

# **Penggunaan Metode Bootstrap dalam Regresi *Cox Proportional Hazards* pada Ketahanan Hidup Pasien *Diabetes mellitus***

**Ninuk Rahayu<sup>a</sup>, Adi Setiawan<sup>b</sup>, Tundjung Mahatma<sup>c</sup>**

<sup>a,b,c</sup> Program Studi Matematika Fakultas Sains dan Matematika,

Universitas Kristen Satya Wacana, Jl. Diponegoro No. 52 - 60, Salatiga

<sup>a</sup> [nieralydia@yahoo.com](mailto:nieralydia@yahoo.com), <sup>b</sup> [adi\\_setia\\_03@yahoo.com](mailto:adi_setia_03@yahoo.com), <sup>c</sup> [t.mahatma@gmail.com](mailto:t.mahatma@gmail.com)

## **ABSTRAK**

Model regresi *Cox proportional hazards* pada ketahanan hidup pasien *Diabetes mellitus* telah dikaji dan dihasilkan estimasi koefisien ( $\beta$ ) model umum regresi *Cox proportional hazards* (Rahayu dkk, 2012). Dalam penelitian ini akan digunakan metode bootstrap untuk mengestimasi koefisien ( $\beta$ ) model regresi *Cox proportional hazards*. Dengan bantuan program aplikasi R 2.15.1 menggunakan metode bootstrap menghasilkan interval konfidensi bootstrap persentil untuk estimasi koefisien ( $\beta$ ) model regresi *Cox proportional hazards*. Diperoleh interval konfidensi bootstrap persentil untuk setiap variabel yaitu variabel genetik, usia, diet, olahraga, berat badan. Interval yang signifikan adalah interval yang tidak memuat nol yaitu interval konfidensi bootstrap persentil untuk  $X_1, X_2, X_3$  (genetik, usia dan diet).

**Kata Kunci :** metode bootstrap, cox proportional hazards, regresi cox.

## **ABSTRACT**

*Cox proportional hazards regression model of the survival of Diabetes mellitus patients that have been observed and produced the estimated of coefficient in general model, (Rahayu dkk, 2012). In this study, the method used to estimate the coefficient in Cox proportional hazards model is bootstrap method. Using the application program R 2.15.1 to bootstrap this data and it will result the percentile bootstrap confidence interval for regression coefficients in Cox proportional hazards model. Then, percentile bootstrap confidence interval for each variable can be found. Significant coefficient can be determined by interval that do not contain zero. Based on this condition we found that significant variables are genetic, age and diet.*

**Key Words :** bootstrap method, cox regression proportional hazards.

## **Pendahuluan**

Analisis *survival* adalah salah satu cabang statistika yang mempelajari teknik analisis data *survival*. Untuk menentukan besarnya hubungan antara variabel independen dengan variabel dependennya digunakan model *Cox proportional hazards*. Pada penelitian ini akan dilakukan studi simulasi regresi *Cox proportional hazards* dari sampel yang dibangkitkan berdasarkan metode bootstrap.

Penelitian berkaitan dengan metode bootstrap pada data *survival* sebelumnya telah dilakukan oleh Unzet dan Sanchez (1992) tentang aplikasi metode bootstrap pada analisis *survival* yaitu Kaplan Meier yang memperoleh kesimpulan bahwa metode bootstrap dapat mengestimasi mediannya. Azimmatul Ihwah (2012) menggunakan metode bootstrap dalam mengestimasi koefisien beta model regresi *Cox* dan diperoleh kesimpulan resampling bootstrap dapat diterapkan pada data *survival*. Di samping itu untuk penelitian berkaitan dengan regresi *Cox proportional hazards* pada data ketahanan hidup pasien *Diabetes mellitus* telah dikaji oleh Rahayu dkk (2012) menghasilkan tiga variabel yang signifikan berpengaruh pada ketahanan hidup pasien.

Dalam penelitian ini diambil data pasien *Diabetes mellitus* rawat inap 2011 di RSUD RAA Soewondo Pati, Jawa Tengah, Rahayu dkk (2012). Data pasien meliputi, waktu bertahan hidup, status genetik, usia, diet *Diabetes mellitus*, olahraga dan berat badan. Berdasarkan data pasien tersebut akan dibangkitkan data baru menggunakan metode bootstrap kemudian dicari interval konfidensi bootstrap persentil. Tujuan dari penelitian ini adalah menerapkan metode bootstrap untuk memperoleh interval konfidensi bootstrap persentil untuk koefisien ( $\beta$ ) model *Cox proportional hazards*. Hasil regresi *Cox proportional hazards* tersebut telah dibahas dalam makalah Rahayu, dkk (2012) dengan koefisien konfidensi 95 %. Jika interval konfidensi persentil bootstrap pada variabel  $X_i$  memuat nol maka koefisien regresi *Cox Proportional Hazards* dikatakan tidak signifikan. Jika interval konfidensi bootstrap persentil dengan koefisien konfidensi 95% pada variabel  $X_p$  tidak memuat nol maka dikatakan signifikan.

## **Fungsi Survival**

Misalkan  $T$  adalah waktu bertahan hidup sampai munculnya kejadian tertentu. Kejadian yang dimaksud misalnya kematian, berkembangnya penyakit tertentu, kambuhnya penyakit setelah dilakukan terapi, dan lain-lain.

Fungsi *survival*,  $S(t)$ , mendefinisikan probabilitas dari suatu individu untuk bertahan setelah waktu yang ditetapkan, namakan  $t$ ,  $S(t) = Pr(T > t)$ .

### **Model Regresi Cox Proportional Hazards**

Menurut StatSci (1995) model regresi Cox mengasumsikan bahwa fungsi *hazards* sebagai :

$$h(t; x) = h_0(t) \cdot r_i(t) \quad (1)$$

dengan  $r_i(t) = \exp(\beta_p X_p(t))$ , sebagai skor resiko untuk individu ke- $i$ ,  $\beta$  adalah vektor koefisien regresi berdimensi  $p$ , dan  $h_0(t)$  merupakan fungsi *hazards* dasar (*baseline hazards function*). Fungsi eksponensial menjamin  $h$  positif untuk setiap  $\beta$ , sehingga bentuk umum regresi Cox adalah :

$$h(t) = h_0(t) \cdot \exp(\beta_1 X_1(t) + \dots + \beta_p X_p(t)) \quad (2)$$

### **Metode Bootstrap**

Untuk mengetahui apakah variabel  $X_p$  (untuk  $p = 1, 2, 3, 4, 5$ ) signifikan atau tidak maka dibuat interval konfidensi bootstrap persentil 95 % dari hasil pembentukan sampel baru oleh bootstrap. Kemudian dilakukan ulangan sejumlah bilangan besar  $B$  kali pada sampel baru tersebut. Langkah-langkah dalam membuat interval konfidensi bootstrap persentil adalah sebagai berikut:

1. Melakukan proses bootstrap sebanyak bilangan besar  $B$  kali.

2. Mengurutkan hasil regresi Cox *proportional hazards* dari pembentukan sampel baru.
3. Buat histogram hasil pengulangan bootstrap.
4. Dengan koefisien konfidensi 95% maka dapat ditentukan interval konfidensi 95 % yaitu dengan memilih batas atas sebesar 97.5 % dan batas bawah sebesar 2.5 %.

Keputusan secara statistik dalam metode bootstrap untuk menentukan variabel  $X_1, X_2, X_3, X_4, X_5$  signifikan atau tidak adalah sebagai berikut :

1. Jika interval konfidensi koefisien ( $\beta$ ) regresi Cox *proportional hazards* menggunakan metode bootstrap memuat nilai nol maka koefisien ( $\beta$ ) tidak signifikan.
2. Jika interval konfidensi koefisien ( $\beta$ ) regresi Cox *proportional hazards* menggunakan metode bootstrap tidak memuat nilai nol maka koefisien ( $\beta$ ) tersebut signifikan.

### **Metode Penelitian**

Data yang digunakan adalah data primer dan sekunder pasien *Diabetes mellitus* RSUD RAA Soewondo Pati tahun 2011 sebanyak 65 pasien. Sebelumnya telah diteliti mengenai

regresi *Cox proportional hazards* pada data tersebut dengan hasil yang diperoleh

**Tabel 1** dan **Tabel 2** (Rahayu dkk, 2012).

**Tabel 1.** Hasil regresi *Cox proportional hazards*

Variabel	Coef	exp(coef)	se(coef)	z	Pr(> z )
( $X_1$ ) genetik	1.6103	5.0043	0.7103	2.267	0.0234 *
( $X_2$ ) usia	-2.6284	0.0722	0.8070	-3.257	0.0011 **
( $X_3$ ) diet	0.9236	2.5183	0.3741	2.469	0.0135 *
( $X_4$ ) olahraga	-0.1610	0.8513	0.3492	-0.461	0.6448
( $X_5$ ) berat badan	-0.0286	0.9718	0.0545	-0.525	0.5996

**Tabel 2.** Hasil Uji Parsial Model Terbaik

Variabel	coef	exp(coef)	se(coef)	z	Pr(> z )
( $X_2$ ) usia	-2.799	0.0608	0.779	-3.60	0.0003
( $X_1$ ) gen	1.566	4.7868	0.661	2.37	0.0180
( $X_3$ ) diet	0.885	2.4220	0.363	2.44	0.0150

Untuk melakukan analisis data digunakan alat bantu program aplikasi R 2.15.1. Dalam penelitian ini akan dilakukan analisis koefisien ( $\beta$ ) model regresi *Cox proportional hazards* menggunakan metode bootstrap untuk setiap variabel sehingga akan diperoleh 5 koefisien ( $\beta$ ) pada variabel bebasnya dalam satu kali pengulangan.

Langkah-langkah yang dilakukan dalam penelitian ini adalah sebagai berikut :

1. Meregresikan data pasien untuk memperoleh koefisien ( $\beta$ ) model regresi *Cox proportional hazards*, pada makalah Rahayu dkk (2012).
2. Dilakukan pemilihan variabel dengan memperhatikan nilai  $-2 \text{ Log Likelihood}$ , sehingga diperoleh variabel-variabel yang signifikan (Rahayu dkk, 2012).

3. Memilih satu baris data ( $X_1, X_2, X_3, X_4, X_5$ ) dari data asli secara acak sebanyak 65 baris dengan pengembalian.

4. Menghitung koefisien regresi *Cox* berdasarkan data baru.

5. Melakukan proses bootstrap pada langkah 3 dan 4 sebanyak bilangan besar  $B$  kali (misalkan  $B = 1000$ ), sehingga diperoleh matriks

$$\begin{bmatrix} \beta_{11} & \cdots & \beta_{51} \\ \vdots & \ddots & \vdots \\ \beta_{1B} & \cdots & \beta_{5B} \end{bmatrix}.$$

6. Mengurutkan matriks  $\begin{bmatrix} \beta_{11} & \cdots & \beta_{51} \\ \vdots & \ddots & \vdots \\ \beta_{1B} & \cdots & \beta_{5B} \end{bmatrix}$  tiap kolom.

7. Membuat histogram tiap kolom sebesar  $X_p$  yaitu sebanyak variabel independen yang dipakai.

8. Menghitung interval konfidensi bootstrap persentil 95% dan mean pada data hasil langkah 5 dan 6.

9. Membuat keputusan signifikansi koefisien regresi.

**Hasil dan Pembahasan**

Menurut makalah Rahayu, dkk (2012), koefisien regresi *Cox proportional hazards* telah diperoleh pada **Tabel 1** sehingga model umum regresi *Cox proportional hazards* menurut persamaan (2) adalah :

$$h(t) = h_0(t). \exp (1.6103 X_1 - 2.6284 X_2 + 0.9236 X_3 - 0.1610 X_4 - 0.0286 X_5).$$

Kemudian pemilihan model terbaik adalah pemilihan variabel bebas dengan memperhatikan nilai *-2 Log Likelihood*-nya sehingga diperoleh model terbaik dengan variabel-variabel yang signifikan yaitu :

$$h(t) = h_0(t). \exp(-2.799 X_2 + 1.566 X_1 - 0.885 X_3).$$

Dengan demikian diperoleh variabel signifikan pada tingkat  $\alpha=5\%$  menurut regresi *Cox proportional hazards* yaitu  $X_1, X_2, X_3$ .

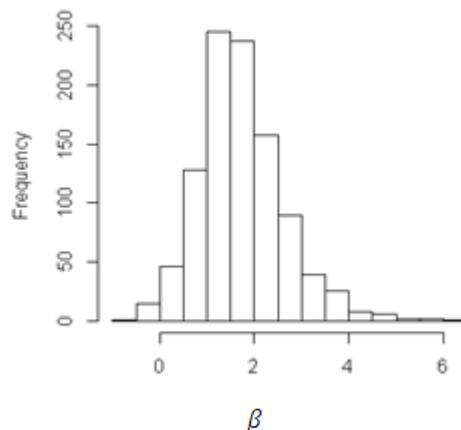
**Metode bootstrap pada lima variabel**

Pertama, dilakukan proses bootstrap pada kelima variabel  $X_1, X_2, X_3, X_4, X_5$ . Proses bootstrap dilakukan dengan pengambilan sampel baru dari data secara berpasangan dengan pengembalian. Data baru tersebut selanjutnya diregresikan dengan regresi *Cox proportional hazards* sehingga dihasilkan **Tabel 3**

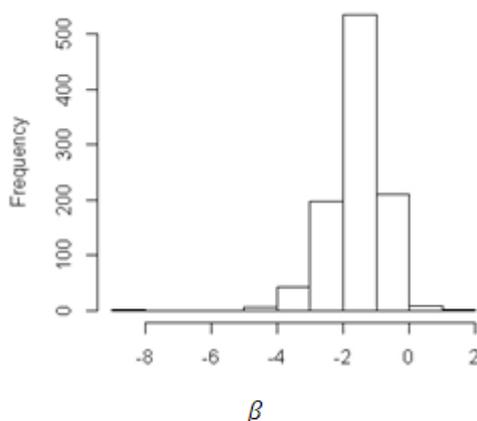
**Tabel 3.** Hasil regresi *Cox* data baru

Variabel	Koefisien
$X_1$	1.1967
$X_2$	-1.6551
$X_3$	1.0802
$X_4$	-0.7996
$X_5$	0.0488

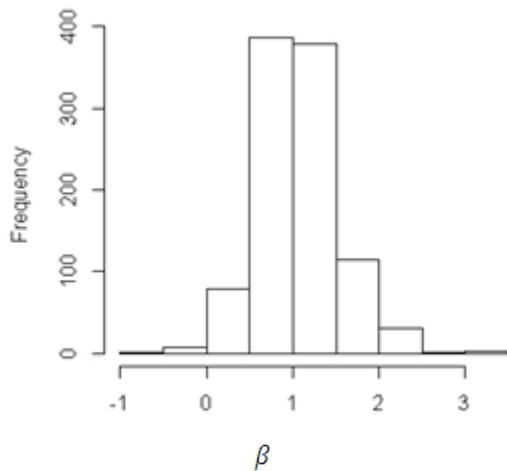
Kemudian dilakukan proses bootstrap untuk menentukan interval konfidensi bootstrap persentil 95% pada sampel baru dengan pengulangan 1000 kali. Histogram dari nilai-nilai koefisien regresi untuk masing-masing variabel yaitu  $X_1, X_2, X_3, X_4, X_5$  dinyatakan pada Gambar 1 sampai Gambar 5.



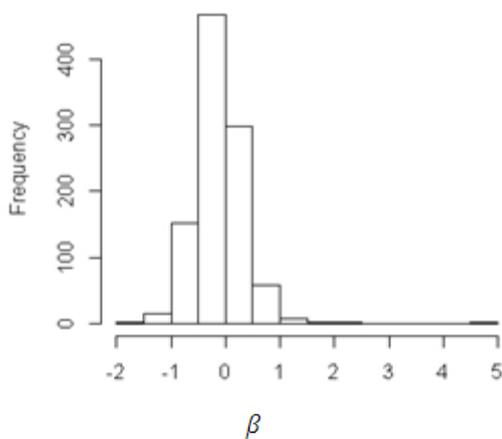
Gambar 1. Histogram nilai koefisien regresi  $X_1$



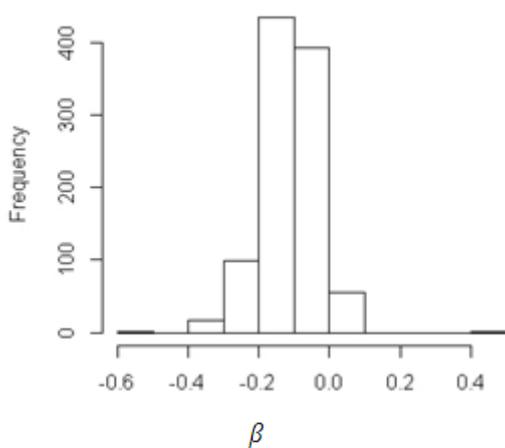
Gambar 2. Histogram nilai koefisien regresi  $X_2$



Gambar 3. Histogram nilai koefisien regresi  $X_3$



Gambar 4. Histogram nilai koefisien regresi  $X_4$



Gambar 5. Histogram nilai koefisien regresi  $X_5$

Hasil interval konfidensi bootstrap persentil tersebut disajikan pada **Tabel 4**.

**Tabel 4.** Interval konfidensi 95% lima variabel

Variabel $X_i$	Interval konfidensi	Rata-rata bootstrap
$X_1$	(0.1243 , 3.8182)	1.7425
$X_2$	(-3.2568 , -0.2716)	-1.5644
$X_3$	(0.1909 , 2.0517)	1.0631
$X_4$	(-0.9497 , 0.8012)	-0.1128
$X_5$	(-0.2839 , 0.0292)	-0.1125

**Tabel 4** menunjukkan bahwa koefisien ( $\beta$ )  $X_1$  (genetik) berada di dalam interval konfidensi persentil (0.1243 , 3.8182), dan interval tersebut tidak memuat nol maka  $X_1$  signifikan. Koefisien ( $\beta$ )  $X_2$  (usia) berada di dalam interval konfidensi persentil (-3.2568 , -0.2716) dan interval tersebut tidak memuat nol maka  $X_2$  signifikan. Koefisien ( $\beta$ )  $X_3$  (diet) berada di dalam interval konfidensi persentil (0.1909 , 2.0517) dan interval tersebut tidak memuat nol maka  $X_3$  signifikan. Untuk koefisien ( $\beta$ )  $X_4$  (olahraga) dan  $X_5$  (berat badan) interval konfidensi persentil bootstrap memuat nol maka tidak signifikan.

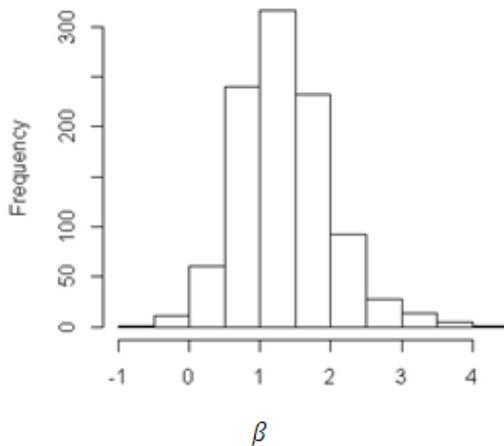
#### Metode Bootstrap pada tiga variabel

Telah diperoleh tiga variabel ( $X_1, X_2, X_3$ ) yang signifikan oleh proses bootstrap pada tahap sebelumnya. Kemudian dilakukan proses bootstrap pada ketiga variabel  $X_1, X_2, X_3$  Dengan cara yang sama diperoleh hasil regresi Cox tiga variabel terhadap data baru pada **Tabel 5**.

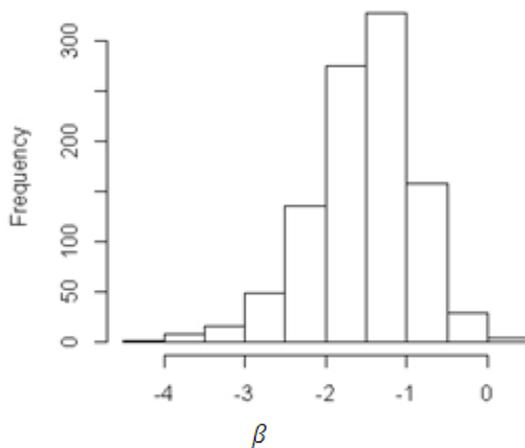
**Tabel 5.** Hasil regresi *Cox* data baru

Variabel $X_i$	Koefisien
$X_1$	1.491
$X_2$	-2.260
$X_3$	0.642

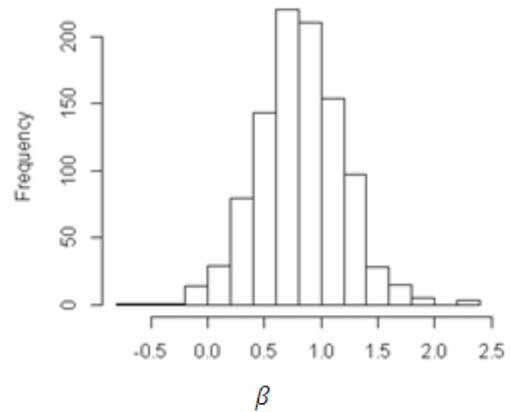
Selanjutnya dilakukan proses bootstrap dengan pengulangan 1000 kali untuk ketiga variabel  $X_1, X_2, X_3$  diperoleh histogram dari nilai-nilai koefisien regresi untuk variabel  $X_1, X_2, X_3$ .



Gambar 6. Histogram nilai koefisien regresi  $X_1$



Gambar 7. Histogram nilai koefisien regresi  $X_2$



Gambar 8. Histogram nilai koefisien regresi  $X_3$

Selanjutnya diperoleh interval konfidensi bootstrap persentil pada **Tabel 6**.

**Tabel 6.** Interval konfidensi 95% tiga variabel

Variabel $X_i$	Interval konfidensi	Rata-rata bootstrap
$X_1$	(0.1907, 2.8155)	1.3476
$X_2$	(-2.9791, -0.4289)	-1.5303
$X_3$	(0.0720, 1.5609)	0.8169

Koefisien hasil regresi tiga variabel pada data asli (**Tabel 2**) dan data baru (**Tabel 5**) berada di dalam interval bootstrap persentil (**Tabel 6**) dan interval tersebut tidak memuat nol untuk  $X_1, X_2, X_3$  dengan nilai rata-rata untuk  $X_1 = 1.3476$ ,  $X_2 = -1.5303$ ,  $X_3 = 0.8169$ .

Dengan demikian variabel  $X_1, X_2, X_3$  signifikan yaitu meliputi variabel genetik, usia dan diet.

### Kesimpulan

Dari hasil pembahasan estimasi koefisien ( $\beta$ ) pada model regresi *Cox proportional hazards* dengan menggunakan metode bootstrap diperoleh interval konfidensi bootstrap persentil pada masing-masing variabel. Interval

konfidensi yang tidak memuat nol adalah interval konfidensi variabel  $X_1, X_2, X_3$  sehingga variabel tersebut signifikan. Interval konfidensi yang memuat nol adalah interval konfidensi pada variabel  $X_4$  dan  $X_5$  maka tidak signifikan.

### **Pustaka**

- Davison A.C & Hinkley, 2003, *Bootstrap Methods and Their Application*. United States of America.
- Ihwah, Azimatul, 2012, *Bootstrap Pada Data Tersensor Untuk Estimasi Beta dalam Model Regresi Cox*. Jurusan Matematika, Universitas Gadjah Mada, Yogyakarta.
- Klein & Moeschberger, 1997, *Survival Analysis. Techniques for Censored and Truncated Data*. Springer, New York, USA.
- Pradeka R. Setiawan. A & L. Linawati, 2012, Uji Koefisien Korelasi Spearman dan Kendall Menggunakan Metode Bootstrap, *Prosiding Seminar Nasional Matematika UNS 6 Oktober 2012*.
- Rahayu N, Setiawan A & Mahatma T, 2012, Analisis Regresi Cox Proportional Hazards pada Ketahanan Hidup Pasien Diabetes mellitus, *Prosiding Seminar Nasional Matematika UNS 6 Oktober 2012*.
- StatSci Division, *S-PLUS Guide to Statistical and Mathematical Analysis*, Seattle Washington, 1995.