendahuluan

Diare bukanlah penyakit yang datang dengan sendirinya. Diare umumnya disebabkan karena asupan makanan yang terkontaminasi bibit penyakit ataupun racun. Diare akibat makanan yang terkena kuman biasanya menimbulkan gejala balita buang air besar kemudian muntah. Sebaliknya, diare karena keracunan, gejala utamanya adalah muntah baru diikuti diare.Penyakit diare dapat menyerang siapa saja, baik itu anak-anak maupun orang dewasa. Kasus ini banyak terjadi di Negara-negara berkembang dengan standar hidup yang rendah, ada beberapa akibat yang terjadi karena diare seperti rendahnya kadar Hb, berat badan menurun, dan akibat terparah pada anak-anak yang mengalami diare adalah dehidrasi. Kebanyakan orang tua kurang memahami adanya gejala dehidrasi yang terjadi pada bayi mereka. Bahkan banyak yang mempercayai terjadinya diare pada anak sebagai pertanda anak akan tambah pintar, tanpa menyadari bahasa dehidrasi yang mengancam dibalik diare itu.

Kondisi inilah yang seringkali menyebabkan diare yang berujung dehidrasi berat bisa menyebabkan kematian. Karena diare masih menjadi masalah yang bahaya, khususnya bagi kesehatan balita, peneliti akan memodelkan waktu sembuh diare pada balita menggunakan metode regresi Cox dan regresi Weibull untuk mengetahui faktor yang berpengaruh signifikan terhadap lama diare.

2 Kajian Teori

2.1 Regresi Cox

Model regresi Cox adalah model regresi hazard proporsional dengan fungsi *baseline* hazardnya dimodelkan secara nonparametrik dan fungsi variabel independennya dimodelkan secara parametrik. Regresi Cox dimodelkan sebagai:

$$h(t|x) = h_0(t)\psi(x,\beta). \tag{1}$$

Dengan $x = (x_1, ..., x_p)$ adalah vektor kovariat dan $\beta' = (\beta_1, ..., \beta_p)$ adalah parameter dari model regresi [3]. Dalam regresi ini, hazard untuk tiap-tiap individu sama dengan *baseline* hazard $h_0(t)$ apabila pengaruh variabel independen tidak diperhatikan. Estimasi parameter pada model regresi didasarkan pada *partial likelihood*:

$$L(\beta) = \prod_{k\in D} \frac{\exp(x_k\beta)}{\sum_{j\in R_k} \exp(x_j\beta)} \tag{2}$$

dengan x adalah vektor kovariat, β adalah parameter regresi, D adalah himpunan indeks j dari semua waktu kejadian,R_k adalah himpunan risiko semua individu yang belum mendapat kejadian pada saat tertentu. Fungsi *log partial likelihood* $l(\beta)$ adalah

$$l(\beta) = \sum_{k\in D} x_k\beta - \sum_{k\in D} log\left(\sum_{j\in R_k} \exp(x_j\beta)\right) \tag{3}$$

Algoritma Newton-Raphson untuk estimasi parameter berdasarkan *log partial likelihood* adalah [1]:

1. Mulai dengan nilai awal $\hat{\beta}^{(0)}$.
2. Pada iterasi ke-k, nilai estimasi di-*update*

$$\hat{\beta}^{(k+1)} = \hat{\beta}^{(k)} + I\left(\hat{\beta}^{(k)}\right)^{-1} U\left(\hat{\beta}^{(k)}\right).$$

3. Iterasi dihentikan dengan kriteria kekonvergenan $l\left(\hat{\beta}^{(k+1)}\right) \approx \left(\hat{\beta}^{(k)}\right)$.
4. Diperoleh estimasi $\hat{\beta}$ dan variansinya $\hat{V}(\hat{\beta}) = I(\hat{\beta})^{-1}$.

2.2 Regresi Weibull

Fungsi survival untuk regresi weibull adalah:

$$S(t|X) = \exp\left(-\left(f_\lambda(X;\beta)t\right)^\alpha\right). \tag{4}$$

Atau dapat ditulis

$$S(t|X) = \exp\left[-\exp\left(\frac{y - X\beta}{\sigma}\right)\right], \tag{5}$$

yang disebut fungsi survival distribusi *Extreme Value*, dengan parameter lokasi $\mu = -X\beta$ dan parameter skala $\sigma = \frac{1}{\alpha}$. Estimasi parameter menggunakan fungsi *likelihood* sebagai berikut:

$$L(\beta;\sigma) = \prod_{i=1}^{n} f(t_i, \theta|X_i)^{\delta_i} S(t_i, \theta|X_i)^{1-\delta_i}. \tag{6}$$

3 Hasil dan Pembahasan

Data yang digunakan dalam penelitian adalah data survival 96 pasien balita yang menderita diare di RS PKU Muhammadiyah Yogyakarta yang diperoleh dari skripsi Arum Ayu Kartika (tahun 2013), Program Studi Pendidikan dokter, Fakultas Kedokteran dan Ilmu Kesehatan, Universitas Muhammadiyah Yogyakarta [2]. Data tersebut meliputi data jenis kelamin ("1" laki-laki, "2" perempuan), lama diare (hari), umur (bulan), berat badan (kg), nilai gizi, kadar Hb (gr/dl), dan delta sama dengan 1. Data kemudian diolah menggunakan software R 3.3.2 dengan *library survival*.

3.1 Pemodelan Regresi Cox

Dengan menggunakan regresi Cox diperoleh hasil berikut. Dari Gambar 1, secara keseluruhan terlihat bahwa *p-value*=0,13 ¿ 0,05, artinya model regresi tidak layak untuk digunakan.

```
Concordance= 0.567  (se = 0.052 )
Rsquare= 0.085   (max possible= 0.999 )
Likelihood ratio test= 8.52  on 5 df,   p=0.13
Wald test            = 8.55  on 5 df,   p=0.1284
Score (logrank) test = 8.58  on 5 df,   p=0.1272
```

Gambar 1: Uji Keseluruhan Regresi Cox

Selanjutnya, dilakukan uji secara parsial dengan hasil sebagai berikut:

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
jk	0.141589	1.152104	0.215392	0.657	0.511
umur	0.005041	1.005054	0.021666	0.233	0.816
bb	-0.030324	0.970131	0.139420	-0.217	0.828
nilai_gizi	0.204127	1.226454	0.178650	1.143	0.253
HB	0.021321	1.021550	0.050967	0.418	0.676

Gambar 2: Uji Parsial Regresi Cox

Pada Gambar 2, semua variabel bebas tidak signifikan hal itu terlihat pada nilai *p-value* lebih besar dari 0,05. Oleh karenanya, perlu dilakukan analisis regresi ulang, tanpa mengikutsertakan variabel yang mempunyai *p-value* terbesar. Proses ini dilanjutkan sampai diperoleh semua variabel signifikan. Berikut hasil regresi ulangnya. Pada Gambar 3 terlihat bahwa nilai *P-value*=0,00489

```
Concordance= 0.571  (se = 0.052 )
Rsquare= 0.079   (max possible= 0.999 )
Likelihood ratio test= 7.92  on 1 df,   p=0.004887
Wald test            = 8.13  on 1 df,   p=0.004365
Score (logrank) test = 8.09  on 1 df,   p=0.004462
```

Gambar 3: Uji Keseluruhan Regresi Cox Ulang

kurang dari 0,05. Dengan demikian, dapat dikatakan bahwa model regresi sudah layak untuk digunakan.

	coef	exp(coef)	se(coef)	z	p
nilai_gizi	0.181	1.2	0.0634	2.85	0.0044

Gambar 4: Uji Parsial Regresi Cox Ulang

Terlihat pada Gambar 4, variabel bebas nilai gizi sudah signifikan $(0,0044 < 0,05)$, sehingga model regresi sudah layak untuk digunakan. Selanjutnya, diperoleh persamaan model regresi Cox-nya adalah sebagai berikut:

$$h(t|X) = h_0(t)\psi(X, \beta)$$

atau

$$h(t|X) = h_0(t)\exp(1,181(nilaigizi)).$$

Karena variabel bebas yang signifikan hanya variabel nilai gizi, dan datanya numerik (bukan kategorik), maka tidak perlu dicari proporsional hazardnya.

3.2 Pemodelan Regresi Model AFT Weibull

Dengan menggunakan regresi model AFT Weibull diperoleh hasil berikut. Pada Gambar 5 terlihat bahwa nilai $p - value = 0,2$ lebih besar dari 0,05, sehingga dikatakan bahwa model regresi tidak layak untuk digunakan.

```
Weibull distribution
Loglik(model)= -153.4   Loglik(intercept only)= -157.1
        Chisq= 7.33 on 5 degrees of freedom, p= 0.2
Number of Newton-Raphson Iterations: 7
n= 96
```

Gambar 5: Uji Keseluruhan Regresi Model AFT Weibull

	Value	Std. Error	z	p
(Intercept)	1.593955	0.28532	5.587	2.32e-08
jk	-0.029024	0.04694	-0.618	5.36e-01
umur	-0.000734	0.00481	-0.152	8.79e-01
bb	0.004981	0.03112	0.160	8.73e-01
nilai_gizi	-0.039141	0.03951	-0.991	3.22e-01
HB	-0.003378	0.01111	-0.304	7.61e-01
Log(scale)	-1.518136	0.08468	-17.928	7.10e-72

Gambar 6: Uji Parsial Regresi Model AFT Weibull

Secara parsial, semua variabel bebas tidak signifikan karena nilai *p-value* lebih besar dari 0,05 pada Gambar 6, sehingga perlu dilakukan analisis regresi ulang. Setelah dilakukan analisis regresi ulang diperoleh sebagai berikut:

```
Weibull distribution
Loglik(model)= -153.6   Loglik(intercept only)= -157.1
        Chisq= 6.87 on 1 degrees of freedom, p= 0.0087
Number of Newton-Raphson Iterations: 6
n= 96
```

Gambar 7: Uji Keseluruhan Regresi Model AFT Weibull Ulang

Berdasarkan Gambar 7, nilai $p - value = 0,0087 < 0,05$, artinya model regresi sudah layak untuk digunakan. Untuk uji parsialnya, diperoleh bahwa nilai konstanta (*intercept*) dan variabel bebas nilai gizi sudah signifikan (lihat Gambar 8), sehingga model regresi sudah layak untuk digunakan. Dengan begitu, persamaan model regresi Weibullnya adalah sebagai berikut:

$$S(t|X) = \exp(-\exp(\beta_0 + \beta_4 X_4)t)^{\alpha})$$

atau

$$S(t|X) = \exp(-\exp(1,5760 - 0,0356(nilaigizi)t)^{\alpha}).$$

Serta dapat dibuat fungsi hazard model AFT Weibull nya adalah sebagai berikut:

$$h(t|X) = \alpha(\exp(\beta_0 + \beta_4 X_4))((\exp(\beta_0 + \beta_4 X_4))t)^{\alpha-1}$$

atau

$$h(t|X) = \alpha(\exp(1,5760 - 0,0356(nilaigizi)))((\exp(1,5760 - 0,0356(nilaigizi)))t)^{\alpha-1}.$$

	Value	Std. Error	z	p
(Intercept)	1.5760	0.0298	52.9	0.00e+00
nilai_gizi	-0.0356	0.0132	-2.7	6.88e-03
Log(scale)	-1.5140	0.0843	-18.0	4.59e-72

Gambar 8: Uji Parsial Regresi Model AFT Weibull Ulang

4 Kesimpulan

Berdasarkan hasil penelitian dapat ditarik kesimpulan bahwa baik menggunakan metode regresi Cox maupun regresi Weibull, faktor yang berpengaruh signifikan terhadap fungsi survival waktu sembuh diare pada balita adalah variabel nilai gizi.

Daftar Pustaka

[1] Danardono. 2012. *Analisis Data Survival.* Diktat Kuliah Jurusan Matematika, FMIPA, UGM. Yogyakarta.

[2] Kartika, A. A. 2013. *Data Balita Menderita Diare di RS PKU Muhammadiyah Yogyakarta.* Skripsi. UMY. Yogyakarta.

[3] Klien, J. P. 2003. *Survival Analysis Techniques for Concored and Truncated Data.* Springer. New York.

[4] Tableman, M. dan Kim, J. S. *Survival Analysis using S.* Chapman and Hall/CRC. New York.