

PERLUASAN REGRESI COX DENGAN PENAMBAHAN PEUBAH TERIKAT-WAKTU

Luh Putu Ari Dewiyanti^{§1}, Ni Luh Putu Suciptawati², I Wayan Sumarjaya³

¹Jurusan Matematika, Fakultas MIPA - Universitas Udayana [Email: ari.dewiyanti@gmail.com]

²Jurusan Matematika, Fakultas MIPA - Universitas Udayana [Email: putusuciptawati@yahoo.co.id]

³Jurusan Matematika, Fakultas MIPA - Universitas Udayana [Email: sumarjaya@unud.ac.id]

[§]Corresponding Author

ABSTRACT

The aim of this study is to model job hunting period in Bali in 2012 using Extended Cox model. Previous study concluded that household status and age variables were not significantly influenced the job hunting period. However, previous study on factors that influence job waiting suggests that both variables should play important role in determining the waiting time for job hunters. Thus incorporating time-dependent covariates into model is necessary. After incorporating time-dependent covariates we found that age with time-dependent covariate is significant. Hence we obtain the following model

$$h(t, X(t)) = h_0(t) \exp[-0,28271x_1 - 0,32952x_2 + 1,51686x_3(\text{divorced}) + 0,37016x_3(\text{married}) - 0,00547x_7g(t)]$$

Keywords: *extended Cox model, job hunting period, survival analysis*

1. PENDAHULUAN

Analisis sintasan merupakan sekumpulan prosedur statistika yang digunakan untuk analisis data, dengan peubah yang diperhatikan adalah waktu sampai terjadinya suatu kejadian (Kleinbaum & Klein [2]). Waktu sampai terjadinya suatu kejadian ini dikenal dengan istilah waktu sintasan yang dapat dinyatakan dalam hari, bulan, maupun tahun. Waktu sintasan ini digunakan dalam mengukur data waktu suatu kejadian tertentu seperti respon, kematian, kegagalan, dan pembebasan bersyarat (Lee & Wang [3]).

Metode regresi yang paling umum digunakan dalam analisis sintasan adalah metode regresi Cox *proportional hazard*. Metode ini umum digunakan karena model tidak harus didasari oleh suatu distribusi tertentu, tanpa ditentukan *baseline hazard*

model ini dapat digunakan (Kleinbaum & Klein [2]). Metode regresi ini mempunyai asumsi yang harus dipenuhi yaitu rasio dari dua angka *hazard* harus konstan terhadap waktu kegagalan atau dengan kata lain *hazard rate* untuk satu individu sebanding dengan *hazard rate* individu lain. Namun terkadang terdapat beberapa kasus yang tidak semua peubah bebas memenuhi asumsi *proportional hazard* yang menyebabkan peubah bebas tersebut tidak akan signifikan terhadap model regresi Cox *proportional hazard*.

Oleh karena itu diperlukan metode lain untuk mendapatkan hasil yang lebih baik untuk menganalisis data sintasan tersebut. Terdapat alternatif cara ketika asumsi *proportional hazard* tidak terpenuhi salah satunya dengan menggunakan perluasan model

regresi Cox dengan penambahan peubah terikat-waktu.

Penelitian ini menggunakan studi kasus faktor-faktor yang memengaruhi lama mencari pekerjaan di Provinsi Bali pada tahun 2012. Studi kasus ini digunakan karena pada penelitian sebelumnya terdapat dua peubah yaitu umur dan status dalam rumah tangga yang tidak signifikan berpengaruh terhadap model lama mencari kerja. Sesungguhnya bila dikaitkan dengan beban tanggung jawab, status rumah tangga dari individu akan memiliki pengaruh terhadap lama mencari kerja. Individu dengan status kepala rumah tangga pastinya akan mencari pekerjaan lebih intensif daripada individu dengan status anggota rumah tangga. Faktor umur juga merupakan salah satu faktor yang memengaruhi cepat lambatnya individu mendapatkan suatu pekerjaan (Setiawan [4]). Oleh karena itu, perlu ditinjau kembali peubah umur dan status dalam rumah tangga ini kedalam metode perluasan model regresi Cox.

Tujuan penelitian ini adalah mengetahui bentuk perluasan model regresi Cox dengan penambahan peubah terikat-waktu. Selain itu, tujuan dari penelitian ini adalah untuk mengetahui diantara kedua faktor yang akan diuji, faktor manakah yang berpengaruh dalam lama mencari pekerjaan di Provinsi Bali dengan menggunakan perluasan model regresi Cox dengan penambahan peubah terikat-waktu.

2. KAJIAN PUSTAKA

Regresi Cox *proportional hazard* merupakan salah satu metode regresi yang paling populer digunakan dalam analisis sintasan ketika hendak mengetahui hubungan antara waktu kejadian dengan peubah bebas tertentu pada saat dilakukannya penelitian (Kleinbaum & Klein [2]). Misalkan risiko individu ke- i mengalami kejadian pada saat t yaitu $h(t)$ bergantung pada nilai x_1, x_2, \dots, x_p dari peubah bebas X_1, X_2, \dots, X_p . Himpunan nilai peubah bebas dinyatakan dalam vektor \mathbf{x}

dengan $\mathbf{x} = (x_1, x_2, \dots, x_p)'$ (Collett [1]). Secara umum persamaan model Cox *proportional hazard* adalah sebagai berikut (Kleinbaum & Klein [2]):

$$h(t, \mathbf{x}) = h_0(t) \exp \sum_{i=1}^p \beta_i x_i.$$

Notasi $h_0(t)$ merupakan *baseline hazard*, vektor $\boldsymbol{\beta}$ adalah parameter nilai peubah bebas x_1, x_2, \dots, x_p dalam model. Nilai *hazard ratio* pada model regresi Cox *proportional hazard* dapat dituliskan sebagai berikut (Kleinbaum & Klein [2]):

$$\begin{aligned} \widehat{HR} &= \frac{\hat{h}(t, \mathbf{x}^*)}{\hat{h}(t, \mathbf{x})} \\ &= \frac{h_0(t) \exp \left[\sum_{i=1}^p \beta_i x_i^* \right]}{h_0(t) \exp \left[\sum_{i=1}^p \beta_i x_i \right]} \\ &= \exp \left[\sum_{i=1}^p \beta_i (x_i^* - x_i) \right] \end{aligned}$$

dengan \mathbf{x}_i^* dan \mathbf{x}_i merupakan vektor peubah bebas dari dua individu yang berbeda, yang mana $\mathbf{x}_i^* = (x_1^*, x_2^*, \dots, x_p^*)$ dan $\mathbf{x}_i = (x_1, x_2, \dots, x_p)$. Terlihat bahwa perbandingan *hazard* kedua objek tetap sepanjang waktu, atau tidak bergantung pada waktu, ini menandakan bahwa asumsi *proportional hazard* terpenuhi.

Berbeda dengan persamaan regresi Cox *proportional hazard*, persamaan perluasan model regresi Cox dengan penambahan peubah terikat-waktu adalah sebagai berikut (Kleinbaum & Klein [2]):

$$h(t, \mathbf{x}(t)) = h_0(t) \exp \left[\sum_{i=1}^{p_1} \beta_i x_i + \sum_{j=1}^{p_2} \delta_j x_j(t) \right]$$

dengan $\mathbf{x}(t) = (x_1, x_2, \dots, x_{p_1}, x_1(t), x_2(t), \dots, x_{p_2}(t))$, peubah x_1, x_2, \dots, x_{p_1} adalah peubah yang bebas oleh waktu, dan $x_1(t), x_2(t), \dots, x_{p_2}(t)$ adalah peubah bebas yang terikat-waktu. Notasi $\boldsymbol{\delta}$ adalah vektor parameter dari peubah bebas terikat-waktu yang berlaku untuk setiap t dan $\boldsymbol{\beta}$ merupakan vektor parameter dari peubah bebas tidak terikat-waktu. Pada penelitian ini menggunakan fungsi waktu adalah $g(t) = t$, sehingga model ini dapat

dituliskan sebagai:

$$h(t, \mathbf{x}(t)) = h_0(t) \exp\left[\sum_{i=1}^{p_1} \beta_i x_i + \sum_{j=1}^{p_2} \delta_j x_j g_j(t)\right].$$

Hazard ratio dalam model tersebut ditunjukkan dengan perbandingan dua kelompok peubah bebas \mathbf{x}^* dan \mathbf{x} ditambahkan dengan pengaruh waktu sehingga didapat $\mathbf{x}^*(t)$ dan $\mathbf{x}(t)$, secara matematis dapat dituliskan sebagai berikut:

$$\begin{aligned} \widehat{HR}(t) &= \frac{\widehat{h}(t, \mathbf{x}^*(t))}{\widehat{h}(t, \mathbf{x}(t))} \\ &= \frac{h_0(t) \exp\left(\sum_{i=1}^{p_1} \beta_i x_i^* + \sum_{j=1}^{p_2} \delta_j x_j^* g_j(t)\right)}{h_0(t) \exp\left(\sum_{i=1}^{p_1} \beta_i x_i + \sum_{j=1}^{p_2} \delta_j x_j g_j(t)\right)} \\ &= \exp\left[\sum_{i=1}^{p_1} \beta_i [x_i^* - x_i] + \sum_{j=1}^{p_2} \delta_j [x_j^*(t) - x_j(t)]\right] \end{aligned}$$

dengan nilai parameter $\widehat{\delta}_j$ akan berbeda-beda pada setiap peubah bebas yang terikat oleh waktu dan parameter $\widehat{\delta}_j$ ini mengartikan pengaruh keseluruhan dari hubungan antara peubah bebas terikat oleh waktu dengan melihat semua waktu pada peubah yang digunakan dalam penelitian. Apabila nilai parameter $\widehat{\delta}_j < 0$ maka *hazard ratio* akan turun bersamaan dengan naiknya waktu. Ini mengakibatkan tidak konstannya *hazard ratio* yang juga berarti bahwa asumsi *proportional hazard* tidak terpenuhi untuk perluasan model ini.

3. METODE PENELITIAN

Penelitian ini menggunakan data SAKERNAS 2012 dengan menggunakan unit observasi penduduk Bali umur produktif (15–64 tahun) yang sedang mencari pekerjaan, selain itu pula penelitian ini hanya mencakup pengangguran yang berstatus mencari pekerjaan saja atau yang dikatakan pengangguran terbuka. Batasan masalah penelitian ini adalah fungsi waktu yang digunakan pada penambahan peubah terikat-waktu adalah $g(t) = t$.

Adapun langkah analisis pada penelitian ini adalah:

1. Melakukan estimasi terhadap perluasan model regresi Cox. Estimasi model ini dilakukan dengan menambahkan peubah

terikat-waktu pada dua peubah yang tidak berpengaruh secara signifikan dalam penelitian. Peubah yang tidak berpengaruh signifikan tersebut ialah peubah X_5 yaitu status dalam rumah tangga dan peubah X_7 yaitu umur. Sehingga perluasan model regresi Cox dapat dituliskan sebagai berikut:

$$\begin{aligned} h(t, \mathbf{x}(t)) &= h_0(t) \exp[\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \\ &\quad + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 \\ &\quad + \beta_7 x_7 + \delta_5 x_5 g(t) \\ &\quad + \delta_7 x_7 g(t)] \end{aligned}$$

2. Melakukan pengujian parameter secara simultan maupun secara parsial. Pengujian parameter ini bertujuan untuk melihat apakah peubah bebas dengan penambahan peubah terikat-waktu berpengaruh signifikan atau tidak terhadap model. Apabila dalam pengujian parameter terdapat peubah bebas yang terbukti berpengaruh signifikan terhadap peubah terikat maka peubah bebas tersebut dapat dimasukkan kedalam perluasan model regresi Cox. Statistik uji yang digunakan dalam pengujian parameter secara simultan adalah uji *likelihood ratio*,

$$\chi_{LR}^2 = -2 \ln L_{PH \text{ model}} - (-2 \ln L_{ext \text{ Cox model}}).$$

3. Menentukan perluasan model regresi Cox dengan melihat *p-value* dari seluruh peubah bebas yang ada, apabila *p-value* kurang dari taraf signifikansi 0,05 maka peubah bebas tersebut dapat masuk kedalam perluasan model regresi Cox

4. HASIL DAN PEMBAHASAN

Model awal perluasan regresi Cox dengan penambahan peubah terikat-waktu pada penelitian ini adalah sebagai berikut:

$$\begin{aligned} h(t, \mathbf{x}(t)) &= h_0(t) \exp[\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \\ &\quad + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 \\ &\quad + \delta_5 x_5 g(t) + \delta_7 x_7 g(t)] \end{aligned}$$

dengan penambahan peubah terikat-waktu pada peubah x_5 yaitu status dalam rumah tangga dan x_7 yaitu umur. Fungsi waktu yang digunakan dalam perluasan model regresi Cox ini adalah $g(t) = t$.

Untuk menghitung nilai dugaan perluasan model regresi Cox dan nilai *likelihood ratio* digunakan pengujian secara serentak pada perluasan model regresi Cox dapat menggunakan bantuan program SAS 9.2. Pengujian parameter peubah terdiri dari dua jenis, yaitu:

1. Pengujian Peubah Bebas Terhadap Waktu

Pengujian parameter dengan peubah bebas terhadap waktu bertujuan untuk melihat apakah peubah bebas yang tidak terikat-waktu memiliki pengaruh atau tidak terhadap peubah terikat.

Pengujian Secara Simultan

Uji simultan bertujuan untuk mengetahui pengaruh peubah bebas dengan peubah terikat secara serentak atau sekaligus. Hipotesis yang digunakan dalam uji simultan dari parameter peubah bebas yang tidak terikat-waktu adalah sebagai berikut:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1: \text{minimal ada satu dari } \beta_i \neq 0, \text{ dengan } i = 1, 2, \dots, p$$

Statistik uji yang digunakan dalam pengujian ini $\chi_{LR}^2 = -2(\ln L_{dsr} - \ln L_{lkp})$, dengan L_{dsr} merupakan nilai *likelihood* pada model dasar dan L_{lkp} adalah nilai dengan model lengkap. Apabila $\chi_{LR}^2 > \chi_{p,\alpha}^2$ dengan p merupakan jumlah dari parameter dan α adalah taraf signifikansi sebesar 0,05 maka keputusan tolak H_0 [2].

Pengujian Secara Parsial

Uji parsial yang digunakan ialah uji Wald dengan pengujian ini peubah bebas yang tidak mempunyai pengaruh signifikan akan dikeluarkan dari model. Hipotesis untuk uji Wald ialah sebagai berikut:

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0, \text{ untuk } i = 1, 2, \dots, p$$

Statistik uji yang digunakan adalah sebagai berikut: $\chi_W^2 = \left[\frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \right]^2$. Jika nilai

$\chi_W^2 > \chi_{1,\alpha}^2$ dengan derajat bebas 1 maka keputusan tolak H_0 .

2. Pengujian Peubah Bebas Terikat-Waktu

Pada umumnya konsep pengujian parameter peubah bebas terikat-waktu dan parameter peubah bebas terhadap waktu adalah sama. Pengujian secara simultan dan secara parsial sama-sama menggunakan uji *likelihood ratio* dan uji Wald.

Pengujian Secara Simultan

Perumusan hipotesis untuk uji *likelihood ratio* adalah sebagai berikut:

$$H_0: \delta_1 = \delta_2 = \dots = \delta_p = 0$$

$$H_1: \text{paling sedikit ada satu parameter } \delta_j \text{ yang tidak sama dengan 0, dengan } j = 1, 2, \dots, p$$

Statistika uji yang digunakan adalah sebagai berikut:

$$\chi_{LR}^2 = -2 \ln L_{PH \text{ model}} - (-2 \ln L_{ext \text{ Cox model}})$$

Apabila hasil dari $\chi_{LR}^2 > \chi_p^2$ dengan p merupakan jumlah dari parameter δ yang digunakan maka keputusan ialah tolak H_0 .

Pengujian Secara Parsial

Hipotesis untuk uji Wald adalah sebagai berikut:

$$H_0: \delta = \delta_j = 0$$

$$H_1: \delta \neq \delta_j \neq 0 \text{ dengan } j = 1, 2, \dots, p$$

Pengujian secara parsial pada perluasan model regresi Cox menggunakan statistik uji dan kriteria uji yang sama dengan model regresi Cox *proportional hazard*. Statistik uji yang

$$\text{digunakan ialah: } \chi_W^2 = \left[\frac{\hat{\delta}_j}{SE(\hat{\delta}_j)} \right]^2.$$

Selanjutnya, nilai estimasi parameter perluasan model regresi Cox dan *likelihood ratio* dapat dihitung dengan bantuan program SAS 9.2. Nilai *likelihood ratio* digunakan untuk menguji parameter perluasan model regresi Cox secara serentak. Pengujian serentak pada perluasan model regresi Cox dapat dilihat dari nilai χ_{LR}^2 yaitu sebesar 92,1720 yang lebih besar dari $\chi_{11;0,05}^2$ yaitu 19,675 sehingga dapat disimpulkan bahwa

secara serentak peubah bebas memiliki pengaruh terhadap peubah terikat.

Selanjutnya akan dilakukan pengujian pada peubah terikat-waktu dengan hipotesis $H_0: \delta_5 = 0$ melawan $H_1: \delta_5 \neq 0$ untuk peubah status rumah tangga terikat-waktu, dan hipotesis $H_0: \delta_7 = 0$ melawan $H_1: \delta_7 \neq 0$ untuk peubah umur terikat-waktu. Taraf signifikansi yang digunakan ialah $\alpha = 0,05$ dengan statistik uji untuk peubah status rumah tangga terikat-waktu $x_5g(t)$ adalah:

$$\chi_W^2 = \left[\frac{\hat{\delta}_5}{SE(\hat{\delta}_5)} \right]^2$$

sedangkan, untuk pengujian parameter peubah umur terikat-waktu $x_7g(t)$ memiliki statistik uji:

$$\chi_W^2 = \left[\frac{\hat{\delta}_7}{SE(\hat{\delta}_7)} \right]^2$$

Jika nilai $\chi_W^2 > \chi_{1;0,05}^2 = 3,841$ atau jika $p\text{-value} < \alpha = 0,05$ maka keputusan tolak H_0 . Hasil perhitungan dari program SAS 9.2 untuk estimasi parameter perluasan model regresi Cox menunjukkan bahwa peubah status rumah tangga yang terikat waktu memiliki $p\text{-value}$ lebih besar dari taraf signifikansi 0,05 yaitu sebesar 0,1746 sehingga keputusan gagal menolak H_0 dan ditarik kesimpulan bahwa peubah status rumah tangga yang terikat-waktu tidak berpengaruh terhadap lama mencari kerja.

Peubah umur yang terikat-waktu memiliki $p\text{-value}$ kurang dari 0,05 yaitu sebesar 0,0136 ini menandakan bahwa peubah umur terikat-waktu memiliki pengaruh terhadap lama mencari kerja di provinsi Bali. Ini juga menandakan bahwa memang peubah umur tidak memenuhi asumsi *proportional hazard*, sehingga ketika dilakukan pengujian dengan metode regresi Cox *proportional hazard* peubah umur dinyatakan tidak berpengaruh terhadap lama mencari kerja.

Selanjutnya akan ditentukan bentuk dari perluasan model regresi Cox dengan melihat $p\text{-value}$ dari masing-masing peubah. Terdapat empat peubah bebas yang memiliki $p\text{-value}$ kurang dari taraf signifikansi 0,05. Empat peubah bebas tersebut adalah peubah daerah

tempat tinggal, peubah jenis kelamin, peubah status perkawinan, dan peubah umur terikat-waktu. Sehingga secara matematis perluasan model regresi Cox yang didapat dapat dituliskan sebagai berikut:

$$h(t, x(t)) = h_0(t) \exp[\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \delta_7 x_7 g(t)]$$

$$h(t, x(t)) = h_0(t) \exp[-0,28271 x_1 - 0,32952 x_2 + 1,51686 x_3(\text{cerai}) + 0,37016 x_3(\text{kawin}) - 0,00547 x_7 g(t)]$$

dengan $h_0(t)$ merupakan *baseline hazard* yang tidak perlu ditentukan, peubah x_1, x_2, x_3 , dan $x_7g(t)$ berturut-turut adalah daerah tempat tinggal, jenis kelamin, status perkawinan dan peubah umur terikat-waktu. Interpretasi dari hasil perluasan model regresi Cox adalah:

1. Peubah daerah tempat tinggal memiliki parameter regresi yang bernilai negatif yaitu $-0,28271$ dan nilai *hazard ratio* 0,754. Ini menandakan bahwa individu yang tinggal di kota memiliki kesempatan mendapatkan pekerjaan 0,754 kali lebih lama dibandingkan dengan individu yang tinggal di desa
2. Peubah jenis kelamin dengan kategori laki-laki sebagai referensi, memiliki parameter regresi negatif dengan nilai 0,32952 dan nilai *hazard ratio* 0,719. Ini menandakan individu berjenis kelamin perempuan memiliki kesempatan mendapatkan pekerjaan sebesar 0,719 kali lebih lama dibandingkan dengan laki-laki.
3. Peubah status kawin memiliki parameter regresi positif pada kategori cerai dan kategori sudah kawin. Peubah status kawin dengan kategori cerai memiliki *hazard ratio* sebesar 4,558 dan *hazard ratio* sebesar 1,448 pada peubah status kawin dengan kategori sudah kawin. Ini berarti individu yang telah cerai memiliki kesempatan mendapatkan pekerjaan 4,558 kali dibandingkan dengan individu yang belum kawin. Sedangkan untuk individu yang sudah kawin memiliki kesempatan mendapatkan pekerjaan 1,448 kali dibandingkan dengan individu yang belum

kawin. Hal ini berarti bahwa individu yang memiliki status belum kawin memiliki masa mencari kerja lebih lama dari pada individu dengan status sudah kawin.

4. Peubah umur dengan nilai *hazard ratio* 0,995, yang menandakan bahwa penambahan umur satu tahun memiliki kesempatan mendapatkan pekerjaan sebesar 0,995 kali dari umur sebelumnya. Dengan kata lain, individu dengan umur setahun lebih muda akan lebih cepat mendapatkan pekerjaan.

5. SIMPULAN

Perluasan model regresi Cox untuk data sintasan lama mencari kerja di Provinsi Bali 2012 adalah

$$h(t, \mathbf{x}(t)) = h_0(t) \exp[-0,28271x_1 - 0,32952x_2 + 1,51686x_3(\text{cerai}) + 0,37016x_3(\text{kawin}) - 0,00547x_7g(t)]$$

dengan $h_0(t)$ merupakan *baseline hazard* yang tidak perlu ditentukan, peubah x_1 yaitu daerah tempat tinggal, peubah x_2 yaitu jenis kelamin, peubah x_3 yaitu status perkawinan, dan peubah $x_7g(t)$ yaitu faktor umur dengan terikat-waktu.

Peubah status rumah tangga terikat-waktu tidak berpengaruh signifikan terhadap lama mencari kerja, sedangkan dengan peubah umur terikat-waktu peubah umur terikat-waktu berpengaruh signifikan terhadap lama mencari kerja.

DAFTAR PUSTAKA

- [1] Collett, D., 2003. *Modelling Survival Data in Medical Research*. 2nd ed. London: Chapman and Hall.
- [2] Kleinbaum, D.G. & Klein, M., 2012. *Survival Analysis A Self-Learning Text*. 3rd ed. New York: Springer.
- [3] Lee, E.T. and Wang, John. 2003. *Statistical Methods for Survival Data Analysis*. Third Edition. New York: John Wiley and Sons.
- [4] Setiawan, Satrio Adi. 2010. Pengaruh Umur, Pendidikan, Pendapatan, Pengalaman dan Jenis Kelamin

Terhadap Lama Mencari Kerja Bagi Tenaga Kerja Terdidik Di Kota Magelang. *Skripsi*. Universitas Diponegoro. Semarang.