PEMODELAN REGRESI *SPLINE* HIERARKI DENGAN METODE *GENERALIZED LEAST SQUARE*

(Studi Kasus : Angka Kematian Ibu Tahun 2020)

MODELLING HIERARCHICAL SPLINE REGRESSION USING THE GENERALIZED LEAST SQUARE METHOD

(*Case Study: Maternal Mortality Rate 2020*)

A.NURHAYATI LATIEF

H062201006

PROGRAM STUDI MAGISTER STATISTIKA

DEPARTEMEN STATISTIK

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

2022

PEMODELAN REGRESI *SPLINE* HIERARKI DENGAN METODE *GENERALIZED LEAST SQUARE*

(Studi Kasus : Angka Kematian Ibu Tahun 2020)

Tesis

Sebagai salah satu syarat untuk mencapai gelar magister

Program Studi Magister Statistika

Disusun dan diajukan oleh

A.NURHAYATI LATIEF

H062201006

Kepada

PROGRAM STUDI MAGISTER STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2022

TESIS

PEMODELAN REGRESI *SPLINE* HIERARKI DENGAN METODE *GENERALIZED LEAST SQUARE*

(Studi Kasus : Angka Kematian Ibu Tahun 2020)

A.NURHAYATI LATIEF

H062201006

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Program Studi Magister Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin
pada tanggal 01 Maret 2023
dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama,

Dr. Anna Islamiyati, S.Si., M.Si.
NIP. 19770808 200501 2 002

Pembimbing Pendamping

Dr. Nurtiti Sunusi, S.Si., M.Si.
NIP. 19720117 199703 2 002

Ketua Program Studi
Magister Statistika,

Dr. Dr. Georgina M. Tinungki, M.Si
NIP. 19620926 198702 2 001

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin,

Dr. Eng. Amiruddin, M.Si.
NIP. 19720515 199702 1 002

PERNYATAAN KEASLIAN TESIS
DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, tesis berjudul Pemodelan Regresi *Spline* Hierarki dengan Metode *Generalized Least Square* adalah benar karya saya dengan arahan dari komisi pembimbing (Dr. Anna Islamiyati, S.Si., M.Si dan Dr. Nurtiti Sunusi, S.Si., M.Si). Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka tesis ini. Sebagian dari isi tesis ini telah dipublikasikan di Jurnal (International Journal of Research Publications (IJRP.ORG) ISSN: 2708-3578 Vol. 113 Issue 1, November - 2022, Pages:323-331) sebagai artikel dengan judul "*The Implementation of Hierarchical Linear Spline Regression Model to Maternal Mortality Rate Data in Indonesia*".

Dengan ini saya melimpahkan hak cipta dari karya tulis saya berupa tesis ini kepada Universitas Hasanuddin.

Makassar, 01 Maret 2023
Yang Menyatakan,

A.Nurhayati Latief
NIM. H0622010006

UCAPAN TERIMA KASIH

Puji syukur atas kehadirat Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya sehingga penulis dapat menyusun dan menyelesaikan tesis ini. Penulis menyadari sepenuhnya bahwa apa yang dikemukakan dalam tesis ini masih jauh dari kesempurnaan yang merupakan sebagai akibat dari keterbatasan kemampuan serta berbagai kesulitan yang penulis hadapi dalam penyusun tesis ini.

Penulis memanjatkan doa kepada Tuhan Yang Maha Esa agar memberikan rahmat-Nya kepada pihak yang banyak membantu dalam penyelesaian tesis ini. Penulis juga percaya tesis ini dapat selesai bukan hanya dengan kekuatan pikiran penulis semata akan tetapi karena bantuan dari berbagai pihak juga, baik selama proses perkuliahan bahkan sampai proses pengerjaan tesis di Program Magister Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin. Namun demikian, penulis dengan senang hati menerima kritik dan saran yang bersifat membangun dari pembaca karya tulis ini demi sempurnanya tesis ini.

Terima kasih yang tak terhingga kepada kedua orang tua tercinta dan saudariku atas doa yang tak pernah putus, dukungan serta segala kebaikan mereka yang sampai kapan pun takkan pernah bisa terbalaskan atas kasih sayang yang tiada henti dalam penyelesaian tesis ini. Selanjutnya, saya ingin menyampaikan juga rasa hormat dan terima kasih kepada:

1. **Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.** selaku Rektor Universitas Hasanuddin.
2. **Dr. Eng. Amiruddin, M.Si**. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam beserta seluruh jajarannya.
3. **Dr. Nurtiti Sanusi, M.Si.** selaku Ketua Departemen Statistika yang menjadi salah satu tim pembimbing tesis sekaligus memberikan ilmu, dukungan, dan motivasi serta kemudahan kepada penulis dalam berbagai hal selama menjalani Pendidikan di Departemen Statistika.

4. **Dr. Dr. Georgina Maria Tinungki, M.Si**. selaku Ketua Program Studi Magister Statistika yang menjadi salah satu tim penguji tesis.
5. **Dr. Anna Islamiyati, S.Si., M.Si** selaku Pembimbing Utama yang telah bersabar dan bersedia meluangkan banyak waktunya untuk membimbing penulis dan memberikan masukan dalam penyelesaian tesis ini.
6. **Dr. Nirwan Ilyas, M.Si** selaku penguji penulis yang telah bersedia memberikan masukan-masukan dan arahan dalam penyusunan tesis.
7. **Sri Astuti Thamrin, S.Si.,** selaku penguji penulis yang telah bersedia memberikan masukan-masukan dan arahan dalam penyusunan tesis.
8. Keluarga **Latief Family** khususnya mama tercinta, **A.Nuraeni Tenriliweng,** kakak-kakak, dan adik-adik yang telah memberi doa dan dukungan baik berupa materiil maupun moril selama ini.
9. Suami tercinta, **Mu'tashim, S.Pd.,M.Pd.** yang telah menjadi rumah ternyaman untuk pulang, yang telah begitu sabar menemani dan mendukung penulis dengan segala pengertian dan perhatian kepada penulis selama ini.
10. Anakku tersayang **Muhammad Zaydan** yang selama ini menjadi sumber kekuatan, semangat hidup, dan sumber kebahagiaan bagi penulis.
11. Keluarga **Sudiang Squad,** ibu mertua, **Hasnawaty**, adik-adikku, **Musfirah**, **Mustaufiqurrahmah**, **Ridha Mustaufida**, **Faidatul Izza**, **Ismayanti**, yang selama ini telah bersabar dan selalu memberi dukungan serta membantu mengasuh Zaydan selama proses penyelesaian studi ini.
12. Kepala Kantor BPMP Provinsi Sulawesi Barat, **Sinar Alam, S.Pd., M.Pd.**, dan **Dr. Haksan Darwangsa, M.Si.**, Kepala BAN S/M Provinsi Sulawesi Barat, **Dr. H. Amran HB, S.Ag., M.Pd**. serta seluruh staf sekretariat BAN S/M Provinsi Sulawesi Barat khususnya **Sudarming, S.Pd.,M.Pd.,** dan **Asman Jufendi, S.Pd.** yang telah memberi kesempatan, bantuan, dan dukungan selama proses penyelesaian studi ini.

13. Teman – teman Mahasiswa Program Magister Statistika FMIPA Unhas Angkatan 2 yaitu, **Nurul Fadhilah, Alfi Nur Khauly, Mita Astuti, Ardiansyah Abubakar, Firman Saputra,** dan **Bakry**, serta seluruh angkatan lainnya terima kasih atas dukungan yang luar biasa kepada penulis.
14. **Teman-teman Lab** yaitu **Siswanto, M.Si.**, **Hedi Kuswanto, M.Si., Andi Isna Yunita, M.Si., Mubasiratul Munawaroh, Ratmila, Nur Aminah, Nurul Azizah, Maktisen Ena, Alimatun Najiha, Ainun Utari, Ummul Auliah Syam, Alya, Samsir Aditya Ania, Muh. Qardawi Hamzah, M.Si. Trigarcia Randa, M.Si.** atas doa, bantuan, semangat serta kebersamaannya selama ini yang banyak membantu penulis.

Semoga Allah SWT memberikan pahala yang berlipat ganda atas segala kebaikan yang telah diberikan kepada penulis dan semoga penulisan tesis ini bermanfaat bagi perkembangan ilmu pengetahuan dan teknologi, khususnya dalam dunia statistika dan data sains.

Makassar, 26 Februari 2023

A.Nurhayati Latief

ABSTRAK

A.NURHAYATI LATIEF. **Pemodelan Regresi *Spline* Hierarki dengan Metode *Generalized Least Square*** (Dibimbing oleh Dr. Anna Islamiyati, S.Si., M.Si dan Dr. Nurtiti Sunusi, S.Si., M.Si)

Analisis regresi nonparametrik merupakan salah satu metode regresi yang digunakan untuk mengetahui pola hubungan antara variabel respon dan variabel penjelas yang tidak diketahui kurvanya. Salah satu pendekatan regresi nonparametrik yang popular digunakan ialah model regresi *spline truncated* karena memiliki sifat fleksibel dan mudah di interpretasikan secara visual. Model regresi *spline* dikembangkan menjadi model *spline* linier hirarki yang telah banyak membantu dalam mengetahui hubungan antara variabel respon dan variabel penjelas untuk data yang memiliki pola data yang tersarang. Model spline linier hirarki juga dapat mengatasi beberapa isu seperti kolinearitas pada pengukuran berulang dan bias karena data yang hilang. Model spline linier hirarki merupakan penggabungan model regresi spline linier truncated ke dalam data yang memiliki struktur hirarki. Penelitian ini, menggunakan metode *Generalized Least Square* (GLS) sebagai estimatornya dan diaplikasikan pada data Angka Kematian Ibu tahun 2020. Hasil yang diperoleh adalah model terbaik penggabungan dari model level 1 menggunakan 1 (satu) titik knot dengan nilai GCV 32,781 dan model level 2 menggunakan 2 (dua) titik knot dengan nilai GCV 89,865 dengan variabel yang signifikan pada level-1 ialah jumlah ibu hamil yang melaksanakan K4 dan jumlah ibu hamil yang menangani komplikasi kebidanan dan variabel yang signifikan pada level-2 ialah persentase penduduk miskin dan angka partisipasi sekolah.

Kata kunci: Model Hirarki, Generalized Least Square, Spline truncated, Angka Kematian Ibu

ABSTRACT

A.NURHAYATI LATIEF. ***Modelling Hierarchical Spline Regression Using The Generalized Least Square Method*** (supervised by Dr. Anna Islamiyati, S.Si., M.Si and Dr. Nurtiti Sunusi, S.Si., M.Si)

Nonparametric regression analysis is one of the regression methods used to determine the relationship pattern between the response variable and the explanatory variable whose curve is unknown. One of the popular nonparametric regression approaches used is the spline truncated regression model because it is flexible and easy to interpret visually. The spline regression model was developed into a hierarchical linear spline model which has helped a lot in knowing the relationship between the response variable and the explanatory variable for data that has a nested data pattern. The hierarchical linear spline model can also overcome some issues such as collinearity in repeated measurements and bias due to missing data. The hierarchical linear spline model is an amalgamation of the truncated linear spline regression model into data that has a hierarchical structure. This study uses the Generalized Least Square (GLS) method as the estimator and is applied to the 2020 Maternal Mortality Rate data. The results obtained are the best model combining the level 1 models using 1 (one) knot point with a GCV value of 32.781 and the level 2 model using 2 (two) knot points with a GCV value of 89.865 with significant variables at level-1 namely the number of pregnant women who carry out K4 and the number of pregnant women who handle obstetric complications and significant variables at level-2 are the percentage of poor people and school enrollment rates

Keywords: Hierarchical Model, Generalized Least Square, Spline Truncated, Maternal Mortality

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

BAB 1
PENDAHULUAN

1.1 Latar Belakang

Analisis regresi merupakan salah satu metode analisis yang digunakan untuk mengetahui pola hubungan suatu peristiwa. Analisis regresi mengkaji pola hubungan antara satu variabel respon dengan satu atau lebih variabel penjelas atau variabel bebas (Montgomery et al., 2012). Analisis regresi memiliki banyak pendekatan diantaranya, analisis regresi parametrik, analisis regresi semiparametrik dan analisis regresi nonparametrik. Dalam analisis regresi parametrik kurva regresi diasumsikan diketahui misalnya kurva tersebut linier, kuadratik, kubik dan lain sebagainya. Akan tetapi, apabila kurva regresi tidak diketahui maka digunakan analisis regresi nonparametrik. Regresi nonparametrik digunakan karena memiliki fleksibilitas yang tinggi (Budiantara, 2011).

Menurut Budiantara (2011), salah satu estimator yang dikembangkan dalam melakukan estimasi kurva regresi nonparametrik adalah *spline* karena metode *spline* dapat menyesuaikan pola data secara efektif. Beberapa jenis *spline* yang telah dikembangkan diantaranya *Spline Truncated* (Islamiyati, 2018), *Spline Smoothing* (Lestari, et al., 2010) dan *Penalized Spline* (Islamiyati, et al., 2020). *Spline Truncated* adalah salah satu bentuk *spline* yang bersifat polynomial, yaitu suatu potongan-potongan polinom yang memiliki sifat tersegmen pada selang yang terbentuk pada titik-titik knot (Wang & Lijian, 2009). Bentuk *spline linear truncated* banyak digunakan dalam analisis data karena kemudahan interpretasi secara visual akibat keterlibatan titik knot dalam estimasi fungsi. Titik knot adalah titik terjadinya pola perubahan data dan terdapat kriteria yang harus dipenuhi dalam pemilihan knot yang optimal. Salah satu cara pemilihan kriteria titik knot optimal yaitu dengan *Generalized Cross Validation (GCV)* (Budiantara, 2011).

Pada perkembangan regresi nonparametrik, beberapa peneliti telah mengembangkan model regresi *spline* yang berhierarki. Pemodelan hierarki telah banyak membantu dalam mengetahui hubungan antara variabel respon dan variabel penjelas untuk data yang memiliki pola tersarang. Menurut Hox (1995) data yang berstruktur hierarki merupakan data yang timbul karena individu-individu terkumpul dalam kelompok-kelompok yang sama dengan karakteristik yang sama pula. Metode analisis yang dapat digunakan untuk menganalisis hal ini adalah dengan pemodelan hierarki. Goldstein (1995) menyebutkan bahwa model hierarki dapat mengatasi berbagai masalah yang muncul dari data dengan struktur hierarki termasuk data pengukuran berulang. Struktur hierarki dalam model hierarki biasanya disebut struktur level secara umum struktur level yang digunakan tidak terbatas meskipun biasanya hanya dua yang digunakan yaitu level rendah atau level 1 dan level tinggi atau level 2. Selain dapat menentukan keragaman levelnya, model hierarki juga dapat menunjukkan korelasi antar dua individu yang pada model lain diasumsikan tidak ada (Goldstein, 2010). Beberapa penelitian mengenai regresi *spline* hierarki ditunjukkan dalam Howe, *et al.*, (2016) yang meneliti mengenai model multilevel *spline linear* pada data perkembangan anak, dan Macdonald-Wallis, et al. (2012) meneliti data pertumbuhan yang pararel dengan menggunakan model *spline* multilevel multivariat. Pemodelan multilevel kernel yang diteliti oleh He, et al. (2017) dan aplikasinya ke klasifikasi *neuroimaging*. Dari hasil penelitian ini diketahui bahwa model *spline linear* hierarki dapat mengatasi beberapa isu seperti kolinearitas pada pengukuran berulang dan bias karena data yang hilang.

Selain itu, beberapa peneliti juga telah mengusulkan metode-metode yang berbeda untuk menaksir parameter dalam model regresi data yang berhierarki. Metode klasik yang popular digunakan yaitu metode kuadrat terkecil atau *Ordinary Least Square* (OLS) yang dapat menaksir parameter dengan menggunakan asumsi data berdistribusi normal. Namun, terdapat beberapa kelemahan pada OLS dalam mengatasi masalah heteroskedastisitas. Untuk mengatasi hal ini, beberapa peneliti mengusulkan suatu transformasi menggunakan pembobot terhadap variabel-variabel asli pada model regresi sehingga sebuah penduga dapat memenuhi asumsi standar pada model statistik. (Zurimi, 2018).

Pada penelitian ini, model regresi *spline* hierarki dikembangkan pada penggunaan GLS dalam estimasi parameternya. Goldstein (1995) dalam Tantular (2011) mengusulkan menggunakan metode kuadrat terkecil umum (*Generalized Least Square*) untuk menaksir parameter tetap pada model hierarki. Metode ini dinilai lebih baik dari metode sebelumnya karena model yang digunakan merupakan model yang telah disubstitusikan sehingga struktur varians-kovarians yang digunakan terdiri dari komponen Level 1 dan Level 2. Selain itu, menurut

Alaba, et al. (2010) yang dikutip dari Laili (2016) bahwa model GLS baik digunakan karena galat (*error*) yang dihasilkan lebih kecil. Metode GLS menghasilkan sifat-sifat penaksir yang baik diantaranya adalah tak bias, konsisten dan varians minimum dibandingkan dengan menggunakan OLS (Widyaningsih, 2014). Metode lain yang juga menggunakan kuadrat terkecil yaitu Iterative Generalised Least Square (IGLS) (Longford, 1987). Metode penaksiran parameter yang melibatkan asumsi distribusi pada model multilevel yaitu metode kemungkinan maksimum yang disebut Restricted Maximum Likelihood (REML) (Kuswanto, 2022) .

Angka Kematian Ibu (*maternal mortality rate*) merupakan jumlah kematian ibu akibat dari proses kehamilan, persalinan, dan pasca persalinan yang dijadikan indikator derajat kesehatan perempuan. Angka Kematian Ibu (AKI) juga merupakan salah satu indikator utama keberhasilan program kesehatan ibu. Selain itu, indikator ini juga mampu menilai derajat kesehatan masyarakat, karena sensitifitasnya terhadap perbaikan pelayanan kesehatan, baik dari sisi aksesibilitas maupun kualitas (Kemenkes RI, 2020). Penelitian mengenai pemodelan regresi nonparametrik *spline truncated* terhadap angka kematian ibu telah banyak dilakukan. Penelitian Widhi (2020) menyatakan bahwa angka kematian ibu dipengaruhi secara signifikan oleh persentase persalinan oleh tenaga kesehatan, persentase komplikasi kebidanan yang ditangani, persentase bayi lahir berat badan rendah, persentase penduduk miskin dan persentase perempuan kawin di bawah 17 tahun. Berdasarkan penelitian tersebut dapat diketahui bahwa variabel yang dapat berpengaruh signifikan terhadap angka kematian ibu ada yang datang dari internal maupun eksternal. Secara internal, angka kematian ibu dipengaruhi oleh hal yang berkaitan dengan ibu secara langsung dan secara eksternal angka kematian ibu dipengaruhi oleh faktor pendidikan dan tingkat kemiskinan. Merujuk pada variabel yang berpengaruh diatas, dapat diidentifikasi bahwa persentase penduduk miskin dan persentase perempuan kawin dibawah 17 tahun berada di level yang berbeda.

Berdasarkan uraian diatas, peneliti mengkaji model *spline* hierarki pada data angka kematian ibu dengan menggunakan estimasi parameter *Generalized Least Square.* Metode diaplikasikan pada data jumlah kematian ibu sebagai variabel respon dan pada variabel penjelas di level 1 (kabupaten) ialah, jumlah ibu hamil yang mendapatkan tablet tambah darah (x_1), jumlah ibu hamil yang melaksanakan K4 (x_2), jumlah penanganan komplikasi kebidanan (x_3) dan di level

2 (provinsi) persentase penduduk miskin (z_1) dan angka partisipasi sekolah (z_2) yang diduga berpengaruh terhadap jumlah kematian ibu.

1.2 Rumusan Masalah

Berdasarkan latar belakang penelitian maka rumusan masalah dalam penelitian ini meliputi:

1. Bagaimana bentuk estimasi parameter model regresi *spline* hierarki dengan metode *Generalized Least Square* ?
2. Bagaimana model regresi *spline* hierarki pada data kematian ibu di beberapa provinsi di Indonesia tahun 2020?

1.3 Batasan Masalah

Agar penelitian ini tidak menimbulkan penafsiran yang lebih luas maka diberikan beberapa asumsi sebagai batasan masalah dari penelitian ini yaitu antara lain:

1. Pada penggunaan regresi *spline* dibatasi pada *spline truncated* dengan orde satu dan dua titik knot
2. Model hierarki yang digunakan dibatasi hingga dua level, yaitu level 1 (Kabupaten) dan level 2 (Provinsi)
3. Data yang digunakan adalah data Profil Kesehatan Indonesia 2020 dan data profil kesehatan pada 16 (enam belas) provinsi di Indonesia

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan maka tujuan penulisan penelitian ini antara lain:

1. Untuk mengetahui bentuk estimasi parameter model regresi *spline* hierarki dengan metode *Generalized Least Square.*
2. Untuk mengetahui model regresi *spline* hierarki pada data angka kematian ibu di beberapa provinsi di Indonesia tahun 2020.

1.5 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat memberi pemahaman tentang bentuk penerapan Model Regresi *Spline* Hierarki dengan pendekatan *Generalized Least Square* pada data angka kematian ibu di beberapa provinsi di Indonesia serta dapat memberi informasi kepada *stakeholder* terkait maupun pemerintah dalam mengambil suatu kebijakan atau keputusan yang lebih baik.

BAB 2
TINJAUAN PUSTAKA

2.1 Model Regresi

Model regresi merupakan suatu model matematis yang digunakan untuk melihat hubungan antara variabel respon dengan satu atau lebih variabel penjelas (Hosmer dan Lemeshow, 2000). Hubungan yang menunjukkan hubungan yang linear antara variabel respon dan variabel penjelas disebut model regresi linear. Terdapat dua persamaan model dalam model regresi linear, yaitu regresi linear sederhana (*simple regression*) dan regresi linear berganda (*multiple regression*). Regresi linear sederhana merupakan model regresi yang hanya memiliki satu variabel penjelas, sedangkan, regresi linear berganda merupakan model regresi yang memiliki lebih dari satu variabel penjelas.

Secara umum, persamaan model regresi dituliskan sebagai berikut:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_r x_{ri} + \varepsilon_i \\ &= \beta_0 + \textstyle\sum_{p=1}^{r} \beta_p x_{pi} + \varepsilon_i \end{aligned} \quad (2.1)$$

dimana:

y_i : variabel respon untuk data ke-*i* , *i* = 1, 2, …, n

x_{1i} : variabel penjelas ke-*1* pada data ke-*i*

ε_i : *error*

β_0 : konstanta/intersep

β_{pi} : koefisien dari variabel penjelas ke-p pada pengamatan ke-i

i : banyaknya data, $i = 1,2, \dots, n$

p : variabel penjelas $p = 1,2, \dots, r$

2.2 Model Regresi Nonparametrik

Regresi Nonparametrik adalah satu metode yang digunakan untuk menjelaskan suatu pola hubungan antara variabel respon dan variabel penjelas dimana kurva fungsi regresi tidak diketahui. Model regresi nonparametrik secara umum dapat ditunjukkan pada persamaan (2.2) sebagai berikut.

$$y_i = f(x_i) + \varepsilon_i \quad , i = 1,2, \dots, n \quad (2.2)$$

Dimana :

y_i : variabel respon pada data ke-*i*

x_i : variabel penjelas pada data ke-*i*

$f(x_i)$: fungsi regresi nonparametrik yang memuat variabel penjelas

ε_i : *error*

Pendekatan nonparametrik digunakan untuk mengestimasi kurva regresi karena model-modelnya tidak ditentukan terlebih dahulu seperti pada regresi parametrik. Apabila bentuk kurva regresi tidak diketahui polanya, maka model regresi nonparametrik lebih disarankan untuk digunakan. Pendekatan nonparametrik juga lebih fleksibel karena tidak dibatasi oleh asumsi-asumsi seperti pada pendekatan parametrik.

2.3 Model Regresi *Spline Truncated*

Salah satu model regresi dengan pendekatan nonparametrik yang dapat digunakan adalah *Spline*. *Spline* adalah salah satu jenis polinomial tersegmen yang memiliki sifat fleksibilitas. Sifat fleksibilitas ini yang membedakan *spline* dari polinomial, sehingga memungkinkan untuk menyesuaikan diri secara lebih efektif terhadap karakteristik suatu fungsi atau data (Budiantara, 2006). Metode *spline* dalam regresi nonparametrik dapat ditemui dalam banyak bentuk, diantaranya adalah *smoothing spline* dan *spline truncated*. Kedua metode *spline* tersebut masing-masing menggunakan parameter yang berbeda untuk membuat estimasi kurva regresi lebih fleksibel, yaitu parameter penghalus pada *smoothing spline* dan titik knot pada *spline truncated*. Perbedaan jenis parameter tersebut, menyebabkan optimasi untuk mendapatkan estimator pada kedua metode *spline* tersebut berbeda.

Spline truncated mempunyai sifat fleksibilitas yang tinggi dan mempunyai kemampuan mengestimasi perilaku data yang cenderung berbeda pada interval berlainan (Eubank, 1999). Kemampuan tersebut ditunjukkan dengan fungsi *truncated* (potongan-potongan), dimana titik perpaduan bersama dari potongan-potongan tersebut, menunjukkan terjadinya perubahan pola perilaku fungsi *spline* pada selang yang berbeda dan disebut sebagai titik knot (Hardle, 1990).

Secara umum fungsi *spline linear truncated* berorde d dengan titik-titik knot pada $c_1, c_2, \dots, c_b$ dapat dinyatakan dalam bentuk sebagai berikut: (Rodrıguez, 2001):

$$f(x_i) = \sum_{h=0}^{d} \beta_h x_i^h + \sum_{g=1}^{b} \beta_{d+g}(x_i - C_g)_+^d, \tag{2.3}$$

dengan $\beta_0, \beta_1, \dots, \beta_d, \beta_{d+1}, \dots, \beta_{d+g}$ adalah parameter regresi, C_g adalah titik knot ke-g, $(g = 1,2, \dots, b)$. Untuk $(x_i - C_g)_+^d$ adalah fungsi polinomial *truncated* yang dijabarkan sebagai berikut:

$$(x_i - C_g)_+^d = \begin{cases} (x_i - C_g)^d, \text{jika } x_i \geq C_g \\ 0, \qquad\qquad \text{jika } x_i < C_g \end{cases}.$$

Jika fungsi yang menyatakan hubungan antara r penjelas dengan respon tunggal maka fungsi *spline* $f(x_i)$ dalam Persamaan (2.3) dapat dinyatakan dalam bentuk sebagai berikut:

$$\begin{aligned} y_i &= f(x_{1i}) + f(x_{2i}) + \dots + f(x_{ri}) + \varepsilon_i \\ &= \sum_{p=1}^{r} f(x_{pi}) + \varepsilon_i; i = 1,2, \dots, n, \end{aligned} \tag{2.4}$$

Fungsi $f(x_{pi})$ dalam persamaan (2.4) dapat dinyatakan sebagai berikut:

$$f(x_{pi}) = \beta_0 + \sum_{h=1}^{d} \beta_{ph} x_{pi}^h + \sum_{g=1}^{b} \beta_{p(d+g)}(x_{pi} - C_{pg})_+^d \tag{2.5}$$

dengan

y_i : variabel respon pada pengamatan ke-i,

x_{pi} : variabel penjelas ke-p pada pengamatan ke-i,

β_{0p} : intersep penjelas ke-p,

β_{ph} : parameter polinomial pada penjelas ke-p dan orde ke-h,

C_{pb} : nilai titik knot pada penjelas ke-p dan titik knot ke-b,

b : banyaknya titik knot,

d : orde polinomial *Spline Linear Truncated* ,

p : banyaknya variabel penjelas,

$\beta_{p(d+b)}$: parameter *truncated* pada penjelas ke-p dan titik knot ke-$(d + b)$,

ε_i : galat *(error)* pada pengamatan ke-*i* yang diasumsikan saling bebas berdistribusi normal dengan mean 0 dan variansi σ^2 *.*

Persamaan (2.4) untuk n data pengamatan dan orde $d = 1$ dapat dinyatakan dalam bentuk matriks sebagai berikut:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$= \begin{bmatrix} 1 & x_{11} & \cdots & x_{r1} & (x_{11} - C_{11})_+ & \cdots & (x_{r1} - C_{r1})_+ & \cdots & (x_{11} - C_{1b})_+ & \cdots & (x_{r1} - C_{rb})_+ \\ 1 & x_{12} & \cdots & x_{r2} & (x_{11} - C_{11})_+ & \cdots & (x_{r2} - C_{r1})_+ & \cdots & (x_{12} - C_{1b})_+ & \dots & (x_{r2} - C_{rb})_+ \\ \vdots & \vdots & \ddots & & & \vdots & & & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \dots & x_{rn} & (x_{1n} - C_{11})_+ & \cdots & (x_{r2} - C_{r1})_+ & \cdots & (x_{1n} - C_{1b})_+ & \dots & (x_{rn} - C_{rb})_+ \end{bmatrix}$$

$$\begin{bmatrix} \beta_0 \\ \beta_{r1} \\ \beta r_2 \\ \vdots \\ \beta_{rd} \\ \beta_{r(d+1)} \\ \vdots \\ \beta_{r(d+b)} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

dengan y adalah vektor yang memiliki ukuran $n \times 1$, matriks X yang memiliki ukuran $n \times (1 + d + b)$, vektor β ukuran $(1 + d + b) \times 1$, dan vektor ε ukuran $n \times 1$ (Budiantara, 2006)

2.4 Pemilihan Titik Knot Optimum

Dalam regresi *spline*, estimator *spline* terbaik dipilih berdasarkan pemilihan titik knot optimal. Titik knot adalah titik temu dimana terdapat sebuah perubahan dari pola kebiasaan atau fungsi kurva. Salah satu metode yang sering digunakan dalam memilih titik knot yang optimal adalah dengan menggunakan metode *General Cross Validation* (GCV) (Hardle, 1990). Metode GCV memiliki sifat optimal asimtotik yang artinya dapat digunakan untuk sampel besar yang mana hal ini tidak dimiliki oleh metode lain, misalnya *Cross Validation* (Wahba, 1990). Sampel besar dalam hal ini ialah jumlah data yang lebih dari 30. Metode ini juga memiliki sifat invarian terhadap transformasi yang artinya data dapat ditransformasi jika asumsi residual tidak terpenuhi. Model regresi *spline* terbaik diperoleh dari titik knot optimal dengan melihat nilai GCV terkecil. Metode GCV bisa dituliskan sebagai berikut (Eubank, 1988).

$$GCV\ (c) = \frac{MSE\ (c)}{[n^{-1} trace\ (\boldsymbol{I} - \boldsymbol{A})]^2} \tag{2.6}$$

Dimana matriks ***A*** adalah $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ dan matriks ***I*** adalah matriks Identitas, n merupakan banyaknya observasi, $c = (c_1, c_2, \ldots, c_g)$ merupakan titik-titik knot, dan

$$MSE\ (c) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}(x_i)\right)^2 \tag{2.7}$$

2.5 Model Hierarki

Model hierarki merupakan teknik yang digunakan untuk melakukan analisis data yang memiliki struktur hierarki. Data yang berstruktur hierarki merupakan data yang individu-individunya terkumpul dalam kelompok-kelompok yang sama dimana level yang lebih rendah tersarang pada level yang lebih tinggi dan banyaknya individu-individu dalam kelompok bisa saja sama atau berbeda untuk masing-masing kelompok. Data hierarki disebut juga data multilevel atau data bersarang, (Tantular, 2016). Pada data berstruktur hierarki atau data berjenjang, variabel respon diukur pada level 1 atau di tingkat terendah saja, sedangkan variabel penjelas diukur pada level 1 dan level yang lebih tinggi (Goldstein, 2011). Pada pemodelan multilevel dengan 2-level, terdapat sebanyak n individu yang berasal dari j kelompok. $y_{1j}, y_{2j}, \ldots, y_{n_ij}$ adalah peubah respon masing-masing n_j individu pada kelompok ke-j, $j = 1,2, \ldots, k$ dan jika $x_{1j}, x_{2j}, \ldots, x_{rj}$ adalah peubah penjelas pada level 1 untuk kelompok ke-j, serta $z_1, z_2, \ldots, z_l$ adalah peubah penjelas pada level 2, maka struktur pemodelan multilevel dengan 2-level. Data hierarki disebut juga sebagai data multilevel (Hox, 2002). Jika terdapat data dengan kondisi struktur hierarki atau data berjenjang yang memiliki 2 level maka:

a. Terdapat k buah kelompok dengan banyaknya pengamatan pada masing-masing kelompok sebesar $n_1, n_2, \ldots, n_k$.

b. Variabel respon (Y) diamati hanya pada level 1.

c. Variabel penjelas yang memiliki 2 level dengan level 1 adalah variabel penjelas x_1 sampai x_r dan level 2 adalah variabel penjelas z_1 sampai z_l.

Adapun struktur dari data hierarki 2 level ditunjukkan pada Tabel 2.1 sebagai berikut:

Tabel 2.1 Struktur Data Multilevel dengan 2 level

J	*I*	Peubah Respon	Peubah Penjelas Level 1				Peubah Penjelas Level 2			
		y_{ij}	x_{1i}	x_{2i}	...	x_{ri}	z_1	z_2	...	z_l
1	1	y_{11}	x_{111}	x_{211}	...	x_{r11}	z_{11}	z_{21}	...	z_{l1}
	2	y_{21}	x_{121}	x_{221}		x_{r21}				
	⋮	⋮	⋮	⋮		⋮				
	n_1	$y_{n_1 1}$	$x_{1n_1 1}$	$x_{2n_1 1}$		$x_{rn_1 1}$				
2	1	y_{12}	x_{112}	x_{212}	...	x_{r12}	z_{12}	z_{22}	...	z_{l2}
	2	y_{22}	x_{122}	x_{222}		x_{r22}				
	⋮	⋮	⋮	⋮		⋮				
	n_2	$y_{n_2 2}$	$x_{1n_2 2}$	$x_{2n_2 2}$		$x_{rn_2 2}$				
⋮	⋮	⋮	⋮	⋮	...	⋮	⋮	⋮	...	⋮
K	1	y_{1k}	x_{11k}	x_{21k}	...	x_{r1k}	z_{1k}	z_{2k}	...	z_{lk}
	2	y_{2k}	x_{12k}	x_{22k}		x_{r2k}				
	⋮	⋮	⋮	⋮		⋮				
	n_k	$y_{n_k k}$	$x_{1n_k k}$	$x_{2n_k k}$		$x_{rn_k k}$				

Dengan $i = 1,2, \dots, n_j$; $p = 1,2, \dots, r$; $j = 1,2, \dots, k$; $q = 1,2, \dots, l$

y_{ij} : variabel respon untuk pengamatan ke-*i* pada level 1 dalam kelompok ke-j pada level 2.

x_{pi} : variabel penjelas ke-*p* di level 1 untuk pengamatan ke-*i* pada level 2

z_{qj} : variabel penjelas ke-*q* di level 2 untuk kelompok ke-*j* pada level 2

Suatu variabel bebas dikatakan memiliki efek tetap jika koefisien regresinya bernilai sama bagi seluruh anggota sampel dan suatu variabel bebas dikatakan memiliki efek acak jika nilai koefisien regresinya berbeda antar dua atau lebih grup anggota sampel (Harlan, 2016). Adapun Persamaan model linear campuran dalam bentuk sederhana adalah (Rencher & Schaalje, 2007):

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\varepsilon} \tag{2.8}$$

dengan:

$\boldsymbol{y}$ **:** vektor variabel respon

$\boldsymbol{X}$ **:** matriks variabel penjelas untuk parameter tetap

$\boldsymbol{\beta}$: vektor parameter efek tetap

$\boldsymbol{Z}$ **:** matriks variabel penjelas untuk parameter acak

$\boldsymbol{u}$ **:** vektor efek acak

$\boldsymbol{\varepsilon}$ **:** vektor galat acak

dengan asumsi antar galat $\boldsymbol{\varepsilon}$ dan $\boldsymbol{u}$ tidak saling bebas dan berdistribusi normal dengan variansi θ yang dapat ditulis sebagai berikut:

$$\begin{bmatrix}\boldsymbol{u}\\ \boldsymbol{\varepsilon}\end{bmatrix} \sim N\left(\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}, \sigma^2\begin{bmatrix}\boldsymbol{G} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{R}\end{bmatrix}\right) \tag{2.9}$$

dengan $\boldsymbol{G} = \boldsymbol{G}(\gamma)$ dan $\boldsymbol{R} = \boldsymbol{\Sigma}$ dengan $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\phi)$**.** γ dan ϕ adalah vektor parameter variansi yang terikat dengan u dan ε sedangkan σ^2 adalah parameter skala variansi. Parameter variansi yaitu adalah σ^2, ϕ dan γ**.** $\boldsymbol{G}, \boldsymbol{R}$ dan $\boldsymbol{\Sigma}$ merupakan matriks variansi yang diasumsikan definit positif.

$$\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \boldsymbol{H}) \tag{2.10}$$

dengan

$$\boldsymbol{H} = \boldsymbol{ZGZ'} + \boldsymbol{R} \tag{2.11}$$

Secara umum model regresi multilevel hanya terdiri dari 2-level. Gabungan level 1 dan level 2 merupakan model campuran dengan *fixed* dan *random* koefisien regresi. Secara umum model regresi linier 2 level mendefinisikan variabel penjelas pada level individu (level 1) dan pada level kelompok (level 2). Berdasarkan struktur data pada Tabel 2.1 pemodelan multilevel dengan 2 level adalah sebagai berikut (Goldstein, 2001):

1. Model Level-1

Model 1 adalah model yang disusun tanpa memperhatikan pengaruh dari level kelompok. Pemodelan multilevel untuk tiap kelompok dapat ditulis sebagai beikut.

$$\begin{aligned} y_{ij} &= \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \cdots + \beta_{rj}x_{rij} + \varepsilon_{ij} \\ &= \beta_{0j} + \textstyle\sum_{p=1}^{r} \beta_{pj}x_{pij} + \varepsilon_{ij} \end{aligned} \tag{2.12}$$

dengan $p = 1,2,\dots,r$, individu-$i = 1,2,\dots,n_j$ dan kelompok-$j = 1,2,\dots,k$

y_{ij} : variabel respon untuk pengamatan ke-i pada level 1 dalam kelompok ke-j pada level 2

β_{0j} : intersep kelompok ke-j pada level 2

β_{pj} : intersep untuk kelompok ke-j pada level 2 yang berasal dari koefisien regresi untuk variabel predikto ke-p pada level 1

x_{pij} : variabel penjelas ke-p di level 1 untuk pengamatan ke-i pada level 1 dalam kelompok ke-j pada level 2

ε_{ij} : eror untuk pengamatan ke-i pada level 1 dalam kelompok ke-j pada level 2

atau dapat dinotasikan dalam matriks berikut

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}, \text{ dengan } \boldsymbol{\varepsilon} \sim N(0, \sigma^2 I) \qquad (2.13)$$

$$\begin{bmatrix} y_{1j} \\ y_2 j \\ \vdots \\ y_{n_k j} \end{bmatrix} = \begin{bmatrix} 1 & x_{11j} & x_{21j} & \cdots & x_{r1j} \\ 1 & x_{12j} & x_{22j} & \cdots & x_{r2j} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n_1 j} & x_{2n_2 j} & \cdots & x_{rn_k j} \end{bmatrix} \begin{bmatrix} \beta_{0j} \\ \beta_{1j} \\ \vdots \\ \beta_{rj} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1j} \\ \varepsilon_{2j} \\ \vdots \\ \varepsilon_{n_k j} \end{bmatrix}$$

dengan:

$\boldsymbol{y}$: vektor variabel respon

$\boldsymbol{X}$: matriks variabel penjelas untuk parameter tetap

$\boldsymbol{\beta}$: vektor parameter efek tetap

$\boldsymbol{\varepsilon}$: vektor galat

2. Model Level 2

Koefisien regresi pada level-1, β_{pj} dengan $p = 0,1,2,\ldots,r$ dalam model level-1 memiliki nilai yang berbeda antar kelompok. Variasi nilai β_{pj} akan dijelaskan dengan membentuk model level 2. Pembentukan model level 2 dilakukan untuk setiap koefisien regresi sebagai respon ke-*q* dengan menggunakan variabel penjelas pada level-2. Bentuk pemodelan pada level-2 dapat ditulis sebagai berikut: (Zulvia, 2017):

$$\beta_{0j} = \gamma_{00} + \gamma_{01} z_{1j} + \gamma_{02} z_{2j} + \cdots + \gamma_{0l} z_{lj} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} z_{1j} + \gamma_{12} z_{2j} + \cdots + \gamma_{1l} z_{lj} + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21} z_{1j} + \gamma_{22} z_{2j} + \cdots + \gamma_{2l} z_{lj} + u_{2j}$$

$$\beta_{rj} = \gamma_{r0} + \gamma_{r1} z_{1j} + \gamma_{r2} z_{2j} + \cdots + \gamma_{rl} z_{lj} + u_{rj}$$

$$\beta_{pj} = \gamma_{p0} + \sum_{q=1}^{l} \gamma_{pq} z_{qj} + u_{pj} \qquad (2.14)$$

dengan $p = 0,1,2,\ldots,r$, dan $q = 0, 1,2,\ldots,l$

β_{pj} : intersep untuk kelompok ke-j pada level 2 yang berasal dari koefisien regresi untuk variabel penjelas ke-p pada level 1

γ_{0q} : koefisien regresi untuk variabel penjelas ke-q pada level 2 yang memprediksi koefisien regresi pada intersep persamaan regresi level 1

γ_{pq} : koefisien regresi untuk variabel penjelas ke-q pada level 2 yang memprediksi keragaman pada koefisien regresi variabel ke-p persamaan regresi level 1

Z_{qj} : variabel penjelas ke-q untuk kelompok ke-j pada level 2

u_{pj} : eror untuk kelompok ke-j pada variabel penjelas ke-p, diasumsikan menyebar $N(0,\sigma_\varepsilon^2)$

atau dinyatakan dalam matriks berikut

$$\boldsymbol{\beta} = \boldsymbol{Z\gamma} + \boldsymbol{u}, \text{ dengan } \boldsymbol{u}_p \sim N(0,\sigma_u^2 I) \qquad (2.15)$$

$$\begin{bmatrix} \beta_{p1} \\ \beta_{p2} \\ \vdots \\ \beta_{pk} \end{bmatrix} = \begin{bmatrix} 1 & z_{11} & z_{21} & \dots & z_{l1} \\ 1 & z_{12} & z_{22} & \dots & z_{l2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & z_{1k} & z_{2k} & \dots & z_{lk} \end{bmatrix} \begin{bmatrix} \gamma_{p1} \\ \gamma_{p2} \\ \vdots \\ \gamma_{pl} \end{bmatrix} + \begin{bmatrix} u_{p1} \\ u_{p2} \\ \vdots \\ u_{pk} \end{bmatrix}$$

Jika Persamaan (2.13) disubstitusikan pada Persamaan (2.15) maka diperoleh model sebagai berikut:

$$y_{ij} = \beta_{0j}x_{0ij} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \dots + \beta_{rj}x_{rij} + \varepsilon_{ij}$$

$$y_{ij} = (\gamma_{00} + \gamma_{01}z_{1j} + \gamma_{02}z_{2j} + \dots + \gamma_{0l}z_{lj} + u_{0j}) + (\gamma_{10} + \gamma_{11}z_{1j} + \gamma_{12}z_{2j} + \dots + \gamma_{1l}z_{lj} + u_{1j})x_{1ij} + \dots + (\gamma_{20} + \gamma_{21}z_{1j} + \gamma_{22}z_{2j} + \dots + \gamma_{2l}z_{lj} + u_{2j})x_{2ij} + \dots + (\gamma_{r0} + \gamma_{r1}z_{1j} + \gamma_{r2}z_{2j} + \dots + \gamma_{rl}z_{lj} + u_{rj})x_{rij} + \varepsilon_{ij}$$

$$y_{ij} = \gamma_{00} + \sum_{q=1}^{l} z_{qj}\gamma_{0q} + \sum_{p=1}^{r} x_{pij}\gamma_{pq} + \sum_{p=1}^{r}\sum_{q=1}^{l} x_{pij}z_{qj}\gamma_{pq} + \sum_{p=1}^{r} x_{pij}u_{pj} + \varepsilon_{ij} \qquad (2.16)$$

atau dinyatakan dalam bentuk matriks sebagai berikut:

$$\boldsymbol{y}_j = \boldsymbol{X}_j\boldsymbol{Z}_j\boldsymbol{\gamma} + \boldsymbol{X}_j\boldsymbol{u}_j + \boldsymbol{\varepsilon}_j \qquad (2.17)$$

dengan $\boldsymbol{X}_j\boldsymbol{Z}_j\boldsymbol{\gamma}$ adalah efek tetap dan $[\boldsymbol{X}_j\boldsymbol{u}_j + \boldsymbol{\varepsilon}_j]$ adalah efek acak.

Hox (2010) menyatakan bahwa matriks $\boldsymbol{Z}$ dalam Persamaan (2.17) merupakan variabel perantara untuk menghubungkan $\boldsymbol{\gamma}$ dan $\boldsymbol{X}$. Oleh karena itu, variasi hubungan antara $\boldsymbol{\gamma}$ dan $\boldsymbol{X}$ tergantung pada $\boldsymbol{R}$. Interpretasi dari koefisien regresi pada model level 1 dan koefisien regresi model level 2 terhadap y tergantung pada tanda positif dan negatif dari kedua koefisien regresi tersebut. Jika koefisien regresi model level 1 dan koefisien regresi model level 2 bertanda sama berarti kedua koefisien regresi tersebut memiliki pengaruh berbanding lurus terhadap y. Pada $\boldsymbol{X}_j\boldsymbol{Z}_j$ merupakan faktor interaksi dalam model sebagai variansi *slope* variabel $\boldsymbol{X}$.

Adapun model linier hierarki disebut juga *Hierarchial Level Model* (HLM), adalah model koefisien random, model covariance component, serta model dengan efek acak tersarang yaitu model linier dengan variabel respon yang nilainya tidak hanya dipengaruhi oleh variabel tetap (fixed) saja namun juga dipengaruhi oleh variabel acak (random). Istilah HLM memberikan dua fitur dalam mendefinisikan model. Hal pertama, data yang sesuai untuk model berstruktur hierarki, dengan unit level pertama tersarang dalam unit level kedua, unit level kedua tersarang di dalam unit level ketiga, dan seterusnya. Sedangkan hal kedua, parameter model tersebut terlihat seperti memiliki struktur hierarki (Raudenbush dan Bryk, 2002).

2.6 Estimasi Parameter *Generalized Least Square*

Pada estimasi OLS (*Ordinary Least Square*) salah satu asumsi yang harus dipenuhi adalah asumsi homoskedastisitas, namun apabila terjadi pelanggaran asumsi tersebut, yakni variansinya tidak sama (heteroskedastisistas), maka metode yang dapat digunakan untuk menduga koefisien regresi adalah metode *Generalized Least Square* (GLS). Menurut Gujarati dan Porter (2010), GLS merupakan OLS yang telah mengalami proses transformasi menggunakan pembobot pada variabel-variabel asli sehingga menghasilkan penduga yang bersifat BLUE (*Best Linear Unbiased Estimator*). Sebuah penduga dikatakan bersifat BLUE jika memenuhi asumsi standar model statistik linier yaitu linier, tak bias, dan mempunyai varians yang minimum (Aziz, 2007). Dalam GLS, diperoleh jumlah kuadrat galat yang telah diberi pembobot. Perhatikan kembali persamaan

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2.18}$$

Menurut Baltagi (2008) pada kasus autokorelasi $\boldsymbol{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0}$ dan $\boldsymbol{var}(\boldsymbol{\varepsilon}) = \sigma^2\,\mathbf{V}$ dan $\boldsymbol{\varepsilon} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{\sigma}^2\mathbf{V})$ dimana **V** merupakan matriks simetris dan definit positif, sehingga **V** dapat difaktorisasikan menjadi

$$\boldsymbol{V} = \boldsymbol{CDC}' \tag{2.19}$$

Matriks ***D*** adalah matriks diagonal yang elemennya merupakan nilai-nilai eigen dari $var(\varepsilon)$.

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

dan $\boldsymbol{C}$ adalah matriks *orthogonal*. Selanjutnya invers persamaan (2.19), Maka diperoleh:

$$\begin{aligned}\boldsymbol{V^{-1}} &= \boldsymbol{C^{-1}D^{-1}C'^{-1}}\\ &= \boldsymbol{C^{-1}D^{-1/2}D'^{1/2}C'^{-1}}\\ &= \boldsymbol{PP'} \qquad (2.20)\end{aligned}$$

$\boldsymbol{P} = \boldsymbol{C^{-1}D^{-1/2}}$ dan $\boldsymbol{D^{-1/2}}$ adalah sebagai berikut

$$D^{-1/2} = \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{\lambda_n}} \end{bmatrix}$$

Dan $\boldsymbol{PVP'} = \boldsymbol{I}$**.** Dari persamaan model statistik linier diperoleh transformasi model dengan mengalikan matriks $\boldsymbol{P}$ ke persamaan (2.18) sehingga menjadi,

$$\boldsymbol{Py} = \boldsymbol{P}\,(\boldsymbol{X\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{PX\beta} + \boldsymbol{P\varepsilon} \qquad (2.21)$$

Atau

$$\boldsymbol{y}^* = \boldsymbol{X}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^* \qquad (2.22)$$

Dimana $\boldsymbol{E}(\boldsymbol{\varepsilon}^*) = \boldsymbol{E}\,(\boldsymbol{P\varepsilon}) = \boldsymbol{PE}(\boldsymbol{\varepsilon}) = \boldsymbol{0}$ dan

$$\begin{aligned}\boldsymbol{var}\,(\boldsymbol{\varepsilon}^*) &= \boldsymbol{var}\,(\boldsymbol{P\varepsilon})\\ \boldsymbol{var}\,(\boldsymbol{\varepsilon}^*) &= \boldsymbol{P}\,\boldsymbol{var}\,(\boldsymbol{\varepsilon})\,\boldsymbol{P'}\\ \boldsymbol{var}(\boldsymbol{\varepsilon}^*) &= \sigma^2\,\boldsymbol{PVP'}\\ \boldsymbol{var}(\boldsymbol{\varepsilon}^*) &= \sigma^2\boldsymbol{I}\end{aligned}$$

Setelah ditransformasi ternyata model regresi memenuhi asumsi regresi klasik $\boldsymbol{E}(\boldsymbol{\varepsilon}^*) = \boldsymbol{0}$ dan homoskedastisitas $\boldsymbol{var}(\boldsymbol{\varepsilon}^*) = \sigma^2\boldsymbol{I}$**.**

Matriks varian-kovarian $\boldsymbol{V}$ ukuran $n \times n$ adalah *block diagonal* yang dinyatakan dalam bentuk sebagai berikut :

$$\boldsymbol{V} = \begin{bmatrix} V_1 & 0 & \cdots & 0 \\ 0 & V_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & V_k \end{bmatrix} \qquad (2.23)$$

k adalah banyaknya kelompok level 2 yang diobservasi, dan $\boldsymbol{V}_1, \boldsymbol{V}_2, \dots, \boldsymbol{V}_k$ adalah matriks varian-kovarian untuk masing-masing kelompok level 2, yang didefinisikan sebagai berikut:

$$\boldsymbol{V_1} = \sigma_u^2\boldsymbol{J}_{(\boldsymbol{n_1})} + \sigma_\varepsilon^2\boldsymbol{I}_{(\boldsymbol{n_1})}$$

$$\boldsymbol{V_2} = \sigma_u^2 \boldsymbol{J}_{(\boldsymbol{n_2})} + \sigma_\varepsilon^2 \boldsymbol{I}_{(\boldsymbol{n_2})}$$
$$\vdots$$
$$\boldsymbol{V_2} = \sigma_u^2 \boldsymbol{J}_{(\boldsymbol{n_2})} + \sigma_\varepsilon^2 \boldsymbol{I}_{(\boldsymbol{n_2})} \tag{2.24}$$
$$\vdots$$
$$\boldsymbol{V_k} = \sigma_u^2 \boldsymbol{J}_{(\boldsymbol{n_k})} + \sigma_\varepsilon^2 \boldsymbol{I}_{(\boldsymbol{n_k})}$$

Jika dijabarkan, matirks varian-kovarian untuk kelompok ke-*j*, V_j, dijabarkan sebagai berikut:

$$\boldsymbol{V_j} = \begin{bmatrix} \sigma_{u_j}^2 + \sigma_{\varepsilon_j}^2 & \sigma_{u_j}^2 & \cdots & \sigma_{u_j}^2 \\ \sigma_{u_j}^2 & \sigma_{u_j}^2 + \sigma_{\varepsilon_j}^2 & \cdots & \sigma_{u_j}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{u_j}^2 & \sigma_{u_j}^2 & \cdots & \sigma_{u_j}^2 + \sigma_{\varepsilon_j}^2 \end{bmatrix} \tag{2.25}$$

$\boldsymbol{V_j}$ berukuran $n_j \times n_j$, dengan $\boldsymbol{I}_{(\boldsymbol{n_j})}$ adalah matriks identitas berukuran $n_j \times n_j$, $j = 1,2,\ldots,16$ dan $\boldsymbol{J}_{(\boldsymbol{n_j})}$ adalah matriks yang entri-entrinya berisi konstanta 1 ukuran $n_j \times n_j$ dari persamaan (2.24)dan persamaan (2.25), matriks varian-kovarian untuk n obervasi, yang mana $n = \sum_{j=1}^{k} n_j$, dinyatakan dalam bentuk:

$$\boldsymbol{V} = \begin{bmatrix} \sigma_{u_j}^2 \boldsymbol{J}_{(\boldsymbol{n_1})} + \sigma_{\varepsilon_j}^2\, \boldsymbol{I}_{(\boldsymbol{n_1})} & 0 & \cdots & 0 \\ 0 & \sigma_{u_j}^2 \boldsymbol{J}_{(\boldsymbol{n_2})} + \sigma_{\varepsilon_j}^2\, \boldsymbol{I}_{(\boldsymbol{n_2})} & \cdots & \sigma_{u_j}^2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{u_j}^2 \boldsymbol{J}_{(\boldsymbol{n_k})} + \sigma_{\varepsilon_j}^2\, \boldsymbol{I}_{(\boldsymbol{n_k})} \end{bmatrix} \tag{2.26}$$

$$\hat{\sigma}_\varepsilon^2 = \frac{\hat{\boldsymbol{\varepsilon}}\prime\hat{\boldsymbol{\varepsilon}}}{(\,n_j - k - p + \lambda)} \tag{2.27}$$

$$\tilde{\sigma}_u^2 = \frac{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}} - (N-p)\hat{\sigma}_\varepsilon^2}{(N - trace\ [(\boldsymbol{X}'\boldsymbol{X})^{-1} \sum_{j=1}^{k} n_j^2 \overline{\boldsymbol{x}}\prime_{.j} \overline{\boldsymbol{x}}_{.j}])} \tag{2.28}$$

$$\hat{\sigma}_u^2 = \begin{cases} \tilde{\sigma}_u^2, & \tilde{\sigma}_u^2 > 0 \\ 0, & lainnya \end{cases}$$

dengan

k : banyaknya kelompok level 2

p : banyaknya variabel penjelas level 1

$\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}$: jumlah kuadrat eror dari regresi Y terhadap X

λ : jumlah kombinasi linier dari variabel X

$\overline{\boldsymbol{x}}_{.j}$: vektor $(1 \times p)$ yang memiliki elemen ke-*p* $\overline{\boldsymbol{x}}_{p.j},\ p = 1,2,\ldots,r$

N : $\sum_{j=1}^{k} n_j$

2.7 Pengujian Parameter

Secara umum dalam regresi dengan variabel penjelas lebih dari satu (*multivariate*), uji signifikansi dari parameter β yang diperoleh menggunakan dua cara yaitu pengujian secara simultan dan pengujian secara parsial. Pengujian ini dilakukan dengan tujuan mengetahui variabel penjelas yang signifikan mempengaruhi variabel respon.

1. Uji Simultan

   Uji simultan merupakan cara mengetahui apakah variabel penjelas dalam model secara simultan (bersama-sama) berpengaruh signifikan (nyata) atau tidak. Uji rasio likelihood dapat digunakan dalam uji simultan karena uji ini melihat apakah model tanpa variabel penjelas (hanya konstanta) signifikan lebih baik dari model dengan variabel penjelas (Hosmer dan Lemeshow, 2000). Berikut hipotesis dari uji rasio likelihood:

   Hipotesis:

   H_0 : $\beta_1 = \beta_2 = \ldots = \beta_p = 0$ dengan $j = 1,2,\ldots,p$

   H_1 : $minimal\ ada\ satu\ \beta_p \neq 0$ dengan $j = 1,2,\ldots,p$

   Statistik uji dalam pengujian ini menggunakan uji *F* seperti pada persamaan berikut*:*

   $$F_{hitung} = \frac{R^2/(k-1)}{(1-R^2)\ /\ (n-k)} \quad (2.29)$$

   dengan:

   R^2 : Kooefisien determinasi

   n : jumlah data

   k : jumlah variabel bebas

   Kriteria uji : tolak H_0 jika $F_{hit} > F_{tabel}$ berarti bahwa hubungan antara semua variabel bebas dan variabel terikat berpengaruh signifikan. (Ramdhanny, 2021)

2. Uji Parsial

   Uji parsial merupakan cara untuk mengetahui variabel-variabel penjelas mana (secara individu) yang memiliki peranan atau hubungan yang lebih kuat terhadap variabel respon. Berikut hipotesis dari uji ini:

   Hipotesis:

   H_0 : $\beta_j = 0$ dengan $j = 1,2,\ldots,p$

   H_1 : $\beta_j \neq 0$ dengan $j = 1,2,\ldots,p$

Statistik uji:

$$t_{hit} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \quad (2.30)$$

dengan:

$\hat{\beta}_j$: penaksir dari β_j

$se(\hat{\beta}_j)$: penaksir galat baku dari β_j

Kriteria uji: tolak H_0 jika nilai $|t_{hit}| > t_{\alpha/2;(n-(p+r)-1)}$ atau *p-value* < α. Sehingga dihasilkan kesimpulan variabel penjelas ke-n berpengaruh signifikan terhadap variabel respon (Putri,2018).

2.8 Angka Kematian Ibu (AKI)

Angka Kematian Ibu (maternal mortality rate) merupakan jumlah kematian ibu akibat dari proses kehamilan, persalinan, dan pasca persalinan yang dijadikan indikator derajat kesehatan perempuan. Angka Kematian Ibu (AKI) juga merupakan salah satu indikator utama keberhasilan program kesehatan ibu. Selain itu, indikator ini juga mampu menilai derajat kesehatan masyarakat, karena sensitifitasnya terhadap perbaikan pelayanan kesehatan, baik dari sisi aksesibilitas maupun kualitas (Kemenkes, 2020)

Menurut Kemenkes (2020), berbagai pelayanan kesehatan dasar yang dilakukan sebagai upaya percepatan penurunan AKI, diantaranya:

1. Pelayanan Kesehatan Ibu Hamil

   Pelayanan kesehatan ibu hamil terdiri dari kunjungan pertama (K1) dan kunjungan keempat (K4). Cakupan K1 adalah jumlah ibu hamil yang telah memperoleh pelayanan kesehatan pertama kali oleh tenaga kesehatan sesuai standar. Sedangkan cakupan K4 adalah jumlah ibu hamil yang telah mendapatkan pelayanan kesehatan sesuai standar, minimal empat kali kunjungan selama masa kehamilannya.

2. Pemberian Tablet Tambah Darah pada Ibu Hamil

   Anemia pada ibu hamil dapat meningkatkan risiko kelahiran prematur, kematian ibu dan anak, serta penyakit infeksi. Anemia defisiensi besi pada ibu dapat mempengaruhi Pertumbuhan dan perkembangan janin/bayi saat kehamilan maupun setelahnya. Hasil Riskesdas 2018 menyatakan bahwa di Indonesia sebesar 48,9% ibu hamil mengalami anemia. Sebanyak 84,6% anemia pada ibu hamil terjadi pada kelompok umur 15-24 tahun. Untuk mencegah anemia setiap ibu hamil diharapkan mendapatkan tablet tambah darah (TTD) minimal 90 tablet selama kehamilan.

3. Penanganan Komplikasi Kebidanan

   Komplikasi kebidanan adalah kesakitan pada ibu hamil, ibu bersalin, ibu nifas dan atau janin dalam kandungan, baik langsung maupun tidak langsung termasuk penyakit menular dan tidak menular yang dapat mengancam jiwa ibu dan atau janin, yang tidak disebabkan oleh trauma/kecelakaan. Adapun penanganan komplikasi itu sendiri adalah penanganan terhadap komplikasi/ kegawatdaruratan yang mendapat pelayanan kesehatan sampai selesai.

2.9 Persentase Penduduk Miskin dan Angka Partisipasi Sekolah (APS)

1. Persentase Penduduk Miskin

   Persentase penduduk miskin yang berada di bawah garis kemiskinan. Headcount Index secara sederhana mengukur proporsi yang dikategorikan miskin. Persentase penduduk miskin ini digunakan untuk mengetahui persentase penduduk yang dikategorikan miskin persentase penduduk miskin yang tinggi menunjukkan bahwa tingkat kemiskinan di suatu wilayah juga tinggi.

2. Angka Partisipasi Sekolah

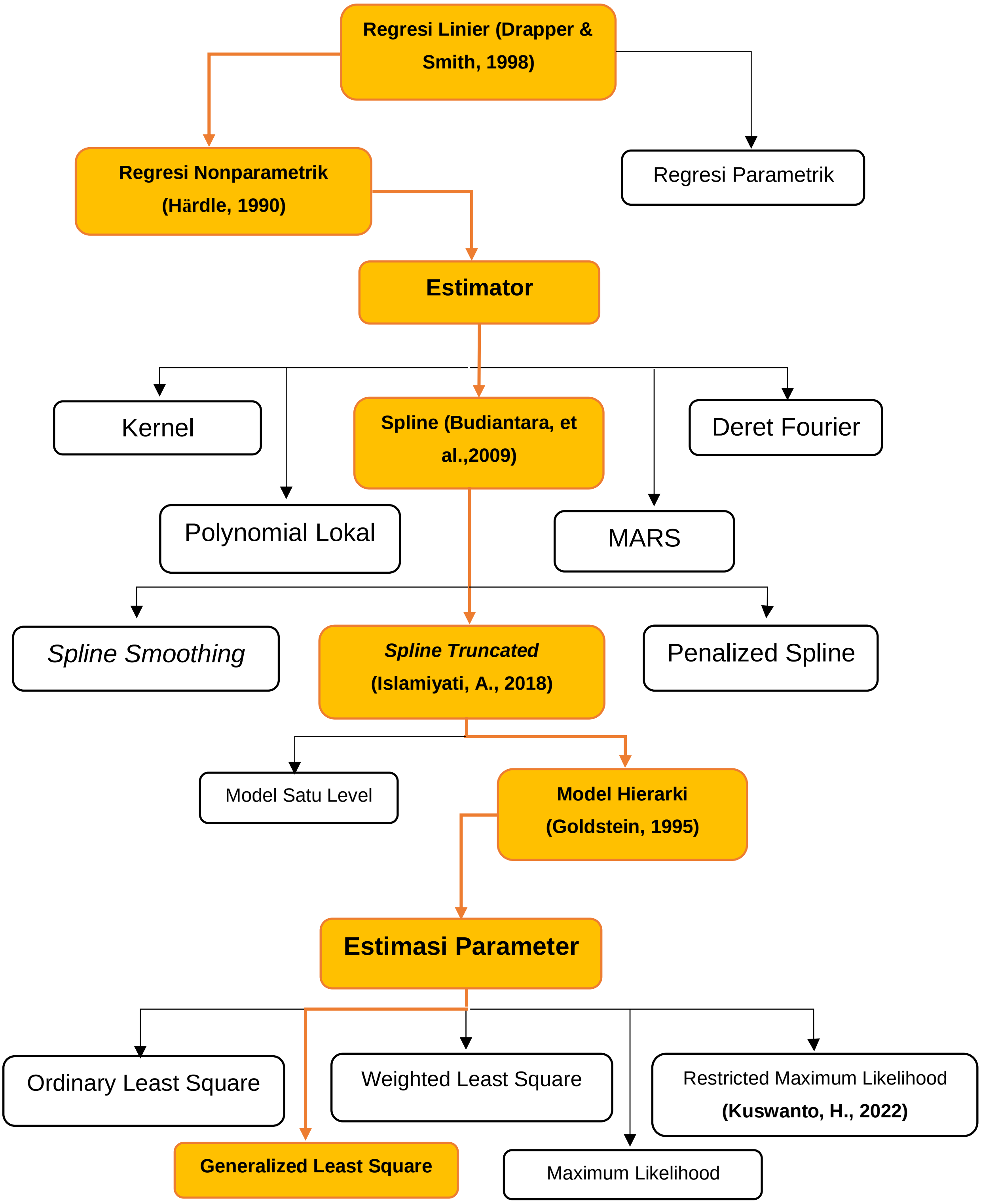
   Angka Partisipasi Sekolah (APS) adalah perbandingan antara jumlah murid kelompok usia sekolah tertentu yang bersekolah pada berbagai jenjang pendidikan dengan penduduk kelompok usia sekolah yang sesuai dan dinyatakan dalam persentase. Makin tinggi APS berarti makin banyak usia sekolah yang bersekolah di suatu daerah

Berikut ialah definsi operasional untuk variabel penelitian yang diuraikan pada table 2.2 berikut:

Tabel 2. 2 Definisi Operasional Peubah Penelitian yang digunakan

No	**Nama Variabel**	**Definsi**
(1)	**(2)**	**(3)**
1	Jumlah Kematian Ibu (Y)	Kematian Ibu adalah Semua Kematian Ibu yang disebabkan oleh kehamilan, persalinan, dan nifas atau pengelolaannya tetapi bukan karena kecelakaan atau incidental
2	Jumlah Ibu Hamil yang Mendapatkan Tablet Tambah Darah (X_1)	Data Jumlah Ibu Hamil yang memperoleh Tablet Tambah Darah (90 Tablet) sebagai upaya penceahan anemia akibat kekurangan zat besi dan asam folat
2	Jumlah Ibu Hamil yang Melaksanakan K4 (X_2)	Jumlah Ibu Hamil yang memperoleh pelayanan antenatal sesuai dengan standar paling sedikit empat kali sesuai jadwal yang dianjurkan di tiap semester
3	Jumlah Penanganan Komplikasi Kebidanan (X_3)	Jumlah kehamilan yang mengalami komplikasi behasil ditangani dengan baik dan tepat waktu
4	Penduduk Miskin (Z_1)	Penduduk Miskin adalah penduduk yang memiliki rata-rata pengeluaran konsumsi per kapita per bulan di bawah garis kemiskinan
5	Angka Partisipasi Sekolah (Z_2) (16-18 tahun)	Angka Partisipasi Sekolah menggambarkan ukuran daya serap system Pendidikan terhadap penduduk usia sekolah

2.10 **Kerangka Konseptual**



Gambar 2. 8 Kerangka Konseptual