BAB III

METODOLOGI PENELITIAN

3.1 Jenis Sumber Data

Jenis data yang digunakan oleh peneliti merupakan data sekunder. Data yang digunakan sebanyak 150 objek. Sumber data pada penelitian ini yaitu data Acute Myeloid Leukimia di North West England pada tahun 1982-1998.

3.2 Variabel Penelitian

Data yang digunakan dalam analisis model regresi Cox Hazard dengan unsur spasial geostatistik akan mencakup observasi (individu) yang memiliki informasi tentang waktu tahan hidup, *censoring* status, koordinat spasial tempat tinggal, usia individu saat menderita sakit(X_1), jenis kelamin (X_2), jumlah sel darah putih individu saat diagnosa (X_3), dan tingkat ketidakmakmuran daerah (X_4). Dengan data ini, analisis akan memodelkan bagaimana prediktor ini memengaruhi risiko kematian sambil memperhitungkan struktur spasial dan hubungan spasial antar lokasi observasi. Model ini akan membantu mengidentifikasi faktor-faktor yang signifikan dalam prediksi waktu tahan hidup dan memahami dampak spasial pada risiko kematian. Pengolahan data akan menggunakan bantuan software R.

1. Waktu Tahan Hidup (*Survival Time*):
   Ini adalah variabel dependen yang mewakili waktu yang diperlukan sejak awal pengamatan hingga kejadian kematian atau waktu terakhir yang diamati (*censoring*). Waktu tahan hidup diukur dalam satuan hari.
2. *Censoring* (*Cens*):
   Variabel ini digunakan untuk mengindikasikan apakah observasi telah berakhir atau tercensor (1, jika mengalami *event* atau meninggal, dan 0, jika masih hidup)
3. Koordinat Tempat Tinggal (Xcoord, Ycoord):
   Ini adalah variabel spasial yang menggambarkan koordinat geografis atau ruang-temporal di mana setiap observasi (individu) berada. Variabel ini memungkinkan analisis spasial geostatistik untuk memperhitungkan struktur

spasial dan hubungan spasial antar observasi. Dengan rentang tertentu mewakili syarat wilayah yang saling berdekatan. Data ini merupakan lokasi titik tempat tinggal individu.

4. Usia individu saat senderita sakit leukimia (X_1):

   Variabel ini mengukur usia individu saat mereka pertama kali menderita penyakit. Rentang diukur dengan tahun.

5. Jenis Kelamin (X_2):

   Ini adalah variabel yang mengkode jenis kelamin individu, 1 untuk laki-laki dan 0 untuk Perempuan.

6. Jumlah Sel Darah Putih Individu Saat Diagnosa (X_3):

   Ini adalah prediktor yang mengukur jumlah sel darah putih individu pada saat diagnosis penyakit. 1 unit = 50.10^9/L

7. Tingkat Ketidakmakmuran Daerah (*tpi)* (X_4):

   Variabel ini mengukur tingkat ketidakmakmuran atau indeks sosio-ekonomi daerah tempat tinggal individu. Faktor-faktor sosial dan ekonomi daerah dapat berkontribusi pada risiko penyakit atau kematian. Nilai *tpi* yang semakin tinggi menunjukkan daerah yang kurang Makmur.

Tabel 3. 1 Variabel kovariat Data sampel

Variabel	Keterangan	Jenis	Kategori
Y	Waktu ketahanan hidup (dalam hari)	Kontinu	-
X_1	Usia	Kontinu	-
X_2	Jenis Kelamin	Kategorikal	1 = laki-laki; 0 = Perempuan
X_3	Jumlah sel darah putih	Kontinu	-
X_4	Tingkat Ketidakmakmuran Daerah	Kontinu	-

3.3 Metode Analisis Data

Metode analisis data dalam penelitian ini adalah analisis deskriptif yang digunakan untuk melihat gambaran dari masing-masing variabel serta analisis regresi *cox proportional hazard* dengan metode MLE (*Maximum Likelihood Estimation*) dan metode *Newton Rhapson* berdistribusi Rayleigh yang digunakan untuk mengetahui faktor-faktor apa saja yang diduga memengaruhi waktu tahan hidup penderita AML.

3.4 Model Survival Spasial

Dalam analisis survival, T merupakan variabel acak waktu kegagalan. Setiap sampel individu diasumsikan berkaitan dengan waktu kegagalan. Sebagian besar data survival merupakan observasi pada individu yang terdiri dari kovariat atau efek acak yang saling berkaitan dengan individu.

Fungsi kumulatif hazard dari model geostatistik berdistribusi Rayleigh adalah sebagai berikut.

$$H(t_{ij}; x_{ij}) = \int_o^{t_{ij}} h(u; x_{ij}) du$$

$$= \int_o^{t_{ij}} h_o(u_{ij}) \exp(\beta^T x_{ij} + W_j)\, du$$

$$= \int_o^{t_{ij}} t_{ij} \exp(\beta^T x_{ij} + W_j)\, du$$

$$= \frac{{t_{ij}}^2}{2} \exp(\beta^T x_{ij} + W_j)$$

Maka, hubungan antara fungsi survival spasial dengan hazard kumulatifnya adalah:

$$S(t_{ij}; x_{ij}) = \exp[-H(t_{ij}; x_{ij})]$$

$$= exp\left[-\frac{{t_{ij}}^2}{2} \exp(\beta^T x_{ij} + W_j)\right]$$

Untuk fungsi densitas, dapat diperoleh:

$$f(t_{ij}; x_{ij}) = h(t_{ij}; x_{ij}) S(t_{ij}; x_{ij})$$

$$= t_{ij}\exp\left(\beta^T x_{ij} + W_j\right)\exp\left[-\frac{{t_{ij}}^2}{2}\exp\left(\beta^T x_{ij} + W_j\right)\right]$$

$$= t_{ij}\{\exp\left(\beta^T x_{ij} + W_j\right) - \frac{{t_{ij}}^2}{2}\mathrm{e}^{\left(\beta^T x_{ij} + W_j\right)}\}$$

3.5 Estimasi Parameter dengan Optimisasi Menggunakan Metode Maximum Likelihood

3.3.1 Menentukan Fungsi Likelihood dari Model Regresi Cox

Berdasarkan model survival spasial, maka diperoleh:

$$L(\beta, W, t) = \prod_u f(t_u)\prod_c S(t_c)$$

$$L(\beta, W, t) = \prod_{i=1}^{l}\prod_{j=1}^{n}\left(f\left(t_{ij}; x_{ij}\right)\right)^{\gamma_{ij}}\left(S\left(t_{ij}; x_{ij}\right)\right)^{1-\gamma_{ij}}$$

$$L(\beta, W, t) = \prod_{i=1}^{l}\prod_{j=1}^{n}\left\{t_{ij}\{\exp\left(\beta^T x_{ij} + W_j\right) - \frac{{t_{ij}}^2}{2}\mathrm{e}^{\left(\beta^T x_{ij} + W_j\right)}\}\right\}^{\gamma_{ij}}\left\{exp\left[-\frac{{t_{ij}}^2}{2}\exp\left(\beta^T x_{ij} + W_j\right)\right]\right\}^{1-\gamma_{ij}}$$

$$L(\beta, W, t) = \prod_{i=1}^{l}\prod_{j=1}^{n}\frac{\left\{\frac{t_{ij}\{\mathrm{e}^{\beta^T x_{ij} + W_j}}{e^{-\frac{{t_{ij}}^2}{2}\mathrm{e}^{\left(\beta^T x_{ij} + W_j\right)}}}\right\}^{\gamma_{ij}}\left\{e^{-\frac{{t_{ij}}^2}{2}\exp\left(\beta^T x_{ij} + W_j\right)}\right\}^{1}}{\left\{e^{-\frac{{t_{ij}}^2}{2}\exp\left(\beta^T x_{ij} + W_j\right)}\right\}^{\gamma_{ij}}}$$

$$L(\beta, W, t) = \prod_{i=1}^{l}\prod_{j=1}^{n}(t_{ij}\{\mathrm{e}^{\beta^T x_{ij} + W_j})^{\gamma_{ij}}\; e^{-\frac{{t_{ij}}^2}{2}\exp\left(\beta^T x_{ij} + W_j\right)}$$

3.3.2 Menentukan Fungsi log-Likelihood dari Model Regresi Cox

$$l\,(\beta, W, t) = \ln L(\beta, W, t)$$

$$l\,(\beta, W, t) = \ln(\prod_{i=1}^{l}\prod_{j=1}^{n}(t_{ij}\{\mathrm{e}^{\beta^T x_{ij} + W_j})^{\gamma_{ij}}\; e^{-\frac{{t_{ij}}^2}{2}\exp\left(\beta^T x_{ij} + W_j\right)}$$

$$l\,(\beta, W, t) = \sum_{i=1}^{l}\sum_{j=1}^{n}[\,\gamma_{ij}\ln(t_{ij}\left\{\mathrm{e}^{\beta^T x_{ij} + W_j}\right) + \ln\left(e^{-\frac{{t_{ij}}^2}{2}exp\left(\beta^T x_{ij} + W_j\right)}\right)$$

$$= \sum_{i=1}^{l}\sum_{j=1}^{n}[\,\gamma_{ij}\left\{\ln(t_{ij}) + \ln\left(e^{\beta^T x_{ij} + W_j}\right)\right\}(\ + \ln\left(e^{-\frac{t_{ij}^2}{2}exp(\beta^T x_{ij} + W_j)}\right)]$$

$$= \sum_{i=1}^{l}\sum_{j=1}^{n}[\,\gamma_{ij}\{\ln(t_{ij}) + \beta^T x_{ij}\ +\ W_j\} - \frac{t_{ij}^2}{2}exp(\beta^T x_{ij}\ +\ W_j)]$$

$$= \sum_{i=1}^{l}\sum_{j=1}^{n}[\,\gamma_{ij}\ln t_{ij} +\ \gamma_{ij}(\beta^T x_{ij}\ +\ W_j) - \frac{t_{ij}^2}{2}exp(\beta^T x_{ij}\ +\ W_j)]$$

3.3.3 Menentukan Turunan Pertama Fungsi log-Likelihood terhadap Parameter β_k

Untuk β_1:

$$\frac{\partial l}{\partial \beta_1} = 0$$

$$\frac{\partial(\sum_{i=1}^{l}\sum_{j=1}^{n}[\gamma_{ij}\, lnt_{ij} + \gamma_{ij}(\beta^T x_{ij} + W_j) - \frac{t_{ij}^2}{2}exp(\beta^T x_{ij} + W_j)}{\partial \beta_1} = 0$$

$$\frac{\partial(\sum_{i=1}^{l}\sum_{j=1}^{n}[\gamma_{ij}\, lnt_{ij} + \gamma_{ij}(\beta_1 x_{1ij} + \dots + \beta_4 x_{4ij} + W_j) - \frac{t_{ij}^2}{2}e^{\beta_1 x_{1i}\ + \dots + \beta_4 x_{4ij} + W_j}}{\partial \beta_1} = 0$$

$$\sum_{i=1}^{l}\sum_{j=1}^{n}[0 + 0 + \gamma_{ij}(x_{1ij} + 0 + 0 + 0) - \frac{t_{ij}^2}{2}x_{1ij}e^{\beta_1 x_{1ij} + \dots + \beta_4 x_{4ij} + W_j} = 0$$

$$\sum_{i=1}^{l}\sum_{j=1}^{n}\left[x_{1ij}\left(\gamma_{ij} - \frac{t_{ij}^2}{2}e^{\beta_1 x_{1ij} + \dots + \beta_4 x_{4i}\ \ + W_j}\right)\right] = 0$$

Untuk β_2:

$$\frac{\partial l}{\partial \beta_2} = 0$$

$$\frac{\partial(\sum_{i=1}^{l}\sum_{j=1}^{n}[\gamma_{ij}\, ln\ \ _{ij} + \gamma_{ij}(\beta^T x_{ij} + W_j) - \frac{t_{ij}^2}{2}exp(\beta^T x_{ij} + W_j)}{\partial \beta_2} = 0$$

$$\frac{\partial(\sum_{i=1}^{l}\sum_{j=1}^{n}[\gamma_{ij}\, lnt_{ij} + \gamma_{ij}(\beta_1 x_{1ij} + \dots + \beta_4 x_{4ij} + W_j) - \frac{t_{ij}^2}{2}e^{\beta_1 x_{1ij} + \dots + \beta_4 x_{4ij} + W_j}}{\partial \beta_2} = 0$$

$$\sum_{i=1}^{l}\sum_{j=1}^{n}[0 + 0 + \gamma_{ij}(0 + x_{2ij} + 0 + 0) - \frac{t_{ij}^2}{2}x_{2i}\ e^{\beta_1 x_{1ij} + \dots + \beta_4 x_{4i}\ \ + W_j} = 0$$

$$\sum_{i=1}^{l}\sum_{j=1}^{n}\left[x_{2i}\ \left(\gamma_{ij} - \frac{t_{ij}^2}{2}e^{\beta_1 x_{1ij} + \dots + \beta_4 x_{4ij} + W_j}\right)\right] = 0$$

Untuk β_3:

$$\frac{\partial l}{\partial \beta_3} = 0$$

$$\frac{\partial (\sum_{i=1}^{l} \sum_{j=1}^{n} [\gamma_{ij}\, ln \;\; _{ij} + \gamma_{ij}(\beta^T x_{ij} + W_j) - \frac{t_{ij}{}^2}{2} exp(\beta^T x_{ij} + W_j)}{\partial \beta_3} = 0$$

$$\frac{\partial (\sum_{i=1}^{l} \sum_{j=1}^{n} [\gamma_{ij}\, lnt_{ij} + \gamma_{ij}(\beta_1 x_{1ij} + \dots + \beta_4 x_{4i} \;\; + W_j) - \frac{t_{ij}{}^2}{2} e^{\beta_1 x_{1i} \; + \dots + \beta_4 x_{4ij} + W_j}}{\partial \beta_3} = 0$$

$$\sum_{i=1}^{l} \sum_{j=1}^{n} [0 + 0 + \gamma_{ij}\big(0 + 0 + x_{3ij} + 0\big) - \frac{t_{ij}{}^2}{2} x_{3i} \;\; e^{\beta_1 x_{1i} \; + \dots + \beta_4 x_{4ij} + W_j} = 0$$

$$\sum_{i=1}^{l} \sum_{j=1}^{n} \left[x_{3ij} \left(\gamma_{ij} - \frac{t_{ij}{}^2}{2} e^{\beta_1 x_{1ij} + \dots + \beta_4 x_{4ij} + W_j} \right) \right] = 0$$

Untuk β_4:

$$\frac{\partial l}{\partial \beta_4} = 0$$

$$\frac{\partial (\sum_{i=1}^{l} \sum_{j=1}^{n} [\gamma_{ij}\, ln\, t_{ij} + \gamma_{ij}(\beta^T x_{ij} + W_j) - \frac{t_{ij}{}^2}{2} exp(\beta^T x_{ij} + W_j)}{\partial \beta_4} = 0$$

$$\frac{\partial (\sum_{i=1}^{l} \sum_{j=1}^{n} [\gamma_{ij} ln\, t_{ij} + \gamma_{ij}(\beta_1 x_{1ij} + \dots + \beta_4 x_{4ij} + W_j) - \frac{t_{ij}{}^2}{2} e^{\beta_1 x_{1i} \; + \dots + \beta_4 x_{4ij} + W_j}}{\partial \beta_4} = 0$$

$$\sum_{i=1}^{l} \sum_{j=1}^{n} [0 + 0 + \gamma_{ij}\big(0 + 0 + 0 + x_{4ij}\big) - \frac{t_{ij}{}^2}{2} x_{4i} \;\; e^{\beta_1 x_{1ij} + \dots + \beta_4 x_{4ij} + W_j} = 0$$

$$\sum_{i=1}^{l} \sum_{j=1}^{n} \left[x_{4ij} \left(\gamma_{ij} - \frac{t_{ij}{}^2}{2} e^{\beta_1 x_{1ij} + \dots + \beta_4 x_{4i} \;\; + W_j} \right) \right] = 0$$

3.3.4 Menentukan Turunan kedua Fungsi log-Likelihood terhadap Parameter β_k

Untuk β_1:

$$\frac{\partial^2 l}{\partial {\beta_1}^2} = \frac{\partial^2 (\sum_{i=1}^{l} \sum_{j=1}^{n} \left[x_{1ij} \left(\gamma_{ij} - \frac{t_{ij}{}^2}{2} e^{\beta_1 x_{1ij} + \dots + \beta_4 x_{4ij} + W_j} \right) \right]}{\partial {\beta_1}^2}$$

$$\frac{\partial^2 l}{\partial {\beta_1}^2} = \sum_{i=1}^{l} \sum_{j=1}^{n} \left\{ (x_{1ij})^2 \frac{t_{ij}{}^2}{2} e^{\beta_1 x_{1ij} + \dots + \beta_4 x_{4ij} + W_j} \right\}$$

Untuk β_2:

$$\frac{\partial^2 l}{\partial \beta_2{}^2} = \frac{\partial^{2(\sum_{i=1}^{l}\sum_{j=1}^{n}\left[x_{2ij}\left(\gamma_{ij}-\frac{t_{ij}{}^2}{2}e^{\beta_1 x_{1ij}+\,\dots\,+\beta_4 x_{4ij}+W_j}\right)\right]}}{\partial \beta_2{}^2}$$

$$\frac{\partial^2 l}{\partial \beta_2{}^2} = \sum_{i=1}^{l}\sum_{j=1}^{n}\left\{(x_{2i}\ )^2\frac{t_{ij}{}^2}{2}e^{\beta_1 x_{1ij}+\,\dots\,+\beta_4 x_{4ij}+W_j}\right\}$$

Untuk β_3:

$$\frac{\partial^2 l}{\partial \beta_3{}^2} = \frac{\partial^{2(\sum_{i=1}^{l}\sum_{j=1}^{n}\left[x_{3i}\left(\gamma_{ij}-\frac{t_{ij}{}^2}{2}e^{\beta_1 x_{1ij}+\,\dots\,+\beta_4 x_{4ij}+W_j}\right)\right]}}{\partial \beta_2{}^2}$$

$$\frac{\partial^2 l}{\partial \beta_3{}^2} = \sum_{i=1}^{l}\sum_{j=1}^{n}\left\{(x_{3ij})^2\frac{t_{ij}{}^2}{2}e^{\beta_1 x_{1ij}+\,\dots\,+\beta_4 x_{4ij}+W_j}\right\}$$

Untuk β_4:

$$\frac{\partial^2 l}{\partial \beta_4{}^2} = \frac{\partial^{2(\sum_{i=1}^{l}\sum_{j=1}^{n}\left[x_{4ij}\left(\gamma_{ij}-\frac{t_{ij}{}^2}{2}e^{\beta_1 x_{1ij}+\,\dots\,+\beta_4 x_{4ij}+W_j}\right)\right]}}{\partial \beta_2{}^2}$$

$$\frac{\partial^2 l}{\partial \beta_4{}^2} = \sum_{i=1}^{l}\sum_{j=1}^{n}\left\{(x_{4ij})^2\frac{t_{ij}{}^2}{2}e^{\beta_1 x_{1ij}+\,\dots\,+\beta_4 x_{4ij}+W_j}\right\}$$

3.6 Tahapan Penelitian

Dalam penelitian ini, tahapan penelitian dilakukan sebagai berikut.

1. Eksplorasi data

   Tahapan awal penelitian melibatkan data sekunder. Data sekunder digunakan sebagai dasar untuk melakukan analisis statistik dan pemodelan.

2. Analisis deskriptif terhadap variabel respon serta variabel prediktor.

   Analisis deskriptif dilakukan terhadap variabel respon dan variabel prediktor. Ini mencakup menghitung statistik deskriptif seperti mean, median, simpangan baku, dan visualisasi data melalui grafik.

3. Uji distribusi tertentu dari waktu survival menggunakan *Anderson-Darling*.

   Langkah selanjutnya adalah menguji distribusi waktu survival data. Uji *Anderson-Darling* adalah salah satu uji statistik yang digunakan untuk menguji apakah data waktu survival mengikuti distribusi Rayleigh.

4. Memasukkan efek acak spasial (frailty).

   Untuk memodelkan variabilitas yang tidak dapat dijelaskan oleh variabel prediktor yang diamati, efek acak spasial atau efek frailty dapat dimasukkan ke model survival. Efek ini memungkinkan penanganan ketidakpastian atau variabilitas yang terkait dengan lokasi geografis atau elemen-elemen tidak terukur lainnya.

5. Estimasi parameter variabel prediktor.

   Tahapan ini melibatkan estimasi parameter dari variabel prediktor. Dalam konteks regresi survival, ini sering melibatkan penggunaan metode statistik yaitu metode *Maksimum Likelihood Estimation* (MLE) untuk mengestimasi parameter efek variabel prediktor terhadap waktu survival.

6. Uji Signifikansi parameter.

   Setelah estimasi parameter dilakukan, uji signifikansi dapat digunakan untuk mengevaluasi apakah efek variabel prediktor tersebut signifikan dalam mempengaruhi waktu survival. Uji hipotesis, yaitu uji Wald digunakan untuk tujuan ini.

7. Pemodelan regresi *Cox Proportional Hazard* dan survival spasial.

   Pada tahap ini, model regresi *Cox Proportional Hazard* dan model survival spasial dapat dikembangkan. Dari hasil uji signifikansi parameter, variabel prediktor yang memiliki hasil yang signifikan dimasukkan ke model. Model ini memungkinkan analisis hubungan antara variabel prediktor, efek frailty, dan waktu survival. Model ini dapat digunakan untuk memprediksi waktu survival individu berdasarkan variabel prediktor.

8. Menghitung peluang hidup dan mati individu.

   Setelah pemodelan selesai, peluang hidup dan peluang kematian individu pada waktu tertentu dapat dihitung menggunakan model yang dikembangkan. Hal ini berguna untuk mendapatkan pemahaman yang lebih dalam tentang risiko dan peluang individu.

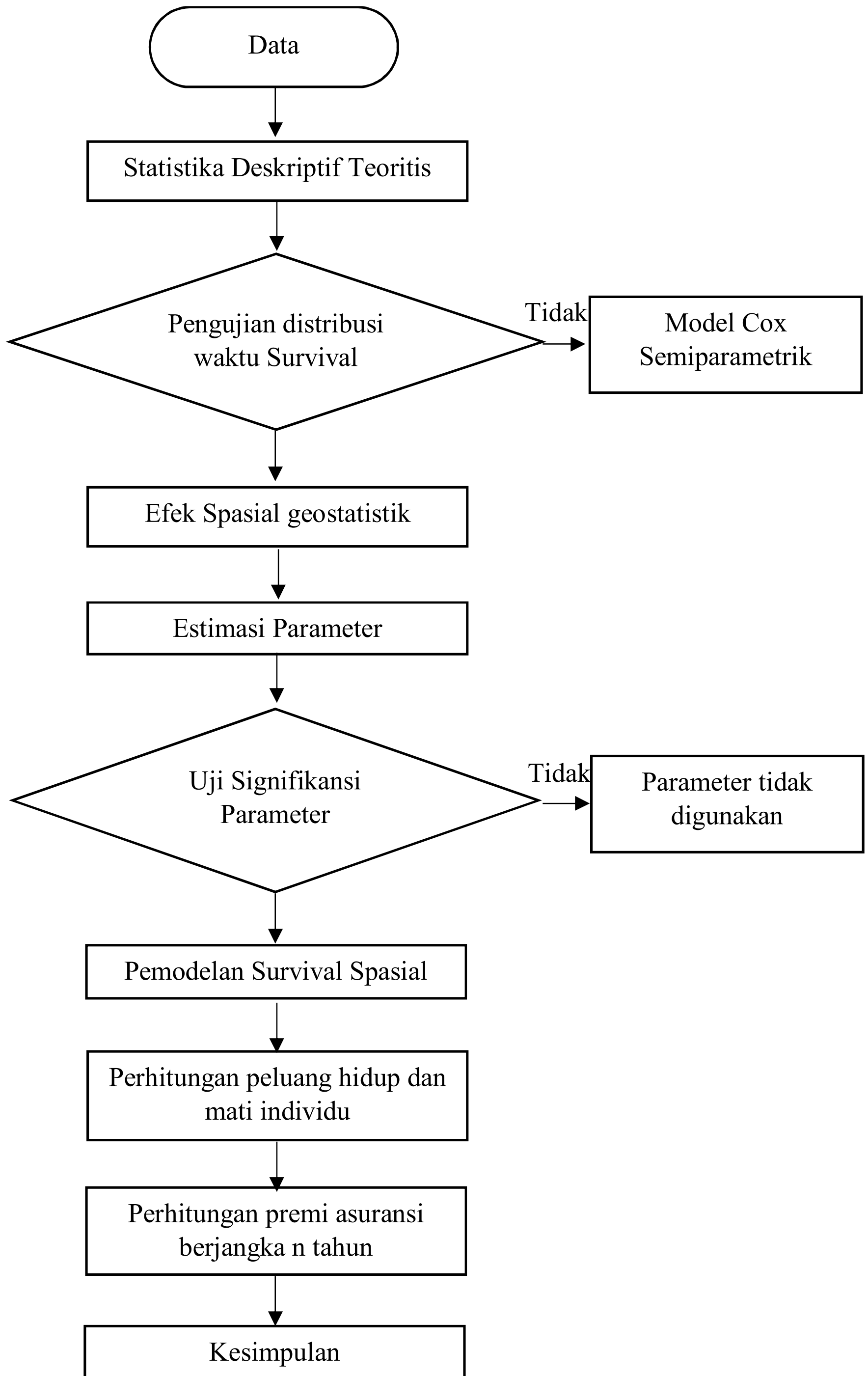
9. Menghitung nilai premi asuransi jiwa berjangka n tahun individu.

   Dalam penelitian ini, ada kebutuhan untuk menghitung nilai premi asuransi jiwa berjangka selama n tahun untuk individu berdasarkan model survival yang telah dikembangkan. Ini melibatkan penghitungan premi yang harus dibayarkan oleh individu untuk mendapatkan manfaat asuransi sesuai dengan jangka waktu yang ditentukan.

10. Kesimpulan.

Tahapan penelitian diakhiri dengan menyusun kesimpulan berdasarkan hasil analisis dan temuan dari penelitian ini. Kesimpulan dapat merangkum temuan penting, implikasi, dan saran untuk penelitian selanjutnya atau implementasi praktis.

Berikut diagram alir mengenai tahapan penelitiannya.



Gambar 3. 1 Diagram Alir Penelitian