

BAB 2

TINJAUAN PUSTAKA

2.1. Diabetes Melitus

Diabetes melitus merupakan suatu kelompok penyakit metabolik dengan karakteristik hiperglikemia yang terjadi karena kelainan sekresi insulin, kerja insulin, atau keduanya. Diabetes adalah penyakit serius kronis yang terjadi baik ketika pankreas tidak menghasilkan cukup insulin atau ketika tubuh tidak dapat secara efektif menggunakan insulin yang dihasilkan. Insulin adalah hormon yang mengatur gula darah, atau glukosa (*World Health Organization*, 2016).

Menurut *American Diabetes Association* (ADA) 2010, DM merupakan suatu kelompok penyakit metabolik dengan karakteristik hiperglikemia yang terjadi karena kelainan sekresi insulin, kerja insulin atau kedua-duanya. Lebih dari 90% dari semua populasi diabetes adalah diabetes mellitus tipe 2 yang di tandai dengan penurunan sekresi insulin karena berkurangnya fungsi sel beta pankreas secara progresif yang di sebabkan oleh resistensi insulin. Diabetes mellitus dapat dikategorikan menjadi empat tipe yaitu diabetes mellitus tipe 1, diabetes mellitus tipe 2, diabetes mellitus gestational dan diabetes mellitus tipe lain yang di sebabkan oleh faktor-faktor lain (Kemer and bruckel, 2014)

2.2. Analisis Survival

Analisis survival atau analisis ketahanan hidup adalah analisis data yang berhubungan dengan waktu, mulai dari awal sampai terjadinya suatu peristiwa khusus (Collett,2003). Menurut Kleinbaum & Klein, sebagaimana dikutip oleh Iskandar (2015:10), Analisis survival telah menjadi alat penting untuk menganalisis

data waktu antar kejadian (*time to event data*) atau menganalisis data yang berhubungan dengan waktu, mulai dari *time origin* sampai terjadinya suatu peristiwa khusus. Kejadian khusus (*failure event*) tersebut dapat berupa kegagalan, kematian, kambuhnya suatu penyakit, respon dari suatu percobaan, atau peristiwa lain yang dipilih sesuai dengan kepentingan peneliti. Peristiwa khusus tersebut dapat berupa kejadian positif seperti kelahiran, kelulusan sekolah, kesembuhan dari suatu penyakit.

Fungsi-fungsi dari pada distribusi waktu survival merupakan suatu fungsi yang menggunakan variabel random waktu survival. Variabel random biasa dinotasikan dengan T . Untuk T suatu variabel acak positif dan menunjukkan waktu survival setiap subjek, maka nilai-nilai yang mungkin untuk T yaitu $T \geq 0$. Menurut Lee dan Wang (2003), distribusi dari T dapat dinyatakan dalam 3 cara yaitu sebagai berikut:

a. Fungsi kepadatan peluang (*fungsi density*)

Fungsi kepadatan peluang atau PDF (Probability Density Function) adalah peluang suatu individu mati atau mengalami kejadian sesaat dalam interval waktu t sampai $t + \Delta t$. Fungsi padat peluang $f(t)$ dirumuskan sebagai berikut (Lee dan Wang, 2003)

$$f(t) = \lim_{\Delta t \rightarrow 0} \left[\frac{P(\text{kegagalan individu dalam interval})}{\Delta t} \right]$$

$$f(t) = \lim_{\Delta t \rightarrow 0} \left[\frac{P(t \leq T < (t+\Delta t))}{\Delta t} \right] = \lim_{\Delta t \rightarrow 0} \left[\frac{F(t+\Delta t) - F(t)}{\Delta t} \right] \quad (2.1)$$

Jika T merupakan variabel acak positif interval $(0, \infty)$, maka $F(t)$ merupakan

fungsi distribusi komulatif kontinu dari T . Didefinisikan sebagai peluang suatu individu mengalami kejadian kurang dari sama dengan waktu t , yaitu :

$$F(t) = P(T \leq t) = \int_0^t f(x) dx$$

$$F(t) = \frac{d(\int_0^t f(x) dx)}{dt} = f(t) \quad (2.2)$$

b. Fungsi ketahanan hidup (*fungsi survival*)

Fungsi survival (fungsi ketahanan hidup) dinotasikan dengan $S(t)$ yang menunjukkan probabilitas suatu individu bertahan hidup lebih dari waktu t , dimana $t > 0$. $S(t)$ didefinisikan :

$$S(t) : P(\text{individu bertahan hidup lebih dari waktu } t)$$

$$S(t) : P(T > t)$$

$$S(t) : 1 - P(\text{individu gagal atau mati sebelum waktu } t)$$

$$S(t) : 1 - P(T > t) = 1 - F(t)$$

Fungsi ketahanan hidup $S(t)$ memiliki sifat sebagai berikut :

$$S(t) = \begin{cases} 1, & \text{untuk } t = 0 \\ 0, & \text{untuk } t = \infty \end{cases}$$

Yang artinya, peluang individu dapat bertahan hidup pada waktu nol adalah 1 dan pada saat waktu tak terbatas seorang individu dapat bertahan nol.

c. Fungsi kegagalan (*fungsi hazard*)

Fungsi Hazard adalah peluang suatu individu mati dalam interval waktu dari t sampai $t + \Delta t$, jika diketahui individu tersebut masih dapat bertahan hidup sampai dengan waktu t . Fungsi Hazard secara matematika dinyatakan sebagai :

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t \leq t)}{\Delta t} = \frac{f(t)}{S(t)} \quad (2.3)$$

2.3. Data Tersensor

Data tersensor merupakan data yang tidak dapat teramati secara utuh, karena adanya individu yang hilang ataupun dengan alasan yang lain, sehingga tidak dapat diambil datanya sampai akhir penelitian. Dengan kata lain, pada akhir pengamatan individu tersebut belum mengalami *failure event*. Berbeda ketika sampai akhir pengamatan individu tersebut telah mengalami *failure event*, maka individu tersebut tidak tersensor (Collet, 2003).

Menurut Lee Wang (2003) Data tersensor merupakan data yang tidak bisa diamati secara utuh, karena adanya individu yang hilang ataupun dengan alasan lain, sehingga tidak dapat diambil datanya sampai akhir pengamatan. Dengan kata lain, pada akhir pengamatan individu tersebut belum mengalami peristiwa tertentu dalam keadaan sebaliknya maka data tersebut disebut data tidak tersensor. Tiga penyebab data dikatakan tersensor antara lain :

1. *Loss to follow up*, yaitu subjek menghilang selama masa pengamatan, misal subjek pindah atau bisa saja menolak untuk diamati.
2. Subjek tidak mengalami kejadian penelitian.
3. Subjek terpaksa dihentikan dari pengamatan karena meninggal sebelum pengamatan berakhir atau karena faktor lain.

Menurut Kleinbaum dan Klein (2005), dalam analisis survival terdapat empat jenis penyensoran yaitu :

1. Penyensoran Kanan (*Right Censor*)

Penyensoran ini terjadi jika objek pengamatan atau individu yang diamati masih tetap hidup pada saat waktu yang telah ditentukan. Dengan kata lain, individu tersebut belum mengalami kejadian failure event sampai akhir periode pengamatan, sedangkan waktu awal dari pengamatan dapat di amati secara penuh.

2. Penyensoran Kiri (*Left Censor*)

Penyensoran ini terjadi jika semua informasi yang diinginkan diketahui dari seseorang inidvidu telah diperoleh pada awal pengamatan. Dengan kata lain, pada saat waktu awal pengamatan individu tidak teramati pada awal pengamatan sementara kejadian dapat teramati secara penuh sebelum penelitian berakhir.

3. Penyensoran Selang (*Interval Censoring*)

Penyensoran ini terjadi jika informasi yang dibutuhkan telah dapat diketahui pada kejadian peristiwa di dalam selang pengamatan atau penyensoran yang waktu daya tahannya berada dalam suatu selang tertentu. Sebagai contoh, jika dalam catatan rekam medis menunjukkan bahwa pada usia 45 tahun pasien penyakit ginjal kronis dalam contoh kondisinya masih sehat dan belum terdiagnosa penyakit ginjal kronis, kemudian pasien melakukan tes pertama pada saat umur 50 tahun dan terdiagnosis terkena penyakit penyakit ginjal kronis, dengan demikian umur saat terdiagnosa penyakit ginjal kronis adalah antara 45 sampai 50 tahun.

4. Penyensoran Acak (*Random Censoring*)

Penyensoran ini terjadi jika individu yang diamati meninggal atau mengalami kejadian karena sebab yang lain, bukan disebabkan dari tujuan awal penelitian.

2.4. Kaplan Meier Estimator

Menurut Hanni dan Wuryandari (2013), metode *kaplan meier* digunakan untuk menaksir fungsi survival dan fungsi hazard. Metode ini disebut juga metode nonparametrik karena tidak membutuhkan asumsi distribusi dari waktu survival.

a. Taksiran waktu survival

Misalkan terdapat n individu dengan waktu survival yaitu t_1, t_2, \dots, t_n .

Beberapa pengamatan ini tersensor jika terdapat r waktu failure diantara n individu, dimana $r \leq n$, maka waktu failure ke- j ditunjukkan sebagai $t_{(j)}$, untuk $j = 1, 2, \dots, r$, dimana $k \leq r$. Estimasi fungsi survival pada waktu ke- k adalah :

$$\hat{S}(t) = \prod_{j=1}^k \frac{n_j - d_j}{n_j} \quad (2.4)$$

Dengan, n_j : jumlah nasabah yang berisiko gagal (tidak mampu membayar) pada $t_{(j)}$

d_j : jumlah nasabah yang gagal (tidak mampu membayar) pada waktu t_1, t_2, \dots, t_n .

b. Taksiran Fungsi Hazard

Menaksir fungsi hazard dari waktu survival menggunakan rasio jumlah failure terhadap jumlah individu yang berada pada risiko failure. Apabila d_j merupakan jumlah individu pada $t_{(j)}$, waktu survival ke- j dan n_j adalah individu yang berisiko failure pada waktu $t_{(j)}$. Maka estimasi fungsi hazard adalah :

$$\hat{h}(t) = \frac{d_j}{n_j} \quad (2.5)$$

Fungsi survival memiliki hubungan dengan fungsi hazard yaitu pada fungsi hazard kumulatif. Nilai taksiran dari fungsi survival dapat digunakan untuk mencari nilai fungsi hazard kumulatif yaitu,

$$\hat{H}(t) = -\log \hat{S}(t) \quad (2.6)$$

2.5. Regresi Weibull

Fungsi ketahanan hidup $S(t)$ adalah peluang dari ketahanan hidup dalam waktu t untuk distribusi *weibull* dirumuskan sebagai berikut :

$$S(t_i) = \exp\left(-\left(\frac{t_i}{\lambda_i}\right)^\gamma\right) \quad (2.7)$$

Fungsi kepadatan peluang distribusi *weibull* untuk *likelihood* :

$$f(t_i) = -\frac{dS(t)}{dt} = \frac{\gamma}{\lambda_i} t_i^{\gamma-1} \exp\left(-\left(\frac{t_i}{\lambda_i}\right)^\gamma\right) \quad (2.8)$$

Pada model tahan hidup parametrik λ diparameter ulang dalam variabel bebas dan parameter regresi. Sedangkan parameter γ (*Shape* parameter) dibuat tetap. Model regresi *weibull* adalah :

$$\hat{\lambda} = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_j X_j) \quad (2.9)$$

Keterangan :

$\beta_0, \beta_1, \beta_2, \dots, \beta_j$: Koefisien parameter

X_1, X_2, \dots, X_j : Variabel bebas

Setelah diperoleh regresi *Weibull* dan estimasi fungsi ketahanan hidup, maka dapat

diperoleh estimasi fungsi *hazard* pada regresi *Weibull* yaitu :

$$h(t_i) = \frac{\gamma}{a_i^\gamma} t_i^{\gamma-1} \quad (2.10)$$

2.6. Regresi *Cox Proportional Hazard*

Cox Proportional Hazard merupakan model regresi yang digunakan untuk melihat faktor-faktor yang menyebabkan terjadinya suatu peristiwa biasa dikenal dengan nama (*time-dependent covariate*) dengan peubah respon adalah waktu ketahanan hidup. Model regresi *Cox* merupakan model regresi yang menyatakan tingkat *hazard* (risiko) dari individu dengan karakteristik tertentu yang disebut kovariat (Cox dan Oakes, 1984). Salah satu tujuan model *Cox Proportional Hazard* adalah untuk memodelkan hubungan antara waktu *survival* dengan variabel-variabel yang diduga mempengaruhi waktu *survival*. Melalui model *cox* dapat dilihat hubungan antara variabel bebas (variabel independen) terhadap variabel terikat (variabel dependen) yaitu waktu *survival* melalui fungsi *hazard*. Risiko kematian individu pada waktu tertentu bergantung pada nilai x_1, x_2, \dots, x_p dari p variabel bebas X_1, X_2, \dots, X_p . Himpunan nilai variabel bebas pada model *cox* dipresentasikan oleh vektor x , sehingga

$x = (x_1, x_2, \dots, x_p)$. Diasumsikan X merupakan variabel bebas yang independen terhadap waktu. Model *cox* dapat dituliskan sebagai berikut (Kleinbaum Klein.2005) :

$$h(t, x) = h_o(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) \quad (2.11)$$

dengan memisalkan :

$h_0(t)$: fungsi dasar hazard

$\beta_1, \beta_2, \dots, \beta_p$: parameter regresi

x_1, x_2, \dots, x_p : nilai dari variable bebas X_1, X_2, \dots, X_p .

Rumus model cox pada persamaan (2.5) memiliki sifat bahwa jika semua $X = 0$, maka rumus tereduksi menjadi fungsi hazard dasar (baseline hazard) $h_0(t)$. Dengan demikian $h_0(t)$ dianggap sebagai awal atau dasar dari fungsi hazard, dapat dituliskan sebagai berikut:

$$\begin{aligned} h(t, x) &= h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) \\ &= h_0(t) \exp(\beta_1 \cdot 0 + \beta_2 \cdot 0 + \dots + \beta_p \cdot 0) \\ &= h_0(t) \exp(0) \\ &= h_0(t)(1) \\ h(t, x) &= h_0(t) \end{aligned} \quad (2.12)$$

Hazard ratio merupakan ukuran untuk mengetahui tingkat risiko (kecenderungan) yang dapat dilihat dari perbandingan antara individu dengan kondisi variabel bebas X pada kategori sukses dengan kategori gagal. Misalnya, variabel bebas X dengan dua kategori yaitu 0 dan 1. Nilai tersebut mempunyai arti bahwa tingkat kecepatan terjadinya failure event pada individu dengan kategori $X = 0$ adalah sebesar kali dari individu dengan kategori $X = 1$ (Fa'rifah dan Purhadi, 2012). *Hazard ratio* untuk individu dengan $X = 0$ dibanding $X = 1$ adalah sebagai berikut (Kleinbaum dan Klein, 2005):

$$\widehat{HR} = \frac{h(t|X_1^*)}{h(t|X_1)} = \frac{h_0(t) \cdot \exp(\beta_1 X_1^*)}{h_0(t) \cdot \exp(\beta_1 X_1)} = \exp[\beta_1 (X_1^* - X_1)] \quad (2.13)$$

Tingkat hazard dari fungsi tersebut bersifat proporsional. Jika pada suatu persamaan bernilai 2 pada titik tertentu, maka risiko kegagalan individu $X = 0$ dua kali lebih besar dari individu $X = 1$.

2.7. Uji Goodness of fit

Setiap variabel bebas harus memenuhi asumsi proporsional sehingga dilakukan pengecekan asumsi dengan menggunakan model *cox proportional hazard*. Dalam penelitian ini untuk pengujian asumsi dilakukan dengan uji *goodness of fit* dengan langkah-langkah pengujian sebagai berikut :

- Menggunakan model cox untuk mendapatkan residual *schoenfeld* setiap variabel bebas, dengan rumus :

$$\hat{r}_{ji} = \delta_i (\hat{x}_{ij} - \hat{a}_{ji}) \text{ dengan } \hat{a}_{ji} = \frac{\sum_{i \in R(t_i)} x_{ji} e^{\beta x_i}}{\sum_{i \in R(t_i)} e^{\beta x_i}} \quad (2.14)$$

2.8. Efron Partial Likelihood

Nurjanah (2015) menguraikan metode dengan pendekatan *efron partial likelihood* ini merupakan metode yang sedikit lebih intensif tingkat komputasinya dibandingkan dengan metode *breslow*. dan metode ini juga memberikan hasil estimasi yang besar jika data kejadian bersama atau *ties* nya dalam ukuran yang besar. Klein dan Moeschberger (2003) menguraikan pendekatan *efron* secara umum memiliki bentuk persamaan sebagai berikut :

$$L(\beta_{\text{Efron}}) = \prod_{i \in D} \frac{\exp(\beta S_k)}{\prod_{i=1}^{d_k} \left[\sum_{i \in R_{t_i}} \exp(\beta X_i) - \frac{j-1}{d_k} \right] \sum_{i \in D_k} \exp(\beta X_i)} \quad (2.15)$$

2.9. Pemilihan Model Terbaik

Akaike's Information Criterion (AIC) adalah salah satu metode yang berguna untuk menemukan model terbaik yang ditemukan oleh *Akaike*. Besarnya

AIC dapat ditentukan dari persamaan berikut :

$$AIC = -2l(\beta) + 2df \quad (2.16)$$

Dengan $l(\beta)$ adalah fungsi $\log(\text{likelihood})$ dan $2df$ adalah derajat bebas yang digunakan dalam model. Model regresi ataupun distribusi terbaik adalah model regresi yang memiliki nilai AIC terkecil. Kelebihan pada AIC terletak pada pemilihan model regresi terbaik untuk tujuan (*forecasting*) yaitu dapat menjelaskan kecocokan model dengan data yang ada (faturahman,2009).

2.10. Uji Multikolinearitas

Multikolinearitas terjadi apabila antar variabel bebas saling berkorelasi dalam satu model data yang bersifat kategori, multikolinearitas dapat dideteksi menggunakan melihat nilai vif. Dengan pernyataan hipotesis yang akan diuji adalah sebagai berikut :

H_0 : tidak terjadi multikolinearitas.

H_1 : terjadi multikolinearitas.

Uji kebebasan di dasarkan pada besaran :

$$\chi^2 = \sum i \frac{(o_i - e_i)^2}{e_i} \quad (2.17)$$

χ^2 adalah nilai bagi peubah acak, χ^2 yang sebaran penarikan contohnya sangat menghampiri sebaran chi-square. Lambang o_i menyatakan frekuensi teramati dan lambang e_i menyatakan frekuensi harapan. Taraf nyata sebesar α , nilai kritiknya χ^2_α dapat diperoleh dari tabel nilai kritik s ebaran chi-square $\chi^2 > \alpha$ maka tolak H_0 bahwa kedua penggolongan ini bebas pada taraf nyata α dan jika selainnya maka H_0 diterima.

2.11. Uji Signifikansi Parameter

Uji signifikansi parameter digunakan untuk mengetahui. Jika ada tidaknya hubungan parameter di dalam model regresi. Uji signifikansi dilakukan secara serentak maupun parsial.

a. Uji Signifikansi secara serentak

Uji serentak dilakukan untuk signifikansi parameter model regresi secara bersama-sama. Prosedur pengujian parameter secara serentak adalah sebagai berikut:

1. Membuat Hipotesis

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

$$H_1: \text{tidak semua } \beta_k = 0, \text{ paling tidak ada satu } \beta_k \neq 0 \text{ untuk } k = 1, 2, \dots, p$$

(Kutner et al., 2004) atau

H_1 : Variabel X_1, X_2, \dots, X_k secara serentak berpengaruh terhadap model

H_0 : Variabel X_1, X_2, \dots, X_k secara serentak tidak berpengaruh terhadap model

2. Taraf signifikan (α) = 0.05

3. Menentukan statistik uji yaitu menggunakan uji rasio likelihood

$$G^2 = -2 \ln \frac{L(\hat{\omega})}{L(\hat{\Omega})} = -2 [\ln L(\hat{\Omega}) - \ln L(\hat{\omega})] \quad (2.18)$$

4. Menentukan daerah kritis (penolakan H_0)

Tolak H_0 jika $G^2_{hitung} > X^2_{p,\alpha}$ atau $p - value < \alpha$

5. Kesimpulan

b. Uji signifikansi parsial

Uji signifikansi secara parsial digunakan untuk mengetahui *covariate* yang berpengaruh terhadap model regresi. Prosedur pengujian parameter secara parsial adalah sebagai berikut:

1. Membuat hipotesis

$$H_0: \beta_k = 0$$

$$H_1: \beta_k \neq 0, \text{ untuk } k = 1, 2, \dots, p - 1 \text{ (kutner et al, 2004)}$$

H_0 : variabel bebas ke $- k$ tidak berpengaruh terhadap model

H_1 : variabel bebas ke

$- k$ berpengaruh terhadap model untuk k

$= 1, 2, \dots, p - 1$

2. Taraf signifikansi (α) = 0.05

3. Menentukan statistik uji yaitu menggunakan uji wald :

$$W^2 = \left(\frac{\hat{\beta}_k}{SE(\hat{\beta}_k)} \right)^2 \quad (2.19)$$

4. Menentukan daerah kritis (penolakan H_0)

Tolak H_0 jika $W^2 > X^2_{p,\alpha}$ atau $p - value < \alpha$

5. Kesimpulan