BAB II

TINJAUAN PUSTAKA

2.1 Tinjauan Non Statistik

2.1.1 Definisi Kanker Payudara

Kanker payudara adalah tumor ganas pada jaringan payudara. Jaringan payudara terdiri dari kelenjar susu (kelenjar pembuat air susu), saluran kelenjar (saluran air susu), dan jaringan penunjang payudara. Oleh Word Health Organization (WHO) penyakit ini juga dimasukkan ke dalam International Classification of Disease (ICD) dengan kode 174-175 . Menurut Sutriston (1992) kanker payudara adalah neoplasma spesifik tempat terlazim perempuan yang merupakan penyebab utama kematian perempuan akibat kanker.

Sedangkan menurut Luwia (2005) kanker payudara adalah kanker yang berasal dari kelenjar, saluran kelenjar ,dan jaringan penunjang payudara. Dari pendapat tersebut maka peneliti dapat menyimpulkan bahwa kanker payudara adalah perubahan sel-sel yang mengalami pertumbuhan tidak normal dan tidak terkontrol pada payudara.Kanker payudara ditemukan di seluruh dunia. Tahun 2003, insidens kanker payudara di Belanda 91 per 100.000 penduduk, Amerika 71,7 per 100.000 penduduk, Swiss 70 per 100.000 wanita, Australia 83,2 per 100.000

penduduk, Kanada 84,7, Indonesia 26 per 100.000 wanita pada tahun 2002 dan Jepang 16 per 100.000 penduduk. Semua wanita memiliki risiko terkena kanker payudara. Kanker payudara juga bisa menyerang pria dengan perbandingan 1 : 100 antara pria dengan wanita.

Kanker payudara lebih sering dijumpai pada umur 40-49 tahun yaitu sebesar 30,35%. Menurut Sukardja yang dikutip oleh Arlinda (2002) di Amerika frekuensi kanker payudara tertinggi ditemukan pada umur 40-50 tahun. Demikian juga di Jepang yaitu sebesar 40,6% kanker payudara ditemukan pada umur 40-49 tahun dan jarang pada umur kurang dari 30 tahun. Kanker payudara terjadi karena adanya kerusakan pada gen yang mengatur pertumbuhan dan diffrensiasi sehingga sel-sel itu tumbuh dan berkembang biak tanpa dapat dikendalikan. Penyebaran kanker payudara terjadi melalui pembuluh getah bening dan tumbuh di kelenjar getah bening, sehingga kelenjar getah bening aksila ataupun supraklavikula membesar. Kemudian melalui pembuluh darah kanker menyebar ke organ lain seperti paru-paru, hati ,jantung, pembuluh darah serta pusat system syaraf otak.

2.1.2 Penyebab Kanker Payudara

Penyebab kanker payudara sampai saat ini belum diketahui. Penyebab kanker payudara termasuk multifaktorial yaitu banyak faktor yang terkait satu dengan yang lainnya. Beberapa faktor risiko yang mempengaruhi timbulnya kanker payudara adalah :

a. Usia

Risiko utama kanker payudara adalah bertambahnya usia. Berdasarkan penelitian *American Cancer Society* tahun 2006 diketahui usia lebih dari 40 tahun mempunyai risiko yang lebih besar untuk mendapatkan kanker payudara yakni 1 per 68 penduduk . Risiko ini akan bertambah seiring dengan pertambahan usia yakni menjadi 1 per 37 penduduk usia 50 tahun, 1 per 26 penduduk usia 60 tahun dan 1 per 24 penduduk usia 70 tahun. Data American Cancer Society (2007) melaporkan 70% perempuan didiagnosa menderita kanker payudara di atas usia 55.

b. Jenis Kelamin

Kanker payudara lebih banyak ditemukan pada wanita. Pada pria juga dapat terjadi kanker payudara, namun frekuensinya jarang hanya kira-kira 1% dari kanker payudara pada wanita.

c. Riwayat Reproduksi

Riwayat reproduksi dihubungkan dengan banyak paritas ,umur melahirkan anak pertama dan riwayat menyusui anak. Wanita yang tidak mempunyai anak atau yang melahirkan anak pertama di usia lebih dari 30 tahun berisiko 2-4 kali lebih tinggi daripada wanita yang melahirkan pertama di bawah usia 30 tahun. Wanita yang tidak menyusui anaknya mempunyai risiko kanker payudara 2 kali lebih besar. Kehamilan dan menyusui mengurangi risiko wanita untuk terpapar dengan hormon estrogen terus.

d. Riwayat Kanker Individu

Penderita yang pernah mengalami infeksi atau operasi tumor jinak payudara berisiko 3-9 kali lebih besar untuk menderita kanker payudara.Penderita tumor jinak payudara seperti kelainan fibrokistik berisiko 11 kali dan penderita yang mengalami operasi tumor ovarium mempunyai risiko 3-4 kali lebih besar.

e. Riwayat Kanker Keluarga

Secara genetik, sel-sel pada tubuh individu dengan riwayat keluarga menderita kanker sudah memiliki sifat sebagai embrio terjadinya sel kanker. Menurut *Sutjipto* (2000) , kemungkinan terkena kanker payudara lebih besar 2 hingga 4 kali pada wanita yang ibu dan saudara perempuannya mengidap penyakit kanker payudara

f. Menstruasi cepat dan Menopause lambat

Wanita yang mengalami menstruasi pertama pada usia kurang dari 12 tahun dan wanita yang mengalami masa menopausenya terlambat lebih dari 55 tahun berisiko 2,5 hingga 5 kali lebih tinggi daripada wanita yang menstruasi pada usia normal . Wanita yang menstruasi pertama di usia kurang dari 12 tahun dan wanita yang mengalami masa menopause terlambat akan mengalami siklus menstruasi lebih lama sepanjang hidupnya yang mengakibatkan keterpaparan lebih lama dengan hormon estrogen.

g. Pajanan Radiasi

Wanita yang terpapar penyinaran (radiasi) dengan dosis tinggi di dinding dada berisiko 2 hingga 3 kali lebih tinggi.

h. Obesitas dan Konsumsi makanan lemak tinggi

Wanita yang mengalami kelebihan berta badan (obesitas) dan individu dengan konsumsi tinggi lemak berisiko 2 kali lebih tinggi terkena kanker payudara dari yang tidak obesitas dan yang tidak sering mengkonsumsi makanan tinggi lemak. Risiko ini terjadi karena jumlah lemak yang berlebihan dapat meningkatkan kadar estrogen dalam darah sehingga akan memicu pertumbuhan sel-sel kanker.

2.1.3 Metode Pengobatan Kanker Payudara

Terdapat beberapa metode dalam pengobatan pada kanker payudara yang sebaiknya disesuaikan dengan kondisi kanker yang diderita oleh pasien. Selain disesuaikan pada kondisi pasien juga berdasarkan persetujuan baik dari keluarga maupun pasien itu sendiri. Metode pengobatan kanker paru ada beberapa yakni sebagai berikut.

a. Pembedahan

Definisi dari pembedahan menurut Promkes RSUD Tugurejo Semarang (2015) adalah tindakan pengangkatan tumor melalui operasi. Namun besar kemungkinan masih ada sel – sel kanker yang tertinggal sehingga biasanya dilanjutkan dengan kemoterapi.

b. Radioterapi

Radioterapi atau terapi sinar menurut Promkes RSUD Tugurejo Semarang (2015) adalah pengobatan kanker dengan menggunakan energi pengion dan non pengion yang ditujukan untuk membunuh sel – sel kanker payudara.

c. Kemoterapi

Kemoterapi menurut Promkes RSUD Tugurejo Semarang (2015) adalah pemberian obat – obatan *neoplastic* yang bertujuan untuk membunuh sel– sel kanker. Biasanya pengobatan ini menggunakan dua obat yang satu berbasis plastina (baik *cisplatin* maupun*carboplatin*). Obat yang lain yang biasa digunakan adalah *gemcitabine, paclitaxel, docetaxel, pemetrexed, etoposide,* atau *vinorelbine.*

Kemoterapi tunggal untuk karsinoma paru yang sudah ditinggalkan dan diganti dengan kombinasi dua obat atau lebih, baik bersama – sama atau berturut – turut. Indikasi utamanya adalah karsinoma yang telah mengadakan *metastasis* dan memberikan keluhan.

2.2 Tinjauan Statistik

2.2.1 Analisis Deskriptif

Analisis deskriptif adalah suatu teknik mengumpulkan, mengolah, menyederhanakan, menyajikan serta menganalisis data kuantitatif secara deskriptif agar dapat memberikan gambaran yang teratur tentang suatu peristiwa ke dalam bentuk tabel dan grafik yang sesuai (Dajan, 1986). Dalam hal ini analisis deskriptif yang digunakan adalah melihat peforma dari pasien atau yang biasa disebut *performance status* dengan menggunakan Skor *karnofsky*. Menurut Karnofsky (1949) indeks skor *karnofsky* memudahkan pengklasifikasian pasien sesuai dengan keadaan gangguan fungsionalnya.

Pengklasifikasian tersebut memudahkan evaluasi hasil terapi, dan penilaian *prognosis* pasien. Skor *karnofsky* terdiri dari nilai 100 – 0, dengan nilai 100 adalah keadaan sehat tanpa gangguan, dan nilai 0 adalah mati. Sistem penilaian ini dinamakan berdasarkan penemunya, Dr. David A. Karnofsky yang pertama kali menjelaskan sistem penilaian ini pada tahun 1949.

1. 100% - normal, tanpa keluhan, tanpa gejala penyakit.
2. 90% - dapat menjalankan aktifitas sehari – hari, sedikit tanda dan gejala dari penyakit.
3. 80% - aktifitas sehari – hari normal namun sedikit kesulitan, dengan beberapa tanda dan gejala penyakit.
4. 70% - mampu merawat diri sendiri, tetapi tidak dapat beraktifitas normal atau bekerja.

5. 60% - butuh dampingan dan bantuan orang lain, masih dapat mengurus kebutuhan dasar pribadi.
6. 50% - membutuhkan lebih banyak bantuan orang lain dan perawatan medis.
7. 40% - terbatas pada tempat tidur dan kursi, membutuhkan perawatan medis khusus.
8. 30% - terbatas pada tempat tidur saja, tidak dapat mengurus diri sendiri.
9. 20% - sakit berat, membutuhkan banyak perawatan dan pengobatan.
10. 10% - keadaan kritis, perjalanan penyakit fatal *rapid progressive.*

0% - meninggal.

2.3 Penyensoran

Menurut Lawless (1982 :31) suatu data dikatakan tersensor jika lamanya hidup seseorang yang ingin diketahui atau diobservasi hanya terjadi pada waktu yang telah ditentukan (interval pengamatan), sedang info yang ingin diketahui tidak terjadi pada interval tersebut, dengan demikian tidak diperoleh informasi apapun yang diinginkan selama interval pengamatan. Tipe – tipe penyensoran menurut Johnson dalam Lawless (1982: 31) ada tiga yakni sebagai berikut.

1. Penyensoran Tipe I

Pada penyensoran sebelah kanan tipe I, penelitian diakhiri apabila waktu pengamatan yang ditentukan tercapai. Jika waktu pengamatan sama untuk semua unit maka dikatakan penyensoran tunggal. Jika waktu pengamatan untuk setiap unit berbeda maka dikatakan penyensoran ganda. Pada penyensoran sebelah kiri tipe I, pengamatan dilakukan jika telah melampaui awal waktu yang ditentukan.Karakteristik penyensoran tipe I adalah bahwa kegagalan adalah acak. Misalkan T_1, T_2, …, T_n adalah sampel random distribusi tahan hidup dengan fungsi kepadatan peluang $f(t)$, fungsi *survival* $S(t)$, sedangkan waktu sensor untuk semua T sama yaitu misalkan L, sampel demikian dikatakan sampel dengan waktu sensor tunggal. Akan tetapi pada umumnya untuk setiap T_i diberikan waktu sensor L_i. Semua komponen dikatakan terobservasi jika $T_i \leq L_i$ diperoleh variabel waktu dan variabel yang menunjukkan semua komponen telah mati yaitu t_i dan δ_i dengan

$$T_i = \min(T_i\ , L_i)\ dan\ \delta_i = \begin{cases} 1, jika\ T_i \leq L_i \\ 0, jika\ T_i > L_i \end{cases}.$$

Maka fungsi kepadatan peluangnya adalah

$$f(t_i)^{\delta i} S(L_i)^{1-\delta i} \qquad (2.3.1)$$

2. Penyensoran Tipe II

Pada penyensoran tipe II, pengamatan diakhiri setelah sejumlah kegagalan yang telah ditetapkan diperoleh, atau dapat dikatakan banyaknya kegagalan adalah tetap dan waktu pengamatan adalah acak. Pada sensor kanan jenis II, jumlah individu pada saat awal ditentukan dan waktu penelitian ditentukan sampai terjadinya kematian dengan jumlah tertentu.

Pada sensor kiri jenis II, titik awal penelitian dilakukan saat waktu kegagalan terurut. Data tersensor tipe II adalah suatu data waktu hidup yang terdapat r buah observasinya dalam sampel random yang berukuran n dengan ($1 \leq r \leq n$). Penyensoran tipe II lebih sering digunakan misalnya dalam data uji hidup, total n item ditempatkan pada tes, tetapi tidak semua n gagal, tes dihentikan jika observasi mengalami kegagalan ke r. Tes tersebut dapat menghemat waktu yang sangat lama untuk semua item gagal dalam beberapa kasus. Akan terlihat bahwa perlakuan statistik data tersensor tipe II, setidaknya langsung pada prinsipnya.Perlu ditekankan bahwa dengan sensor tipe II jumlah observasi r ditentukan sebelum data dikumpulkan. Secara formal, data terdiri dari r terkecil waktu hidup $T_{(1)} \leq T_{(2)} \leq \ldots \leq T_{(r)}$dari sampel acak n waktu hidup T_1, …, T_n dari distribusi hidup dalam pertanyaan. Jika T_1, …, T_1 i.i.d. dan memiliki distribusi kontinu dengan fkp $f(t)$ dan fungsi *survival* $S(t)$,maka hasil umum pada urutan statistik fkp bersama dari $T_{(1)}, \ldots, T_{(r)}$ adalah

$$\frac{n!}{(n-r)!} f\left(t_{(1)}\right) \ldots f\left(t_{(r)}\right) [S\left(t_{(r)}\right)]^{n-r} \quad (2.3.2)$$

3. Penyensoran Maju (*Progresive Censoring*)

Pada penyensoran maju, suatu jumlah yang ditentukan dari banyaknya unit – unit bertahan dikeluarkan dari penelitian berdasarkan kejadian dari tiap kegagalan terurut. Secara konseptual, hal ini sama dengan suatu praktek yang dikenal sebagai *sudden- death testing,* dimana tes secara serempak memuat beberapa pengetesan dan apabila terjadi kegagalan pertama keseluruh pengetesan dianggap gagal.

Tipe penyensoran menurut *Latan* (2014:306) terbagi menjadi tiga yakni sebagai berikut.

1. Tipe I Sensoring

   Observasi mempunyai waktu sensoring yang tetap. Tipe I sensoring diasumsikan tidak terjadi kecelakaan sehingga semua sensoring observasi sama dari awal sampai berakhirnya studi.

2. Tipe II Sensoring

   Peneliti akan menetapkan target yang harus dipenuhi di dalam desain studi. Tipe II sensoring diasumsikan tidak ada kecelakaan dan sensoring observasi sama dengan atau lebih besar dari *uncensored* observasi.

3. Tipe III Sensoring

   Menurut Sukestiyarno (2011: 66) analisis regresi adalah tidak jauh berbeda pengertiannya dengan analisis korelasi. Pada analisis korelasi hanya melihat hubungan antara variabel x dan y, dimana antara variabel x dan y berkedudukan sama artinya bias dipertukarkan antara yang satu mempengaruhi yang lain. Pada analisis, ingin melihat hubungan satu arah antar variabel yang lebih khusus dimana variabel x berfungsi sebagai variabel bebas yang mempengaruhi dan variabel y sebagai variabel terikat adalah yang terpengaruhi. Biasanya variabel x disebut sebagai variabel independent dan variabel y disebut sebagai variabel dependent. Menurut Latan (2014: 162) analisis regresi merupakan teknik analisis statistik yang digunakan untuk menguji hubungan antara satu variabel atau lebih variabel independent (prediktor) dengan satu variabel dependent (kriteria). Analisis regresi mempunyai dua tujuan utama yaitu untuk memprediksi dan untuk menganalisis hubungan kausal.

2.4 Regresi Tobit

Regresi tobit pertama kali diperkenalkan oleh James Tobin pada tahun 1958. Regresi Tobit merupakan analisis regresi yang digunakan untuk variabel terikat yang sebagian datanya memiliki skala pengukuran diskrit dan sebagian yang lain berskala kontinu. Greene (2008) menyebutkan bahwa variabel terikat yang bersifat mixture (campuran) memiliki struktur data dengan skala diskrit untuk yang bernilai nol, dan berskala kontinu untuk yang tidak bernilai nol. Data tersebut disebut juga data tersensor. Tersensor sendiri dalam hal ini dapat berarti nilai dari variabel terikat tersebut terkonsentrasi atau terkelompok pada satu nilai (Novianti, 1993 dalam Salim, 2007). Sebaran data tersensor adalah sebaran normal tersensor, yang mengikuti asumsi $N\ (\mu, \sigma^2)$.

Model regresi tersensor merupakan salah satu metode statistika yang dapat digunakan untuk menentukan model bila terjadi pembatasan pada variabel terikatnya. Pada model regresi tersensor beberapa nilai sampel dicatat sebagai nilai batas dari nilai yang sebenarnya. Data pengamatan pada variabel jenis ini mengelompok akibat adanya batas bawah (tersensor kiri), batas atas (tersensor kanan) atau dapat juga keduanya. Pembatasan tersebut dapat terjadi secara alamiah seperti beberapa nilai yang lebih dekat terhadap suatu nilai tertentu. Pembatasan juga dapat ditentukan oleh peneliti tergantung pada tujuan penelitiannya (Frone, 1997).

Adanya pembatasan terhadap suatu nilai tertentu terhadap variabel terikat y, sebut saja a, mengakibatkan distribusi data tersebut berubah. Jika suatu populasi telah diketahui berdistribusi normal, maka distribusi akibat adanya pemotongan nilai tertentu berubah menjadi distribusi normal tersensor, sehingga model menjadi:

$$y_i = \begin{cases} a\ , \ jika\ y_i^* \ \le\ a \\ y_i^*\ \ ,\ jika\ y_i^*\ >\ a \end{cases}$$

Sampel yang dihasilkan yaitu $y_1, y_2, \ldots\ldots\ldots y_n$ disebut sampel tersensor. Dalam model Tobit standar didefinisikan seperti masalah diatas dengan nilai $a = 0$. Formulasi model Tobit dalam Tobin (1958) dalam Greene (2008) secara umum adalah sebagai berikut.

$$y_i = \begin{cases} y_i^* \ , \ jika\ y_i^* \ > \ 0 \\ 0 \ , \ jika\ y_i^* \ \leq \ 0 \end{cases}$$

Dimana $i = 1,2, \ldots, T$ dan Y_i^* adalah variable terikat dengan persamaan sebagai berikut.

$$y_i^* \ = \ x_i^T \beta + u_i \qquad (2.4.1)$$

Dengan :

y_i^* : adalah nilai variabel terikat yang sebenarnya

x_i^T : $\begin{bmatrix} 1 & x_{1i} & x_{2i} & \ldots & x_{pi} \end{bmatrix}$ adalah vector variabel bebas

β : $[\beta_0 \quad \beta_1 \quad \ldots \beta_p]^T$ adalah vektor parameter, p merupakan banyaknya variabel

u_i : residual model yang mengikuti sebaran normal tersensor $(0, \sigma^2)$

2.4.1 Distribusi Normal Tersensor

Variabel tersensor didefinisikan sebagai berikut, misalkan y^* berdistribusi normal dengan mean μ dan varian σ^2.

$$y = \begin{cases} a, \; jika \; y^* \leq a \\ y^*, \; jika \; y^* > a \end{cases}$$

maka probabilitas tersensor $y = a$ bernilai

$$Prob\,(y = a) = Prob\,(y^* \leq a).$$

$$= \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi\sigma}} exp\left(-\frac{1}{2}(\frac{y^*-\mu}{\sigma})^2\right) dy^*$$

$$= \int_{-\infty}^{\frac{a-\mu}{\sigma}} \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{1}{2}z^2}\, dz$$

dengan nilai $z = \frac{y^*-\mu}{\sigma}$ dan $\frac{1}{\sigma}\, dy^* = dz$.

Sehingga Probabilitas $(y = a) = \Phi\left(\frac{a-\mu}{\sigma}\right)$, sedangkan untuk probabilitas tidak tersensor $y = y^*$ adalah

$$Prob\,(y = y^*) = Prob\,(y^* > a)$$

$$= 1 - Prob\,(y^* \leq a)$$

$$= 1 - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Menurut *Joreskog* (2002) fungsi densitas dari y adalah

$$f(y) = [\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}(\frac{y-\mu}{\sigma})^2}]^{1-j} \, [\Phi\left(\frac{a-\mu}{\sigma}\right)]^j$$

$$= [\frac{1}{\sigma}\phi\left(\frac{a-\mu}{\sigma}\right)]^{1-j} \, [\Phi\left(\frac{a-\mu}{\sigma}\right)]^j, \quad (2.4.2)$$

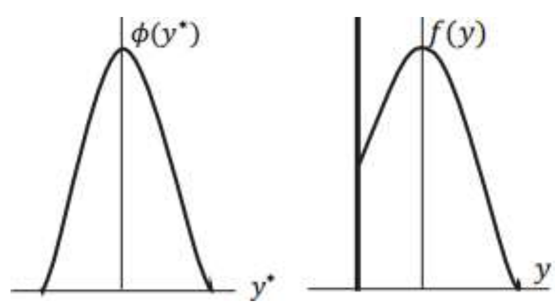
dengan $j = \begin{cases} 1, & jika\, y = a \\ 0, & jika\ lainnya \end{cases}$,

Maka parameter ϕ dan Φ masing-masing adalah fungsi densitas dan fungsi distribusi dari distribusi normal standar. Selanjutnya dari persamaan tersebut dapat diperoleh fungsi densitas untuk nilai $y = y^*$ .Selanjutnya diperoleh fungsi densitas untuk nilai $y = y^*$ atau nilai $y > a$ adalah

$$f(y^*) = \frac{1}{\sigma}\phi\left(\frac{y^*-\mu}{\sigma}\right) \quad (2.4.3)$$

dan densitas dari $y = a$ adalah :

$$f(a) = Prob\ (y = a) = \Phi\left(\frac{a-\mu}{\sigma}\right)$$



Gambar 2.1. Variabel Normal y^* dan variabel tersensor$y$

Pada Gambar di sebelah kiri menunjukkan bahwa data menyebar mengikuti distribusi normal, sedangkan pada gambar disebelah kanan, akibat adanya batas terhadap suatu nilai maka data menyebar mengikuti sebaran distribusi normal tersensor namun dengan total peluang tetap bernilai satu.

2.4.2 Fungsi Likelihood Model Tobit Standar

Fungsi Likelihood dari Model Tobit standar adalah:

$$L = \prod_{y_i=0}^{n_0} \left[1 - \Phi \left(\frac{x_i'\beta}{\sigma} \right) \right] \prod_{y_i>0}^{n_1} [\sigma^{-1} \phi(y_i - x_i'\beta)/\sigma] \qquad (2.4.4)$$

Dimana :

$\prod_{y_i=0}^{n_0} =$ adalah perkalian dari banyaknya pengamatan

dimana $y_i = 0$ atau $y_i^* \leq 0$

$\prod_{y_i>0}^{n_1} =$ adalah perkalian dari banyaknya pengamatan dimana $y_i > 0$ atau $y_i^* > 0$

misalkan :

n_0 = adalah banyaknya pengamatan dimana $y_i = 0$

n_1 = adalah banyaknya pengamatan dimana $y_i > 0$

Dengan $n_0 + n_1 = n$, $\phi(.)$ dan $\Phi(.)$ masing-masing menyatakan fungsi probabilitas densitas normal standar dan fungsi distribusi dari normal standar. Sehingga untuk $y_i = 0$ atau $y_i^* \leq 0$ didapatkan

$$Pr\,(Y = 0) = Pr\,(Y^* \leq 0)$$

$$Pr\,(Y = 0) = 1 - \Phi(x_i'\beta/\sigma)$$

Bila pemisalan yang dilakukan adalah $u = \sigma t$, dengan $\sigma > 0$ maka akan didapatkan $Pr\,(Y = 0) = \Phi(x_i'\beta/\sigma)$, untuk $y_i > 0$ atau $y_i^* > 0$ diperoleh :

$$Pr(Y, y_i > 0) = f(y_i^* | y_i^* > 0)\,\Pr(Y^* > 0)$$

$$= \frac{f(y_i^*)}{Pr(Y^* > 0)}\Pr(Y^* > 0)$$

$$= f(y_i^*)$$

Karena $Y^* \sim N(x_i'\beta, \sigma^2)$ dan juga berlaku $y_i = y_i{}^*$, maka

$$Pr(Y, y_i > 0) = \sigma^{-1}\phi((y_i - x_i'\beta)/\sigma)$$

Fungsi Likelihood untuk model Tobit Standar

$$L = \prod_{y_i=0}^{n_0} Pr(Y, y_i = 0) \prod_{y_i>0}^{n_1} Pr(Y, y_i > 0)$$

$$= \prod_{y_i=0}^{n_0} \left[1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right)\right] \prod_{y_i>0}^{n_1} \sigma^{-1}\phi\left[\frac{y_i - x_i'\beta}{\sigma}\right]$$

(2.4.5)

2.4.3 Estimasi Parameter Model Tobit

Dalam menduga parameter yang ada pada model tobit standar, yakni β dan σ^2 berdasarkan n pengamatan y_i ,tidak digunakan metode Least Squares karena kelinieran hubungan antara variabel terikat dengan variabel bebas tidak terpenuhi. Penggunaan metode Penaksiran Least Squares dalam menaksir parameter model tobit adalah bias. Oleh karena itu dalam menaksir parameter yang ada pada model tobit standar digunakan Maksimum Likelihood. Dengan mensubstitusikan persamaan Fungsi Likelihood dari Model Tobit standar dapat ditulis sebagai:

$$L = \prod_{y_i=0}^{n_0} \left[1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right)\right]$$

$$\prod_{y_i>0}^{n_1}\left[\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(y_i-x_i'\beta)^2}\right]$$

(2.4.6)

Jadi Fungsi Likelihoodnya adalah sebagai berikut :

$$\ln L = \sum_{y_i=0}^{n_0}\ln\left[1-\Phi\left(\frac{x_i'\beta}{\sigma}\right)\right] + \sum_{y_i>0}^{n_1}\ln\left[\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(y_i-x_i'\beta)^2}\right]$$

(2.4.7)

Dimana :

$\sum_{y_i=0}^{n_0}$ = adalah adalah penjumlahan dari banyaknya pengamatan dimana $y_i = 0$

$\sum_{y_i>0}^{n_1}$ = adalah adalah penjumlahan dari banyaknya pengamatan dimana $y_i > 0$, misalkan :

$n_0 =$ adalah banyaknya pengamatan dimana $y_i = 0$

$n_1 =$ adalah banyaknya pengamatan dimana $y_i > 0$

dengan $n_0 + n_1 = n$

$$= \sum_{y_i>0}^{n_1} \ln \left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y_i - x_i'\beta)^2}\right]$$

$$= -n_1 \ln \sigma - \frac{n_1}{2} \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{y_i>0}^{n_1} (y_i - x_i'\beta)^2$$

$$= -\frac{n_1}{2} \ln \sigma^2 - \frac{n_1}{2} \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{y_i>0}^{n_1} (y_i - x_i'\beta)^2$$

Sehingga diperoleh persamaan

$$\ln L = \sum_{y_i>0}^{n_1} \ln \left[1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right) - \frac{n_1}{2} \ln \sigma^2 - \frac{n_1}{2} \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{y_i>0}^{n_1} (y_i - x_i'\beta)^2\right]$$

Untuk mencari nilai ekstrimnya:

$$\frac{\partial \ln L}{\partial \beta} = 0 \; ; \; \frac{\partial \ln L}{\partial \sigma^2} = 0 \qquad (2.4.8)$$

maka diperoleh :

$$0 = -\frac{1}{\sigma} \sum_{y_i=0}^{n_0} \frac{\phi\left(\frac{x_i'\beta}{\sigma}\right) x_i'}{1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right)} + \frac{1}{\sigma^2} \sum_{y_i>0}^{n_1} (y_i - x_i'\beta)\, x_i'$$

$$0 = -\frac{1}{2\sigma^3} \sum_{y_i=0}^{n_0} \frac{(x_i'\beta)\phi\left(\frac{x_i'\beta}{\sigma}\right)}{1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right)} - \frac{n_1}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{y_i>0}^{n_1} (y_i - x_i'\beta)^2$$

Oleh Olsen (1978) didapatkan matriks Hessianya sebagai berikut :

$$= \begin{bmatrix} \sum_{y_i=0}^{n_0} \frac{\phi_i}{1-\Phi_i}\left(x_i' a - \frac{\phi_i}{1-\Phi_i}\right) x_i' x_i & 0 \\ 0 & -\frac{n_1}{h_1} \end{bmatrix}$$

$$- \begin{bmatrix} \sum_{y_i>0}^{n_1} x_i' x_i & -\sum_{y_i>0}^{n_1} x_i' y_i \\ -\sum_{y_i>0}^{n_1} x_i y_i & \sum_{y_i>0}^{n_1} y_i^2 \end{bmatrix}$$

Dengan persamaan diatas akan diperoleh nilai penaksir dari β dan σ^2 sebagai berikut .

$$\sigma^2 = \frac{\sum_{y_i>0}^{n_1} (y_i - x_i'\beta) y_i}{n_1}$$

$$\beta = \beta_{LS} - \sigma (X_1' X)^{-1} X_0' \gamma_0 \qquad (2.4.9)$$

Persamaan secara eksplisit memperlihatkan hubungan antara Penaksir Maksimum Likelihood untuk Model Tobit dan Penaksir Least Squares yang diperoleh dari pengamatan tak nol y.

2.5 Estimasi Pengujian Parameter Regresi Tobit

Parameter model regresi tobit dapat diestimasi dengan metode *Maximum Likelihood Estimation* (MLE). Fungsi likelihood untuk model regresi tobit dibedakan menjadi dua yaitu untuk $y_i^* > \tau$ dan $y_i^* \leq \tau$. Jika $\tau = 0$, fungsi likelihood model regresi tobit dapat dituliskan sebagai persamaan gabungan dua fungsi (Wooldridge, 2002).

$$L(\beta) = \prod_{i=1}^{n} \left[1(y_i > 0)\frac{1}{\sigma}\emptyset\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right]\left[1(y_i = 0)\left\{1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right)\right\}\right]$$

(2.5.1)

$$L(\beta) = \log L(\beta) = \sum_{i=1}^{n} 1(y_i > 0)\log\left\{\frac{1}{\sigma}\emptyset\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right\} + 1(y_i = 0)\log\left\{1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right)\right\}$$

(2.5.2)

Estimator parameter model regresi tobit diperoleh dari turunan β yang disamadengankan nol dengan mengasumsikan asumsikan σ^2 diketahui, akan tetapi karena penyelesaian persamaan ini bersifat nonlinier, maka sulit dilakukan perhitungan secara analitis. Oleh sebab itu, digunakan metode iteratif Newton-Raphson (Greene, 2001).

2.5.1 Uji Linearitas

Menurut Osborne dan Waters (2002), parameter model regresi linier berganda hanya bisa dengan tepat terestimasi apabila hubungan antara prediktor dan respon bersifat linier. Jika hubungan keduanya nonlinier, maka akan menimbulkan underestimate. Salah satu cara untuk menguji linieritas hubungan prediktor dan respon adalah dengan menguji komponen nonlinier yang dimasukkan ke dalam model, seperti *uji reset*.

Tahap pertama dalam uji reset adalah mendapatkan nilai dugaan $\hat{y}_i$, selanjutnya, nilai duga tersebut dimasukkan ke model regresi yang mengandung komponen nonlinier (Shukur dan Mantalos, 2004).

$$y_i = x_i'\beta + \hat{y}_i^2\gamma_1 + \hat{y}_i^3\gamma_2 + \ldots + \hat{y}_i^{H+1}\gamma_H + \delta_i$$

Hipotesis yang diuji adalah sebagai berikut .

H_0 : $\gamma_1 = \gamma_2 = \ldots = \gamma_H = 0$

H_1 : minimal terdapat satu $\gamma_h \neq 0$ (h=1,2,..,H)

Statistik uji reset ditunjukkan oleh persamaan berikut

$$RESET = \frac{(\varepsilon'\varepsilon - \delta'\delta)/p}{(\delta'\delta)/(n-H)} \sim F_{(p,n-H-1)}$$

Dimana :

$\varepsilon'\varepsilon$ = jumlah kuadrat residual model regresi primer

$\delta'\delta$ = jumlah kuadrat residual model regresi auxiliary

H_0 ditolak jika $P(F_{(p,n-H-1)} > RESET) < \alpha$

Di sisi lain, apabila terdapat variabel prediktor bersifat kategorik (dummy), maka uji yang digunakan adalah uji korelasi Point Biserial. Uji ini bertujuan untuk mengetahui adanya hubungan linier antar dua variabel di mana salah satu variabel bersifat biner/dikotomus dan variabel lainnya kontinu. Hipotesis yang diuji sebagai berikut.

H_0 : hubungan antar variabel tidak linier

H_1 : hubungan antar variabel linier

Menurut Glass dan Hopkins (1995), perhitungan korelasi Point biserial ditunjukkan oleh persamaan sebagai berikut.

$$r_{pb} = \frac{\bar{y}_1 - \bar{y}_0}{s_y} \sqrt{\frac{c_1 c_0}{c(c-1)}}$$

$$r_{pb} = \frac{r_{pb}\sqrt{c-2}}{\sqrt{1-{r_{pb}}^2}} \sim t_{(c-2)}$$

dimana :

$\bar{y}_1$ = rata-rata respon kategori 1

$\bar{y}_0$ = rata-rata respon kategori 0

c_1 = banyaknya observasi kategori 1

c_0 = banyaknya observasi kategori 0

c = $c_1 + c_0$

s_y = standar deviasi variabel respon y

H_0 ditolak jika $P(t_{(c-2)} > |t|) < \alpha$

2.5.2 Uji Beda

Menurut *Robinson et al*. (2013), uji beda dapat dilakukan untuk mengetahui pengaruh suatu variabel yang berpotensi menjadi moderator dalam suatu analisis regresi. Apabila variabel yang kemungkinan adalah moderator bersifat biner, dan pengaruhnya diuji terhadap variabel lain yang bersifat kontinu, maka uji beda yang digunakan adalah uji t. Hipotesis yang diuji dengan uji t sebagai berikut.

$$H_0 : \mu_1 - \mu_0 = 0$$

$$H_1 : \mu_1 - \mu_0 \neq 0$$

Bentuk umum statistik uji t

$$t = \frac{\bar{x}_1 - \bar{x}_0}{s_{\bar{x}_1 - \bar{x}_0}} \sim t_{(db)}$$

Jika ragam kelompok sama, maka $s_{\bar{x}_1 - \bar{x}_0} = \sqrt{\frac{(n_0 - 1)s_{x_0}^2 + (n_1 - 1)s_{x_1}^2}{n_0 + n_1 - 2}}$

dengan $db = n_0 + n_1 - 2$. Di sisi lain, jika ragam kelompok berbeda, digunakan $s_{\bar{x}_1 - \bar{x}_0} = \sqrt{\frac{s_{x_0}^2}{n_0} + \frac{s_{x_1}^2}{n_1}}$ dan db efektif.

dimana :

$\bar{x}_1$	=	rata-rata sampel kelompok kategori 1
$\bar{x}_0$	=	rata-rata sampel kelompok kategori 0
$s_{\bar{x}_1 - \bar{x}_0}$	=	$\sqrt{S_P^2 \left(\frac{1}{n_0} + \frac{1}{n_1}\right)}$, jika ragam sama
		$\sqrt{\frac{s_{x_0}^2}{n_0} + \frac{s_{x_1}^2}{n_1}}$, jika ragam berbeda

$S_P^2 \quad = \quad \sqrt{\frac{(n_0-1)s_{x_0}^2+(n_1-1)s_{x_1}^2}{n_0+n_1-2}}$

$db \quad = \quad n_0 + n_1 - 2$, jika ragam sama

$$\frac{\left(\frac{s_{x_0}^2}{n_0}+\frac{s_{x_1}^2}{n_1}\right)^2}{\left(\frac{s_{x_0}^2}{n_0}\right)^2 \Big/ (n_0-1)+\left(\frac{s_{x_1}^2}{n_1}\right)^2 \Big/ (n_1-1)}$$, jika ragam berbeda

n_0 = banyaknya observasi kelompok 0

n_1 = banyaknya observasi kelompok 1

H_0 ditolak jika $P\left(t_{(db)} > |t|\right) < \alpha/2$.

Apabila kedua variabel sama-sama bersifat biner, maka dipakai uji independensi chi-square. Adapun hipotesis untuk uji chisquare sebagai berikut.

H_0 : hubungan antar kedua variabel saling bebas

H_1 : hubungan antar kedua variabel tidak saling bebas

Statistik uji chi-square menurut Agresti, (2002)

$$\chi^2 = \sum_{m=1}^{g} \frac{(O_m-E_m)^2}{E_m} \sim \chi^2{}_{(g-1)}$$

dimana :

O_m = frekuensi observasi dalam kategori ke-m

E_m = frekuensi yang diharapkan dalam kategori ke-m

g = banyaknya kategori

H_0 ditolak jika $P\left(\chi^2{}_{(g-1)} > \chi^2\right) < \alpha$

2.6 Model Regresi Kuantil

Regresi kuantil adalah salah satu metode analisis regresi yang dapat menggambarkan hubungan satu atau beberapa variabel prediktor terhadap satu variabel respon pada berbagai titik kuantil (conditional quantile) dari distribusi variabel respon tersebut, sehingga metode ini dapat digunakan pada kondisi data yang heterogen. Hal ini berbeda dengan analisis regresi linier yang hanya dapat menggambarkan hubungan sebab-akibat pada mean (conditional mean) variabel respon (Koenker dan Hallock, 2001). Persamaan model regresi kuantil menunjukkan bentuk umum model regresi kuantil linier (Buhai, 2005).

$$y_i = x_i^t \beta(\theta) + \varepsilon(\theta)_i \quad 0 < \theta < 1$$

dimana :

y_i = nilai variabel respon ke-i

x_i^t = $(1, x_{1i}, x_{2i}, \dots, x_{pi})$

$\beta(\theta)$ = parameter model regresi pada kuantil ke-θ

$\varepsilon(\theta)_i$ = error model regresi kuantil ke-θ

$i = 1,2,\dots,n$

Berdasarkan Chen (2005), estimasi parameter model regresi kuantil diawali dengan menyatakan fungsi peluang kumulatif dari variabel random Y sehingga kuantil ke θ dari variabel ini dapat ditunjukkan sebagaimana persamaan.

$$F(y) = P(Y \leq y)$$

$$Q_Y(\theta) = \inf\{y; F(y) \geq \theta\}$$

Menurut Koenker dan Basset (1978), jika terdapat sebanyak n observasi $\{y_i : 1,\dots,n\}$ sebagai sampel random dari variabel Y dengan fungsi distribusi F, maka kuantil ke-θ sehingga dapat didefinisikan dengan menggunakan persamaan.

$$\min_{\beta \epsilon R^p} \left[\sum_{i \epsilon \{i; y_i \geq x_i\ '\beta\}} \rho_\theta(y_i - f(x_i))\right]$$

$\rho_\theta(u) = (\theta - 1_{\{u<0\}})u$ merupakan check function

Koenker dan Machado (1999) menyatakan bahwa persamaan tidak memiliki bentuk turunan yang tetap, sehingga metode iterasi numerik biasa tidak dapat digunakan untuk menyelesaikan persamaan tersebut. Oleh karena itu, untuk mendapatkan $\hat{\beta}(\theta)$ digunakan metode pemrograman linear yaitu metode simpleks. Formulasi pemrograman linier model regresi kuantil berdasarkan Yao dan Lee (2010).

2.7 Model Regresi Kuantil Bayesian

Menurut Yu dan Moyeed (2001) regresi kuantil Bayesian prinsip dasar pemodelan dengan pendekatan Bayesian adalah mendapatkan distribusi posterior dari suatu parameter bila diketahui distribusi prior dan fungsi likelihood yang sesuai dengan kaidah Bayes. Apabila pendekatan ini digunakan untuk model regresi kuantil, maka distribusi posterior $\beta(\theta)$ yakni $\pi(\beta(\theta)|y)$ dapat dirumuskan sebagai berikut.

$$\pi(\beta(\theta)|y) \propto L\big(y\big|\beta(\theta)\big)\pi(\beta(\theta))$$

$\pi(\beta(\theta))$ merupakan distribusi prior dari $\beta(\theta)$ dan $L\big(y\big|\beta(\theta)\big)$ adalah fungsi likelihood data dengan asumsi residual berdistribusi *asimetris Laplace.*

$$L\big(y\big|\beta(\theta)\big) = \theta^n(1-\theta)^n \, exp\left[-\sum_{i=1}^{n} \rho_0(y_i - x_i^t\beta(\theta))\right]$$

Distribusi prior yang digunakan oleh Yu dan Moyeed (2001) adalah prior improper uniform. Prior ini dipilih karena model regresi kuantil bayesian tidak memiliki prior konjugasi. Selain itu melalui pembuktian secara matematis, diketahui bahwa meskipun prior ini improper namun dapat menghasilkan posterior yang bersifat proper.

2.7.1 Metode MCMC (Markov Chain Monte Carlo): Algoritma Metropolis Hastings

MCMC adalah metode umum yang digunakan untuk mendapatkan suatu nilai (sampel) β dari distribusi peluang yang diketahui $\pi(\beta(\theta))$, lalu nilai yang didapatkan tersebut dikoreksi sehingga bisa lebih mendekati distribusi posterior yang dikehendaki $\pi(\beta|y)$.

Sampel-sampel β didapatkan secara berurutan (sequentially), atau dengan kata lain distribusi peluang yang digunakan untuk menghasilkan sampel saat ini tergantung dari sampel yang dihasilkan sebelumnya, sehingga membentuk suatu rantai Markov (Gelman et al., 2004).

2.7.2 Bayes Faktor

Bayes faktor pertama kali diperkenalkan oleh Harold Jeffreys pada tahun 1960 sebagai alternatif dari pengujian hipotesis frekuentis. Statistik ini sering digunakan untuk membandingkan beberapa model guna mendapatkan model terbaik (Lavine dan Schervish, 1999).Perbandingan model regresi tobit menggunakan Bayes faktor menurut Kass dan Raftery (1995) adalah sebagai berikut.

$$B = \frac{m(y|M_1)}{m(y|M_2)} = \frac{\int L(y|M_1,\beta_1)\pi(\beta_1|M_1)d\beta_1}{\int L(y|M_2,\beta_2)\pi(\beta_2|M_2)d\beta_2}$$

dimana :

M_1 = model ke-l ($l = 1,2$)

y = $(y_1, y_2, \dots, y_n)$

β_1 = vector parameter model ke-l

$m(y|M_1)$ = fungsi likelihood marginal model ke-l

$L(y|M_1, \beta_1)$ = fungsi likelihood model ke-l

$\pi(\beta_1|M_1)$ = fungsi likelihood prior model ke-l

Jika digunakan hipotesis $M_1 > M_2$, maka interpretasi Bayes faktor dapat dijelaskan seperti pada Tabel 2.1 (Kass dan Raftery, 1995).

Tabel 2.1 Kriteria Interpretasi Bayes Faktor

2log B	**B**	**Kekuatan Pembuktian**	**Model**
< 0	<1	Negatif	M_1
0 s/d 2	1 s/d 3	Tidak ada	
2 s/d 5	3 s/d 12	Positif	
5 s/d 10	12 s/d 150	Kuat	
> 10	> 150	Sangat Kuat	

Menurut Kass dan Raftery (1995) menyatakan bahwa perhitungan Bayes faktor cukup sulit untuk dilakukan. Namun, Chib dan Jeliakoz (2001) memberikan gagasan mereka untuk menghitung estimasi logaritma marginal likelihood, $\log m(y|M_1)$, untuk sembarang nilai β_l^* yang memiliki densitas tinggi.

$$\log m(y|M_1) = \log L(y|M_1, \beta_l^*) + \log \pi\left(\beta_{l\,1}^*\middle|M_1\right) - \log \pi\left(\beta_{l\,1}^*\middle|y, M_1\right)$$

Persamaan diatas dapat dihitung dengan menggunakan output dari algoritma Metropolis-Hastings.

2.8 Root Mean Square Error (RMSE)

RMSE (Root Mean Square Error) merupakan salah satu statistik yang sering digunakan untuk mengevaluasi kebaikan performa model atau estimator. Statistik ini mengukur selisih antara nilai yang diprediksi oleh suatu model/estimator dengan nilai sebenarnya, yang disebut juga sebagai error atau residual. Persamaan berikut menunjukkan formulasi RMSE jika diasumsikan terdapat sebanyak n error model (Chai dan Draxler, 2014).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2}$$