

BAB III

METODOLOGI PENELITIAN

3.1 Pengumpulan Data

3.1.1 Metode Pemilihan Sample

Populasi sample yang diikutsertakan dalam penelitian ini adalah nasabah atau debitur dari perusahaan pembiayaan sepeda motor PT XXX. Karena tujuan dari penelitian ini adalah untuk menguji *Survival Analysis* pada kredit yang dikucurkan oleh perusahaan pembiayaan. Hanya saja tidak seluruh nasabah kredit dari PT XXX yang diikutsertakan dalam penelitian ini. Karena periode pengujian model ini memerlukan waktu yang cukup panjang (sepanjang masa kredit tersebut), maka nasabah yang dimasukkan dalam penelitian ini adalah nasabah yang membeli motor dari 1 Juli 2004 dan berakhirnya waktu kredit sebelum 31 Desember 2007.

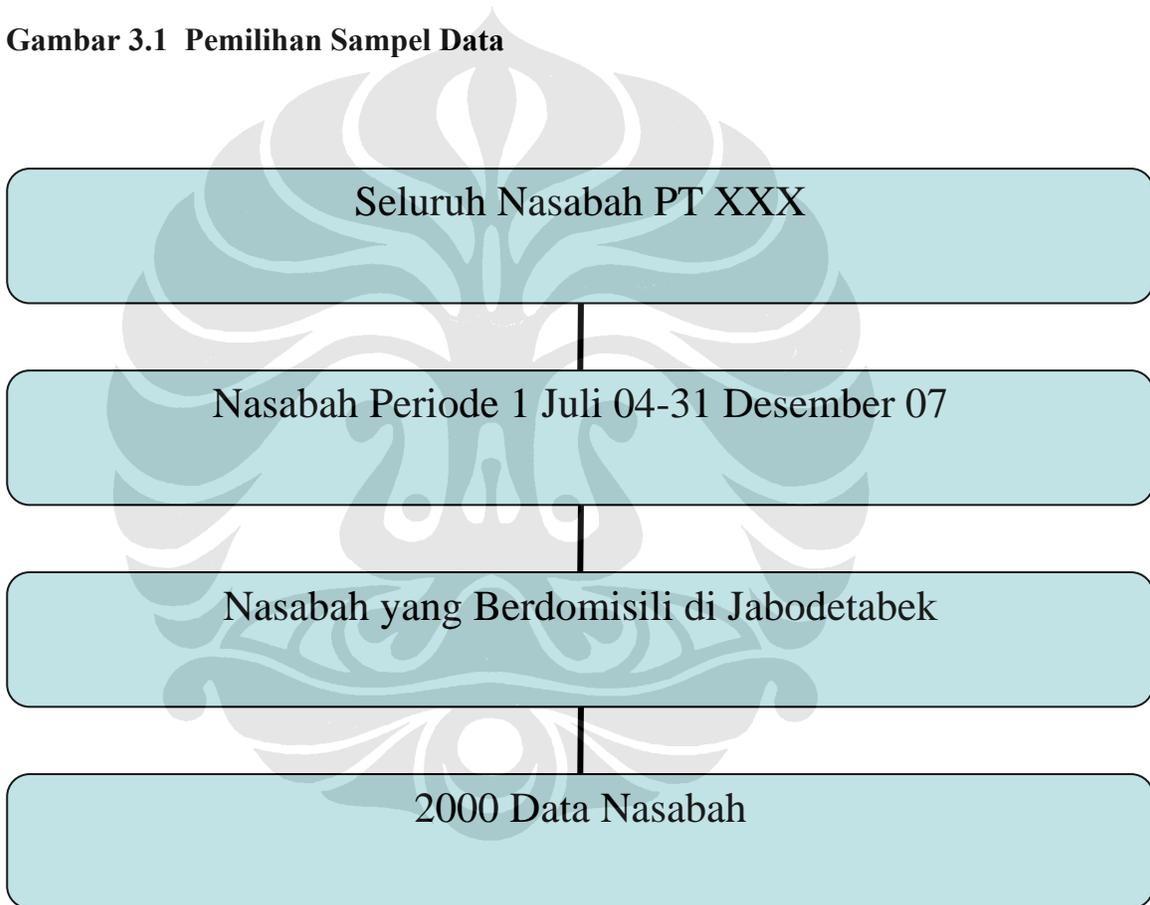
Perusahaan yang menjadi sample penelitian dipilih berdasarkan kriteria sebagai berikut :

- Sampel termasuk sebagai nasabah kredit PT XXX
- Sampel telah terdaftar sebagai nasabah PT XXX setelah 1 Juli 2004 dan berakhirnya masa kredit sebelum 31 Desember 2007
- Sampel yang diambil adalah nasabah PT XXX di wilayah JABODETABEK
- Sampel yang diambil adalah sebanyak 2000 data yang merupakan nasabah PT XXX yang berdomisili di wilayah JABODETABEK

3.1.2 Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder dan bersifat kuantitatif. Seluruh data diambil dari PT XXX yang meliputi data nasabah, suku bunga yang berlaku, Down Payment dan data-data lainnya.

Gambar 3.1 Pemilihan Sampel Data



3.2 Model Penelitian

Survival analysis ini digunakan pada umumnya di bidang medis dan ilmu biologi, tetapi teknik ini juga mulai digunakan secara luas pada bidang ilmu sosial, ekonomi, dan juga teknik. Pada bidang ekonomi kita akan mempelajari *survival* dari bisnis yang baru atau waktu *survival* dari produk seperti otomotif. Pada penelitian ini, penulis mencoba meneliti *survival* kredit dari pembiayaan kredit kendaraan bermotor.

Adapun langkah-langkah dari *Survival analysis* adalah:

1. *Life table*
2. *Kaplan-Meier survival function estimation*
3. *Log rank test* untuk membandingkan *survival* pada masing-masing kelompok di dalam *covariate*
4. *Cox regression model* bersama untuk mengetahui hubungan *covariate* yang signifikan dengan *survival* dan mendapat permodelan yang tepat

3.2.1 Life Table Analysis

Langkah yang paling cepat untuk menjelaskan *survival* pada sampel adalah dengan menggunakan *Life Table*. Teknik *Life Table* adalah metode yang paling tua untuk menganalisa data *Survival (failure time)*. Tabel ini bisa dianggap sebagai table frekuensi distribusi. Distribusi dari *Survival Time* dibagi menjadi beberapa interval. Pada masing-

masing interval kita hitung jumlah dan proporsi dari objek yang “hidup” dan jumlah dan proporsi dari kejadian yang gagal di dalam rentang interval tersebut.

Berdasarkan jumlah dan proporsi tersebut, beberapa statistik tambahan bisa dihitung yaitu :

- Jumlah Kejadian yang Beresiko
- Proporsi Gagal
- Proporsi Sukses
- Akumulasi Proporsi Sukses (*Survival Function*)
- Probabilitas Kepadatan
- Hazard Rate
- Median Waktu *Survival*
- Ukuran Sample

Jumlah Kejadian yang Beresiko. Adalah jumlah kasus yang ada di dalam masing-masing interval

Proportion Gagal. Adalah proporsi yang dihitung sebagai rasio antara kejadian yang gagal pada interval tersebut, dibagi dengan jumlah kejadian yang beresiko di interval tersebut.

Proportion Sukses. Proporsi ini dihitung sebagai 1 dikurangi Proporsi yang gagal

Akumulasi Proporsi Sukses (*Survival Function*). Adalah akumulasi dari proporsi dari kejadian yang sukses di dalam interval tersebut. Karena probabilitas dari *Survival*

diasumsikan independen terhadap interval, probabilitas ini dihitung dengan mengkalikan probabilitas yang sukses pada masing-masing interval. Hasil dari fungsi tersebut disebut juga *Survivorship* atau *Survival Function*

Probabilitas Kepadatan. Adalah estimasi probabilitas gagal pada rentang interval, dihitung pada masing-masing waktu :

$$F_i = (P_i - P_{i+1}) / h_i \dots \dots \dots (12)$$

F_i : adalah probabilitas kepadatan pada interval i

P_i : adalah estimasi akumulasi proporsi sukses pada awal interval i

P_{i+1} : adalah akumulasi proporsi sukses pada akhir interval i

h_i : adalah rentang interval tersebut

Hazard Rate. *Hazard rate* didefinisikan sebagai probabilitas per masing-masing waktu. Misalnya seseorang yang sudah sukses di awal rentang interval tersebut akan gagal pada interval tersebut. Secara spesifik itu diperhitungkan sebagai jumlah kegagalan pada kurun waktu dalam rentang interval tersebut dibagi dengan rata-rata jumlah kejadian yang sukses pada nilai tengah interval.

Median Survival Time. Adalah waktu *survival* dimana akumulasi *Survival Function* sama dengan 0.5. Perlu diketahui bahwa nilai median untuk akumulasi *Survival Function* biasanya tidak sama dengan jumlah dalam waktu dimana 50% sampel sukses. (Ini adalah pengecualian jika tidak ada *Censored Observation* di dalam waktu ini).

Ukuran Sampel. Untuk memperhitungkan secara seksama pada tiga fungsi mayor (*survival, probability density, dan hazard*) dan *standard error* pada masing-masing interval, minimum sampel yang digunakan adalah 30.

3.2.2 Kaplan-Meier Product-Limit Estimator

Ada langkah mudah daripada mengklasifikasi observasi data ke *Life Table* yaitu dengan mengestimasi *Survival Function* langsung dari waktu *continous survival* atau gagal. Dengan mengkalikan probabilitas *survival* pada interval-interval akan kita dapatkan *Survival Function* sebagai berikut :

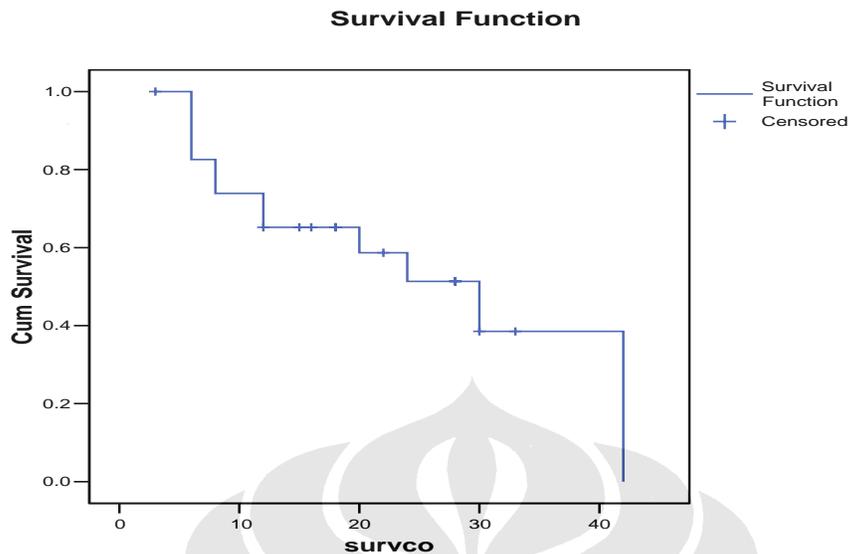
$$S(t) = \prod_{j=1}^t [(n-j)/(n-j+1)]^{\delta(j)} \dots\dots\dots(13)$$

$S(t)$: adalah estimasi *Survival Function*

n : adalah jumlah kejadian

\prod : adalah multiplikasi (jumlah geometri) pada seluruh kejadian kurang dari sama dengan t

$\delta(j)$: adalah konstanta dimana 1 apabila kejadian j *uncensored* dan 0 apabila itu kejadian itu *censored*. Penghitungan *Survival Function* ini disebut juga *Product-limit estimator*, dan diperkenalkan pertama kali oleh Kaplan dan Meier (1958) Sebagai contoh dapat dilihat seperti yang ditunjukkan dibawah ini :



- Keuntungan dari metode Kaplan- Meier *Product Limit Estimator* dibandingkan dengan metode *Life Table* adalah ketika menganalisa data *survival* dan *failure* adalah hasil penghitungan tidak tergantung pada data yang dikelompokkan (pada sejumlah waktu interval). Apabila observasi hanya satu, maka metode *Product Limit* dan metode *Life Table* akan sama.

3.2.3 Log Rank Test

*Log rank test*¹ adalah metode yang paling populer untuk melihat survival di dalam suatu grup. Caranya adalah dengan membandingkan estimasi *hazard function* dari grup yang diobservasi dalam waktu tertentu. Log rank test ini dapat dinotasikan sebagai berikut :

$$E_j = O_j \frac{N_{1j}}{N_j} \text{ dan variance } V_j = \frac{O_j(N_{1j}/N_j)(1 - N_{1j}/N_j)(N_j - O_j)}{N_j - 1} \dots(14)$$

¹ Bland JM, Altman DG. Survival probabilities. The Kaplan-Meier method. *BMJ* 1998

Logrank statistik membandingkan setiap O_{1j} dengan expektasi E_j dalam *null hypothesis* sehingga dapat dinyatakan sebagai berikut :

$$Z = \frac{\sum_{j=1}^J (O_{1j} - E_j)}{\sqrt{\sum_{j=1}^J V_j}} \dots\dots\dots(15)$$

Dengan *log rank test* akan didapat *hazard* dan *ratio* grup di dalam masing-masing *covariate* dan akan diketahui grup mana yang mempunyai *hazard* dan resiko yang terbesar dan terkecil.

3.2.4 Cox Proportional Hazard Model

Proportional hazard model adalah model regresi yang paling umum karena tidak berdasarkan asumsi yang fokus pada *nature* atau bentuk dari *survival distribution* yang ada. Model ini mengasumsikan *underlying hazard rate* adalah fungsi dari variabel independen (*covariate*). Tidak ada asumsi yang dibuat mengenai *nature* atau bentuk dari fungsi *hazard*. *hazard rate* dalam *regression model* dimodelkan sebagai berikut :

Adapun untuk *proportional hazard function* dengan *time varying covariates* dapat dirumuskan sebagai berikut :

$$\log h_i(t) = \alpha(t) + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} \dots\dots\dots(16)$$

atau

$$h_i(t) = h_0(t) \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}) \dots \dots \dots (17)$$

Model inilah yang disebut *proportional hazard model*. Model ini adalah *semi-parametric* karena *baseline hazard* bisa menggunakan bentuk apapun, dan *covariates* memasuki model secara linear.

Cox's regression model dapat dianggap sebagai metode nonparametrik. Model ini dapat diformulasikan sebagai :

$$h\{(t), (z_1, z_2, \dots, z_m)\} = h_0(t) * \exp(b_1 * z_1 + \dots + b_m * z_m) \dots \dots \dots (18)$$

Dimana :

- $h(t, \dots)$ = merupakan *resultant hazard*
- (z_1, z_2, \dots, z_m) = *covariates* dari kasus-kasus
- T = *survival time*
- $h_0(t)$ = *baseline hazard*; adalah *hazard* untuk individu ketika variabel independen sama dengan 0

Dengan begini kita dapat melinearkan model ini dengan membagi kedua persamaan tersebut dengan $h_0(t)$ dan menggunakan logaritma pada kedua persamaan tersebut sehingga menjadi :

$$\log[h\{(t), (z\dots)\}/h_0(t)] = b_1 * z_1 + \dots + b_m * z_m \dots \dots \dots (19)$$

Sekarang kita memiliki model linear yang lebih mudah dan dapat diestimasi

Asumsi : Ketika tidak ada asumsi yang dibuat mengenai bentuk dari *underlying hazard function*, persamaan model diatas menunjukkan dua asumsi. Yang pertama adalah mereka menjelaskan hubungan yang *multiplicative* antara *underlying hazard function* dan *Log Linear function* dari covariate. Asumsi ini disebut juga sebagai *Proportionality assumption*. Pada prakteknya, ini mengasumsikan bahwa dua buah observasi dengan nilai yang berbeda untuk variabel independenn, rasio dari *hazard function* untuk dua buah observasi tersebut tidak tergantung pada waktu. Asumsi yang kedua adalah ada *Log Linear Relationship* antara variabel independent dengan *underlying hazard function*.

Exponential Regression

Pada dasarnya, model ini mengasumsikan bahwa *Survival Time Distribution* adalah eksponensial dan *contigent* terhadap nilai dari sekelompok variabel independent (z_i). *Exponential Distribution* bisa dirumuskan sebagai berikut :

$$S(z) = \exp(a + b_1 * z_1 + b_2 * z_2 + \dots + b_m * z_m) \dots\dots\dots(20)$$

$S(z)$: menjelaskan waktu survival

a : adalah konstanta

b_i : adalah *regression parameters*

Goodness-of-fit. *Chi Square goodness of fit* value menghitung fungsi dari *log-likelihood* untuk model dengan parameter diestimasi ($L1$) dan *log-likelihood* dari model dimana seluruh *covariate* menjadi angka 0 (zero; $L0$). Apabila *Chi square value* signifikan, maka kita menolak *hypothesis* dan mengasumsikan bahwa variabel-variabel independen adalah signifikan terhadap waktu *survival*

Standard exponential order statistic. Satu cara untuk mengecek asumsi exponential dari model ini adalah dengan mem-plot *residual survival time* terhadap standar exponential statistic *theta*. Apabila asumsi exponential tercapai, maka seluruh nilai dari plot ini akan membentuk sebuah garis lurus.

Normal and Log-Normal Regression

Pada model ini, diasumsikan bahwa *survival time (log survival time)* dihasilkan dari distribusi normal. Model yang dihasilkan pada dasarnya akan identik dengan *multiple regression model* pada umumnya dan bisa diformulasikan sebagai berikut :

$$t = a + b_1 * z_1 + b_2 * z_2 + \dots + b_m * z_m \dots\dots\dots(21)$$

t : adalah *survival times*

3.3 Definisi Variabel Penelitian

Dari penelitian tentang *survival analysis* ini, variabel yang digunakan oleh penulis ada beberapa yaitu :

1. Sex : yaitu jenis kelamin dari nasabah kredit tersebut. Pada variabel ini dibagi menjadi 2 kategori yaitu pria dan wanita.
2. Marital : yaitu status perkawinan dari nasabah tersebut. Untuk karegori variabel ini dibagi menjadi 3 yaitu belum menikah, menikah, dan cerai
3. City : yaitu membagi wilayah domisili nasabah ini. Karena data yang diambil adalah nasabah yang berdomisili di wilayah Jabodetabek, maka ini di wilayah Jabodetabek ini dibagi menjadi 8 zona yaitu untuk membagi wilayah Jakata Pusat, Jakarta Utara, Jakarta Barat, Jakarta Timur, Jakarta Selatan, wilayah Tangerang dan sekitarnya, wilayah Bogor dan sekitarnya, dan wilayah Bekasi hingga Cikampek
4. DP : DP atau down payment adalah sejumlah uang muka yang dibayarkan oleh nasabah kepada perusahaan pembiayaan. Di dalam usaha kredit, biasanya DP itu ditetapkan besarnya dibandingkan dengan harga produk tersebut, missal DP 10 persen dari harga unit dan sebagainya. Untuk itu variabel ini dikategorikan menjadi 6 bagian yaitu DP $\leq 10\%$, DP 10-20%, DP 20-30%, DP 30-40%, DP 40-50%, dan DP $\geq 50\%$.
5. Gaji : Gaji atau pendapatan sangat besar pengaruhnya terhadap lancar atau tidaknya kredit yang dibayarkan oleh konsumen. Apabila gaji atau pendapatan semakin kecil, maka seseorang akan semakin sulit untuk membayar cicilan kreditnya dan begitu juga sebaliknya. Karena itu penulis memasukan variabel ini untuk diteliti. Untuk variabel gaji,

dikategorikan menjadi 4 yaitu : ≤ 3 juta rupiah per bulan, 3 juta-6 juta rupiah per bulan, 6 juta-10 juta rupiah per bulan, dan diatas 10 juta rupiah per bulan

6. Bunga : Bunga adalah besaran biaya yang harus dibayarkan atas pinjaman seseorang. Di dalam pemberian kredit kepada konsumen, PT XXX menetapkan perlakuan tingkat suku bunga yang berbeda-beda antara konsumen yang satu dengan yang lainnya. Hal itu dilihat dari besarnya DP, semakin besar DP maka suku bunga yang ditetapkan akan semakin kecil dan begitu juga sebaliknya. Perusahaan juga melihat faktor-faktor resiko lainnya yang dihadapi oleh perusahaan terhadap pemberian kredit kepada konsumen, oleh karena hal itu maka suku bunga yang ditetapkan berbeda-beda. Untuk bunga, penulis mengkategorikannya menjadi 4 yaitu : suku bunga 10%-13%/tahun, 13%-16% per tahun, 16%-19% per tahun dan $\geq 19\%$ per tahun

7. Usia : Usia seseorang juga berpengaruh terhadap resiko dari kredit yang diberikan oleh perusahaan kepada konsumen. Misalnya bila usia seseorang terlalu tua atau terlalu muda, dikhawatirkan orang tersebut tidak mampu membayar cicilan dari kredit yang diambilnya. Karena itu penulis memasukannya sebagai variabel dalam penelitian ini. Penulis mengkategorikan usia ini menjadi 9 yaitu : usia ≤ 20 tahun, 20-25 tahun, 25-30 tahun, 30-35 tahun, 35-40 tahun, 40-45 tahun, 45-50 tahun, 50-55 tahun dan ≥ 55 tahun.

8. Tenor : Tenor atau jangka waktu kredit yang diambil oleh seseorang untuk mencicil pinjamannya juga berpengaruh terhadap resiko yang dihadapi oleh perusahaan. Oleh karena itu penulis memasukannya ke dalam variabel yang akan diteliti.
9. Harga : Harga unit dimasukan ke dalam variabel yang berpengaruh di dalam analisa ini karena, dengan mengetahui harga serta besaran uang muka yang dibayarkan oleh konsumen maka kita dapat mengetahui berapa besar ratio antara uang muka yang dibayarkan terhadap harga. Dengan begitu, dapat dikategorikan di dalam kategori DP
10. Status : yang dimaksud status di dalam variabel ini adalah status kredit dari masing-masing nasabah. Untuk itu dikategorikan lagi menjadi 2 yaitu : status kredit yang berhasil dan yang gagal.
11. Waktu macet: Jangka waktu gagal atau macetnya suatu kredit dari masing-masing nasabah yang gagal bayar. Dengan ini dapat kita ketahui pada masing-masing nasabah dengan segala variabelnya masing-masing mengalami gagal bayar pada bulan ke berapa. Dengan diketahui waktu macetnya, diharapkan penelitian dapat memberikan hasil yang lebih detail lagi.

3.4 Prosedur Pengolahan Data

Tahapan metodologi penelitian ini adalah :

1. Mengkonversi data-data yang ada yaitu data nasabah PT XXX yang mempunyai variabel-variabel sex, marital, city, DP, gaji, bunga, usia, tenor, harga, status, waktu macet menjadi variabel independent dan dependen dengan program Microsoft Excel
2. Setelah data-data ditentukan jenisnya, data lalu dipindahkan ke dalam software SPSS 15. Lalu memasukan variabel tenor sebagai variabel dependen, dan variabel lainnya sebagai variabel independent.
3. Data lalu diolah dengan metode *Likelihood ratio test of model* dengan plot survival, hazard, lalu *Cox Regression* untuk menentukan variabel apa yang paling dominant
4. Setelah didapat variabel yang paling dominan, maka variabel yang lain dilakukan penelitian juga untuk mengetahui penghitungan masing-masing variabel. Adapun caranya adalah dengan memasukan variabel itu secara silih berganti
5. Melakukan regressi lagi untuk terakhir kalinya, dengan memasukan data yang aktual tanpa dikategorikan untuk mengetahui permodelan yang paling tepat
6. Melakukan intepretasi dan pembahasan hasil penelitian

Hal yang perlu diperhatikan adalah SPSS 15 seperti program software statistik lainnya menyediakan estimasi dan standard error untuk misalnya parameter β_j pada *Cox proportional hazard model*. Untuk aplikasi *Cox Regression*, maka yang pertama dilakukan di dalam software SPSS adalah pertama *Analyse*, lalu *Survival* dan pada akhirnya *Cox Regression*.

Ada beberapa hal yang perlu dilihat dari hasil pengolahan data tersebut, yaitu :

- *Hazard ratio*, disebut juga *odds ratio* atau $Exp(B)$. *Hazard ratio* adalah probabilitas dari suatu event yang muncul di waktu $t+1$, dimana *survival* adalah waktu t . *Hazard ratio* dapat dinotasikan sebagai berikut :

$$RR = h(t)/ho(t) = [ho(t)] e^B / ho(t) \dots \dots \dots (22)$$

ho = baseline hazard

Hazard ratio 1.0 menunjukkan variabel dari model tidak mempunyai dampak pada waktu dimana suatu event atau kejadian itu ada pada waktu t . Semakin *hazard ratio* dibawah 1.0, semakin baik *covariate* tersebut, semakin kecil kejadian itu muncul. Semakin besar *hazard ratio* diatas 1.0, semakin besar kemungkinan variabel mengakibatkan suatu kejadian (misal kematian, gagal bayar, dll)

- *Confidence intervals on Exp(B)* adalah output dari SPSS dan software statistik lainnya, memberikan tingkat terendah dan tertinggi 95% tingkat keyakinan. Apabila nilai 1.0 terletak diantara tingkat keyakinan tersebut, seseorang tidak bisa 95% yakin bahwa *covariate* tersebut mempunyai dampak, dan harus mengkategorikannya sebagai tidak signifikan.
- *Likelihood ratio test of the model*, atau disebut juga *omnibus test* atau $-2LL$ atau $-2 \log likelihood$. Apabila $-2LL$ signifikan, maka permodelan secara keseluruhan adalah signifikan. Yaitu apabila Sig. > .05 (social science standar), maka paling tidak 1 dari *covariate* mempunyai kontribusi yang signifikan terhadap penjelasan

dari waktu terhadap suatu kejadian. Ini juga bisa berarti permodelan ini lebih signifikan dibandingkan dengan *null model*, dimana *time-only model* ketika seluruh *covariates* adalah 0.

- *Regression coefficients.* Seluruh software statistik menampilkan *regression coefficient (B)* untuk masing-masing *covariate*, *standard error* dari *B (SE)*, *Wald test significance value*, *the degrees of freedom (df)*, dan *significance value of the coefficient*. Apabila *Sig.* > .05 (*social science standard*), maka *covariate effect* tidak bisa diasumsikan bukan angka 0. Apabila *Sig (Wald)* < .05, maka peneliti menarik kesimpulan bahwa variabel ini berguna untuk permodelan. Positif *regression coefficients* berarti *covariate* meningkatkan *hazard*, apabila *negative coefficients* berarti menurunkan *hazard*. Perlu diingat bahwa -2LL tes lebih diutamakan daripada *Wald test* ketika menguji permodelan secara keseluruhan.
- *Baseline cumulative hazard* adalah *hazard rate* untuk *time-only model* ketika seluruh *covariates* = 0. *Baseline cumulative hazard rates* lebih mudah diinterpretasikan ketika data *covariate numerical* sudah distandarisasi, menyebabkan *covariate* menjadi 0, dengan *baseline hazard* diinterpretasikan sebagai *time-only rates* untuk orang-orang pada rata-rata *covariate*.