

**PEMODELAN RATA-RATA LAMA SEKOLAH DI PROVINSI
NUSA TENGGARA TIMUR (NTT) MENGGUNAKAN
PENDEKATAN REGRESI NONPARAMETRIK
*SPLINE TRUNCATED***

Ludia Ni'matuzzahroh^{1*}, Andrea Tri Rian Dani¹

¹Departemen Statistika, Fakultas Sains dan Analitika Data,
Institut Teknologi Sepuluh Nopember Surabaya, 60111, Indonesia

Corresponding author: ludia.rahma@gmail.com

Abstrak. Misalkan diberikan data berpasangan x_i dan y_i , dimana $i = 1, 2, \dots, n$, kemudian hubungan antara variabel x_i dan y_i mengikuti model regresi nonparametrik sebagai berikut: $y_i = f(x_i) + \varepsilon_i$. Kurva regresi dari $f(x_i)$ dihampiri dengan pendekatan estimator *spline truncated*. Estimator *spline truncated* digunakan jika kurva regresi antara variabel respon terhadap variabel prediktor tidak membentuk suatu pola tertentu. Estimator *Spline Truncated* memiliki bentuk yang lebih fleksibel dalam menangani data yang perilakunya berubah-ubah pada sub-sub interval tertentu, hal ini dikarenakan pada estimator *Spline Truncated* terdapat titik knot. Tujuan dari penelitian ini adalah memodelkan variabel Rata-Rata Lama Sekolah di Provinsi Nusa Tenggara Timur (NTT) dengan menggunakan model regresi nonparametrik *Spline Truncated*. Variabel yang diduga mempengaruhi variabel respon adalah Persentase Penduduk Miskin, Angka Beban Ketergantungan, dan Persentase Rumah Tangga yang Menerima Program Indonesia Pintar. Metode yang digunakan dalam pemilihan titik knot adalah *Generalized Cross-Validation (GCV)*. Berdasarkan hasil analisis, model terbaik yang menghasilkan nilai GCV minimum adalah model regresi nonparametrik *Spline Truncated* dengan 3 titik knot, dimana nilai GCV sebesar 0,24. Berdasarkan hasil pengujian hipotesis secara simultan dan parsial, semua variabel prediktor yang digunakan didalam penelitian ini berpengaruh signifikan terhadap variabel respon yaitu rata-rata lama sekolah Kabupaten/Kota di Provinsi Nusa Tenggara Timur.

Kata Kunci: *Rata-Rata Lama Sekolah, Regresi Nonparametrik, Spline Truncated*

1 PENDAHULUAN

Rata-rata lama sekolah merupakan salah satu indikator yang dapat digunakan untuk mengevaluasi kinerja pemerintah dalam meningkatkan kualitas pendidikan serta kualitas sumber daya manusia di suatu wilayah. Semakin tinggi rata-rata lama sekolah di suatu wilayah, maka semakin tinggi juga jenjang pendidikan yang telah dicapai penduduk di wilayah tersebut. Berdasarkan data publikasi BPS Tahun 2020, rata-rata lama sekolah di Provinsi Nusa Tenggara Timur (NTT) sebesar 8,09 tahun. Hal tersebut dapat diartikan bahwa pendidikan yang telah dicapai oleh penduduk di Provinsi NTT sebagian besar hanya sampai tingkat Sekolah Menengah Pertama (SMP). Dalam proses mengevaluasi rata-rata lama sekolah, juga perlu melihat apa saja faktor-faktor yang diduga dapat mempengaruhinya, dengan harapan apabila terdapat faktor yang dirasa kurang maksimal dapat diatasi, sehingga kualitas pendidikan serta kualitas sumber daya manusia di wilayah tersebut dapat meningkat.

Faktor-faktor yang diduga dapat mempengaruhi rata-rata lama sekolah dan seberapa besar pengaruhnya, dapat diketahui dengan menggunakan analisis regresi. Pada analisis regresi tersebut akan dapat diketahui bentuk pola hubungan antara variabel rata-rata lama sekolah (variabel respon) dengan beberapa variabel yang diduga mempengaruhinya (variabel prediktor). Pola hubungan antara variabel respon dan prediktor dapat didekati dengan tiga cara yaitu regresi parametrik, regresi nonparametrik, dan regresi semiparametrik. Pendekatan regresi parametrik dapat digunakan ketika bentuk kurva regresi diketahui, seperti linier, kuadrat, kubik, dan lain-lain. Pendekatan semiparametrik perlu digunakan pada saat pola hubungan antar variabelnya sebagian mengikuti pola tertentu dan sebagian lainnya tidak diketahui bentuk polanya. Sedangkan pada pendekatan regresi nonparametrik lebih disarankan untuk digunakan apabila bentuk kurva regresi tidak diketahui [1]. Pendekatan regresi nonparametrik tidak tergantung pada asumsi bentuk kurva regresi tertentu. Beberapa estimator yang banyak dikembangkan oleh para peneliti dalam pemodelan menggunakan pendekatan regresi nonparametrik, yaitu diantaranya *Spline*, Kernel, Deret Fourier, Polinomial Lokal, *Wavelets*, dan lain sebagainya [2].

Pada beberapa estimator yang digunakan dalam regresi nonparametrik tersebut, estimator *Spline* lebih banyak diminati karena model regresi nonparametrik dengan estimator *Spline* mempunyai interpretasi statistik dan interpretasi visual sangat khusus dan sangat baik [2]. Pengembangan estimator *Spline* dalam regresi nonparametrik dengan menggunakan basis fungsi keluarga *Spline Truncated*, yang memberikan perhitungan matematik lebih mudah dan sederhana [3]. Selain itu, estimator *Spline Truncated* mampu menangani data yang memiliki pola yang berubah-ubah pada sub-sub interval tertentu [1]. Beberapa penelitian yang menggunakan regresi nonparametrik *Spline Truncated* antara lain [4]-[7].

Dalam suatu analisis statistika, perlu menggunakan statistik deskriptif dan statistik inferensia yang berupa pengujian hipotesis secara simultan maupun parsial terhadap parameter model. Hal tersebut bertujuan untuk mengetahui signifikansi parameter dalam model. Beberapa penelitian yang melakukan pengujian hipotesis dalam regresi nonparametrik *Spline Truncated*, diantaranya [8]-[10].

Berdasarkan uraian di atas, akan dimodelkan rata-rata lama sekolah Kabupaten/Kota di Provinsi Nusa Tenggara Timur (NTT) Tahun 2020 dengan menggunakan pendekatan regresi nonparametrik *Spline Truncated*, kemudian

dilanjutkan analisis statistika inferensia yaitu pengujian hipotesis secara simultan dan secara parsial terhadap parameter-parameter yang dihasilkan dalam model.

2 TINJAUAN PUSTAKA

2.1 Analisis Regresi

Analisis regresi merupakan suatu metode yang digunakan untuk mengetahui hubungan antara variabel respon dan variabel prediktor [11]. Diberikan data berpasangan (x_i, y_i) yang secara umum dapat dimodelkan dengan menggunakan model regresi:

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

dimana y_i merupakan variabel respon, x_i merupakan variabel prediktor, $f(x_i)$ merupakan kurva regresi yang akan didekati, dan ε_i merupakan *error random* yang diasumsikan identik, independen, dan berdistribusi normal dengan *mean* bernilai nol dan varians σ^2 atau $N(0, \sigma^2)$ [2].

2.2 Regresi Nonparametrik *Spline Truncated*

Regresi nonparametrik *Spline Truncated* merupakan salah satu pendekatan analisis regresi yang dapat digunakan pada saat pola hubungan antara variabel respon dan variabel prediktor tidak diketahui bentuk kurva regresinya. Regresi nonparametrik *Spline Truncated* mampu menangani data yang mempunyai perilaku yang berubah-ubah pada sub-sub interval tertentu dan cenderung mencari sendiri estimasi data kemanapun pola data bergerak. Hal tersebut dikarenakan dalam fungsi *truncated* terdapat adanya titik-titik knot, yang merupakan titik perpaduan bersama dimana terdapat perubahan perilaku data atau fungsi yang polanya berubah [12]. Secara umum, model regresi nonparametrik *Spline Truncated* multivariabel dapat dituliskan seperti pada Persamaan (2).

$$y_i = f(x_{qi}) + \varepsilon_i; q = 1, 2, \dots, Q; i = 1, 2, \dots, n \quad (2)$$

dengan,

$$f(x_{qi}) = \sum_{q=1}^Q \left(\sum_{m=0}^M \beta_{qm} x_{qi}^m + \sum_{r=1}^R \beta_{q(m+r)} (x_{qi} - K_{rq})_+^m \right)$$

$$(x_{qi} - K_{rq})_+^m = \begin{cases} (x_{qi} - K_{rq})^m & ; x_{qi} \geq K_{rq} \\ 0 & ; x_{qi} < K_{rq} \end{cases}$$

dimana:

- y_i : variabel respon
- $f(x_{qi})$: fungsi regresi nonparametrik *Spline Truncated*
- x_{qi} : variabel prediktor
- ε_i : *error random* yang diasumsikan identik, independen, dan berdistribusi normal dengan *mean* bernilai nol dan variansi σ^2

$$\sum_{m=0}^M \beta_{qm} x_{qi}^m \quad : \text{komponen polinomial}$$

$$\sum_{r=1}^R \beta_{q(m+r)} (x_{qi} - K_{rq})_+^M \quad : \text{komponen truncated}$$

2.3 Generalized Cross Validation (GCV)

Generalized Cross Validation (GCV) merupakan salah satu metode yang dikembangkan oleh peneliti untuk menentukan titik knot optimal dalam regresi nonparametrik *Spline Truncated* [13]. Rumus GCV untuk model regresi nonparametrik *Spline Truncated* ditampilkan pada Persamaan (3).

$$GCV(K) = \frac{MSE(K)}{(n^{-1}tr(I - A(K)))^2} \quad (3)$$

dengan,

$$MSE(K) = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad \text{dan} \quad A(K) = X(X^T X)^{-1} X^T$$

3 DATA

Pada penelitian ini menggunakan data sekunder yang dipublikasikan oleh Badan Pusat Statistik (BPS), dengan unit observasi Kabupaten/Kota di Provinsi Nusa Tenggara Timur (NTT) Tahun 2020 yaitu sebanyak 22 kabupaten/kota.

Variabel-variabel yang digunakan dalam penelitian ini telah disajikan pada Tabel 1, yang terdiri dari satu variabel respon (Y) dan tiga variabel prediktor (X).

Tabel 1: Variabel Penelitian

Variabel	Simbol Variabel	Deskripsi
Respon	Y	Rata-Rata Lama Sekolah
	X ₁	Persentase Penduduk Miskin
Prediktor	X ₂	Angka Beban Ketergantungan
	X ₃	Persentase Rumah Tangga yang Menerima Program Indonesia Pintar

Definisi operasional dari masing-masing variabel adalah sebagai berikut:

- 1) Rata-Rata Lama Sekolah
 Rata-rata lama sekolah didefinisikan sebagai jumlah tahun yang digunakan oleh penduduk dalam menjalani pendidikan formal, sehingga rata-rata lama sekolah dapat digunakan untuk mengetahui kualitas pendidikan masyarakat dalam suatu wilayah.
- 2) Persentase Penduduk Miskin
 Persentase penduduk miskin merupakan persentase penduduk yang memiliki rata-rata pengeluaran perkapita perbulan berada di bawah Garis Kemiskinan (GK), dimana Garis Kemiskinan (GK) adalah penjumlahan dari Garis Kemiskinan Makanan (GKM) dan Garis Kemiskinan Non Makanan (GKNM).
- 3) Angka Beban Ketergantungan

Perbandingan antara jumlah penduduk umur 0-14 tahun, ditambah dengan jumlah penduduk 65 tahun ke atas (keduanya disebut dengan bukan angkatan kerja/usia non produktif) dibandingkan dengan jumlah penduduk usia 15-64 tahun (angkatan kerja/ usia produktif).

- 4) Persentase Rumah Tangga yang Menerima Program Indonesia Pintar
Jumlah rumah tangga penerima manfaat Program Indonesia Pintar (PIP) menurut kepemilikan kartu keluarga sejahtera dalam setahun terakhir dibagi jumlah rumah tangga dikali seratus.

4 HASIL DAN PEMBAHASAN

4.1 Statistika Deskriptif

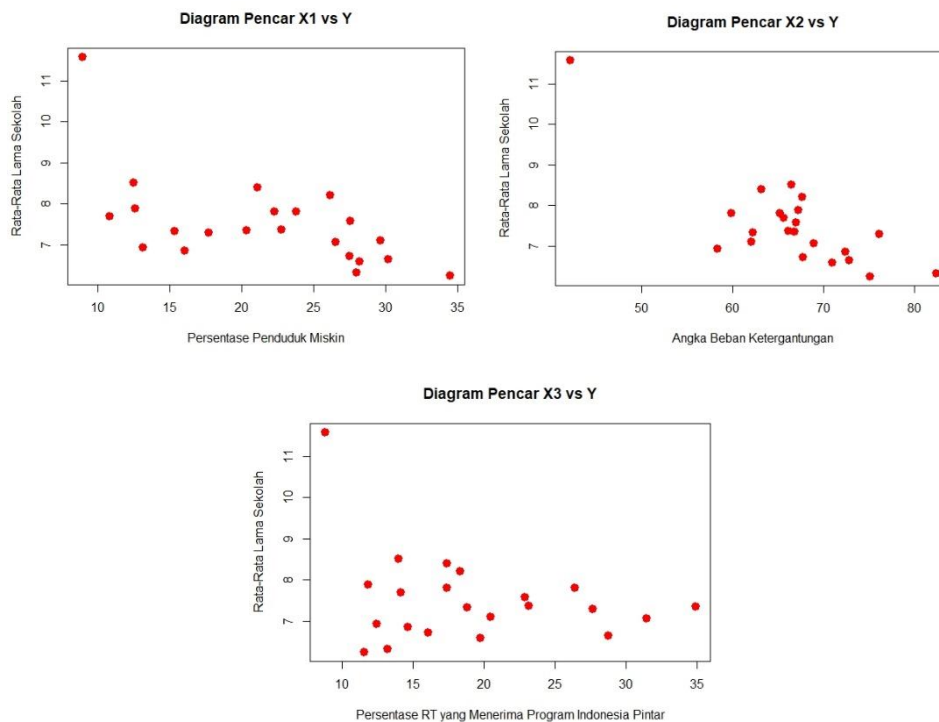
Statistika deskriptif digunakan untuk memberikan gambaran awal mengenai masing-masing variabel yang digunakan, ditampilkan pada Tabel 2.

Tabel 2: Statistika Deskriptif

No.	Variabel	Nilai Minimum	Nilai Maksimum	Rata-Rata
1.	X_1	8,96	34,49	21,62
2.	X_2	42,12	82,36	66,60
3.	X_3	8,78	34,91	19,24
4.	Y	6,25	11,58	7,524

4.2 Scatter Plot

Identifikasi awal yang perlu dilakukan sebelum melakukan pemodelan regresi nonparametrik *spline truncated* adalah dengan melihat pola hubungan antara variabel respon terhadap masing-masing variabel prediktor yang ditampilkan pada Gambar 1.



Gambar 1: *Scatter Plot*

Berdasarkan Gambar 1, terlihat jika pola hubungan antara masing-masing variabel prediktor terhadap variabel respon tidak membentuk suatu pola tertentu (bukan linear, kuadratik, kubik, dan lain sebagainya), sedemikian sehingga dalam proses pemodelannya akan didekati dengan pendekatan regresi nonparametrik *spline truncated*.

4.3 Pemodelan Regresi Nonparametrik *Spline Truncated*

Regresi nonparametrik *spline truncated* terkenal cukup baik dalam memodelkan pola hubungan antara variabel prediktor terhadap variabel respon yang tidak memiliki pola tertentu. Ciri khas dari model regresi nonparametrik *spline truncated* yaitu adanya titik knot. Model regresi nonparametrik *spline truncated* yang terbaik diperoleh dari lokasi titik knot dan jumlah titik knot yang optimal. Pada penelitian ini, jumlah titik knot yang digunakan adalah 1 sampai dengan 3 titik knot, dimana jumlah titik knotnya adalah sama untuk setiap variabel prediktor.

1) Pemilihan Titik Knot Optimal

Metode yang digunakan untuk menentukan lokasi dan jumlah titik knot yang optimal pada penelitian ini adalah *Generalized Cross-Validation (GCV)*. Lokasi titik knot dan jumlah yang optimal dapat dilihat dari nilai GCV yang minimum dari beberapa kombinasi titik knot yang dicobakan. Akan ditampilkan 3 posisi titik knot yang diduga merupakan titik knot yang optimal.

Tabel 3: Nilai GCV dengan 1 Titik Knot

No.	Lokasi Titik Knot			GCV
	X ₁	X ₂	X ₃	
1.	10,02	43,79	9,87	0,37
2.	17,47	55,53	17,49	0,41
3.	11,08	45,47	10,95	0,42

Tabel 4: Nilai GCV dengan 2 Titik Knot

No.	Lokasi Titik Knot			GCV
	X ₁	X ₂	X ₃	
1.	10,17	44,04	10,02	0,33
	22,33	63,19	22,47	
2.	10,17	44,04	10,02	0,35
	23,55	65,11	23,71	
3.	10,17	44,03	10,02	0,36
	19,90	59,36	19,97	

Tabel 5: Nilai GCV dengan 3 Titik Knot

No.	Lokasi Titik Knot			GCV
	X ₁	X ₂	X ₃	
1.	11,08	45,47	10,95	0,24
	13,21	48,83	13,13	
	18,53	57,21	18,58	
2.	11,08	45,47	10,96	0,26
	13,21	48,82	13,13	
	17,47	55,53	17,49	
3.	12,15	47,15	12,04	0,28
	13,21	48,82	13,13	
	18,53	57,21	18,57	

Setelah dilakukan pemodelan regresi nonparametrik *spline truncated*, maka akan ditampilkan ringkasan hasil seperti pada Tabel 6.

Tabel 6: Ringkasan Hasil Pemilihan Titik Knot

Banyak Titik Knot	Lokasi Titik Knot			GCV	R ²
	X ₁	X ₂	X ₃		
1 Titik Knot	10,02	43,79	9,87	0,37	80,66%
2 Titik Knot	10,17	44,04	10,02	0,33	88,43%
	22,33	63,19	22,47		
3 Titik Knot	11,08	45,47	10,95	0,24	93,71%
	13,21	48,83	13,13		
	18,53	57,21	18,58		

Berdasarkan Tabel 6, dapat diketahui jika model regresi nonparametrik *spline truncated* yang menghasilkan nilai GCV minimum dengan nilai Koefisien Determinasi (R^2) terbesar adalah model regresi nonparametrik *spline truncated* dengan 3 titik knot.

2) Model Regresi Nonparametrik yang Terbaik

Ditampilkan hasil estimasi parameter model regresi nonparametrik *spline truncated* 3 titik knot pada Tabel 7.

Tabel 7: Hasil Estimasi Parameter

Parameter	Estimasi Parameter	Parameter	Estimasi Parameter	Parameter	Estimasi Parameter
$\hat{\beta}_0$	-0,11				
$\hat{\beta}_{11}$	20,46	$\hat{\beta}_{21}$	-3,94	$\hat{\beta}_{31}$	-0,61
$\hat{\beta}_{12}$	-23,53	$\hat{\beta}_{22}$	0,81	$\hat{\beta}_{32}$	0,53
$\hat{\beta}_{13}$	3,39	$\hat{\beta}_{23}$	1,14	$\hat{\beta}_{33}$	0,15
$\hat{\beta}_{14}$	-0,41	$\hat{\beta}_{24}$	1,96	$\hat{\beta}_{34}$	-0,11

Sehingga dapat dituliskan modelnya sebagai berikut:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_{11}x_1 + \hat{\beta}_{12}(x_1 - 11,08)_+ + \hat{\beta}_{13}(x_1 - 13,21)_+ + \hat{\beta}_{14}(x_1 - 18,53)_+ + \hat{\beta}_{21}x_2 + \hat{\beta}_{22}(x_2 - 45,47)_+ + \hat{\beta}_{23}(x_2 - 48,83)_+ + \hat{\beta}_{24}(x_2 - 57,21)_+ + \hat{\beta}_{31}x_3 + \hat{\beta}_{32}(x_3 - 10,95)_+ + \hat{\beta}_{33}(x_3 - 13,13)_+ + \hat{\beta}_{34}(x_3 - 18,58)_+$$

3) Pengujian Hipotesis

Setelah diperoleh model regresi nonparametrik *spline truncated* yang terbaik, yaitu model regresi nonparametrik *spline truncated* dengan 3 titik knot, maka langkah selanjutnya adalah melakukan pengujian hipotesis. Pengujian hipotesis terbagi menjadi 2, yaitu secara simultan dan parsial.

a. Pengujian Hipotesis Simultan

Pengujian hipotesis simultan bertujuan untuk melihat apakah secara simultan (serentak), variabel prediktor yang digunakan berpengaruh signifikan atau tidak terhadap variabel respon. Dengan menggunakan statistik uji F , maka dapat dituliskan hipotesisnya sebagai berikut:

Hipotesis

H_0 : $\beta_1 = \beta_2 = \dots = \beta_{12} = 0$

H_1 : paling tidak ada satu $\beta_j \neq 0$, dimana $j = 1, 2, \dots, 12$

Dengan taraf signifikansi sebesar 5% atau 0,05, dan daerah penolakan H_0 yaitu Tolak H_0 jika $F_{hit} > F_{(\alpha,k,n-k-1)}$, diperoleh tabel ANOVA sebagai berikut:

Tabel 8: Tabel ANOVA

Sumber	db	SS	MS	Fhit
Regresi	12	23,96	1,99	27,33
Error	9	1,61	0,07	
Total	21	25,58		

Berdasarkan Tabel 8, didapatkan nilai F_{hit} yaitu sebesar 27,33 dimana diketahui nilai ini lebih besar dari nilai $F_{(0,05,12,9)}$ yaitu 3,07, sehingga diputuskan Tolak H_0 . Kita peroleh kesimpulan jika paling tidak ada satu $\beta_j \neq 0$, atau minimal terdapat satu parameter yang signifikan.

b. Pengujian Hipotesis Parsial

Pengujian hipotesis secara parsial dilakukan dengan tujuan untuk melihat parameter-parameter mana yang berpengaruh signifikan didalam model regresi. Dengan menggunakan statistik uji t , maka dapat dituliskan hipotesisnya sebagai berikut:

Hipotesis

$H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0$, dimana $j = 1, 2, \dots, 12$

Dengan taraf signifikansi sebesar 5% atau 0,05, dan daerah penolakan H_0 yaitu Tolak H_0 jika $|t_{hit}| > t_{(\alpha,n-k-1)}$.

Tabel 9: Uji t

No.	Variabel	Parameter	Thit	Keputusan	Kesimpulan
1.	Konstanta	$\hat{\beta}_0$	-4,58	Tolak H_0	Signifikan
2.	Variabel X ₁	$\hat{\beta}_{11}$	5,58	Tolak H_0	Signifikan
3.		$\hat{\beta}_{12}$	-5,69	Tolak H_0	Signifikan
4.		$\hat{\beta}_{13}$	6,22	Tolak H_0	Signifikan
5.		$\hat{\beta}_{14}$	-4,42	Tolak H_0	Signifikan
6.		$\hat{\beta}_{21}$	-5,20	Tolak H_0	Signifikan
7.	Variabel X ₂	$\hat{\beta}_{22}$	3,55	Tolak H_0	Signifikan
8.		$\hat{\beta}_{23}$	4,67	Tolak H_0	Signifikan

9.		$\hat{\beta}_{24}$	6,62	Tolak H_0	Signifikan
10.		$\hat{\beta}_{31}$	-4,48	Tolak H_0	Signifikan
11.	Variabel X_3	$\hat{\beta}_{32}$	3,48	Tolak H_0	Signifikan
12.		$\hat{\beta}_{33}$	0,69	Gagal Tolak H_0	Tidak Signifikan
13.		$\hat{\beta}_{34}$	-1,65	Gagal Tolak H_0	Tidak Signifikan

Berdasarkan Tabel 9, diperoleh hasil jika hampir semua parameter yang terdapat didalam model adalah signifikan. Dari 12 parameter yang terdapat didalam model, hanya 2 parameter saja yang tidak signifikan. Sehingga dapat kita simpulkan jika semua variabel prediktor yang digunakan di dalam penelitian ini berpengaruh terhadap variabel respon yaitu rata-rata lama sekolah Kabupaten/Kota di Provinsi Nusa Tenggara Timur.

4) Pemeriksaan Asumsi

a. Asumsi Non-Heteroskedastisitas

Pengujian asumsi ini dapat dilakukan dengan menggunakan uji Glejser dengan hipotesis sebagai berikut:

Hipotesis

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$$

$$H_1 : \text{paling tidak ada satu } \sigma_i^2 \neq \sigma^2, \text{ dimana } i = 1, 2, \dots, n$$

Dengan taraf signifikansi sebesar 5% dan daerah penolakan H_0 yaitu Tolak H_0 jika $F_{hit} > F_{(\alpha, k, n-k-1)}$, diperoleh hasil seperti pada Tabel 10.

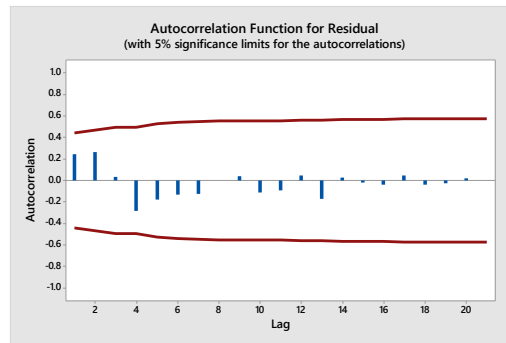
Tabel 10: Tabel ANOVA Uji Glejser

Sumber	db	SS	MS	Fhit
Regresi	12	0,39	0,03	0,78
Error	9	0,37	0,04	
Total	21	0,77		

Berdasarkan Tabel 10, didapatkan nilai F_{hit} yaitu sebesar 0,78 dimana diketahui nilai ini lebih kecil dari nilai $F_{(0,05,12,9)}$ yaitu 3,07, sehingga diputuskan Gagal Tolak H_0 . Di peroleh kesimpulan jika $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$, yang artinya varians dari residual adalah sama (homogen) atau dengan kata lain tidak terjadi masalah heteroskedastisitas.

b. Asumsi Non-Otokorelasi

Pengujian asumsi non-otokorelasi dapat dilakukan dengan membuat grafik *Autocorrelation Function* (ACF). Di indikasikan apabila tidak terdapat lag yang keluar dari batas signifikansi, maka tidak terjadi masalah otokorelasi. Ditampilkan grafik ACF dari residual pada Gambar 2.



Gambar 2: Grafik ACF

Berdasarkan Gambar 2, terlihat jika tidak terdapat *lag* yang keluar dari batas signifikansi sedemikian sehingga dapat kita simpulkan jika tidak terjadi masalah otokorelasi.

c. Asumsi Residual Berdistribusi Normal

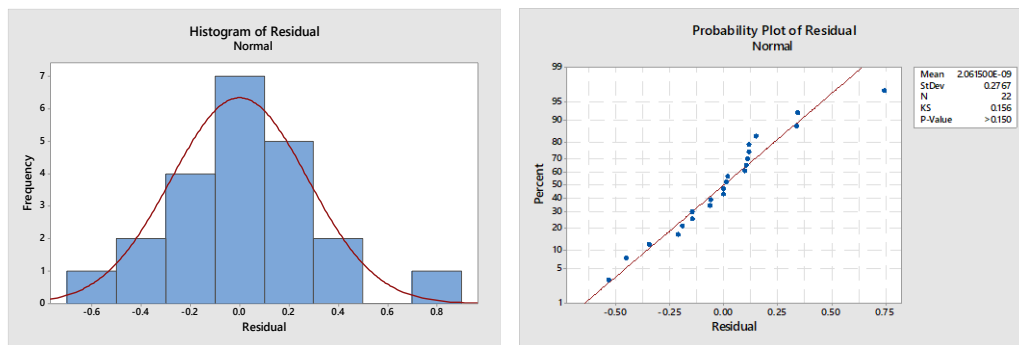
Untuk melakukan pemeriksaan apakah *residual* dari hasil pemodelan mengikuti distribusi Normal atau tidak, dapat dilakukan dengan membuat histogram dan ataupun menggunakan statistik uji. Adapun statistik uji yang digunakan pada penelitian ini adalah Kolmogorov-Smirnov dengan hipotesis sebagai berikut:

Hipotesis

H_0 : Residual berdistribusi Normal

H_1 : Residual tidak berdistribusi Normal

Dengan taraf signifikansi sebesar 5%, diperoleh:



(a) Histogram

(b) Normal QQ-Plot

Gambar 3: Uji Normalitas

Berdasarkan Gambar 3(a), terlihat jika histogram yang terbentuk menyerupai bentuk lonceng yang simetris, sehingga diindikasikan jika residual berdistribusi Normal. Hal ini diyakinkan dengan pengujian hipotesis pada Gambar 3(b) menggunakan uji Kolmogorov-Smirnov, dimana diperoleh nilai *P-Value* yang lebih besar dari 0,05 sehingga diputuskan Gagal Tolak H_0 .

5 KESIMPULAN

Berdasarkan hasil analisis pada Poin 4, maka diperoleh kesimpulan sebagai berikut:

- 3) Diperoleh model regresi nonparametrik *spline truncated* dengan 3 titik knot yang terbaik dalam memodelkan permasalahan rata-rata lama sekolah Kabupaten/Kota di Provinsi Nusa Tenggara Timur Tahun 2020, yang ditunjukkan dengan nilai GCV yang paling minimum yaitu sebesar 0,24.
- 4) Model regresi nonparametrik *spline truncated* dengan 3 titik knot yang merupakan model terbaik dituliskan sebagai berikut:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_{12} (x_1 - 11,08)_+ + \hat{\beta}_{13} (x_1 - 13,21)_+ + \hat{\beta}_{14} (x_1 - 18,53)_+ + \hat{\beta}_{21} x_2 + \hat{\beta}_{22} (x_2 - 45,47)_+ + \hat{\beta}_{23} (x_2 - 48,83)_+ + \hat{\beta}_{24} (x_2 - 57,21)_+ + \hat{\beta}_{31} x_3 + \hat{\beta}_{32} (x_3 - 10,95)_+ + \hat{\beta}_{33} (x_3 - 13,13)_+ + \hat{\beta}_{34} (x_3 - 18,58)_+$$

- 5) Berdasarkan hasil pengujian hipotesis secara simultan, diketahui jika paling tidak ada satu $\beta_j \neq 0$, atau minimal terdapat satu parameter yang signifikan. Kemudian dari hasil pengujian hipotesis secara parsial, dapat disimpulkan jika semua variabel prediktor berpengaruh signifikan terhadap variabel respon yaitu rata-rata lama sekolah Kabupaten/Kota di Provinsi Nusa Tenggara Timur.

Saran yang dapat diajukan untuk penelitian selanjutnya diantaranya:

- 3) Menggunakan jumlah titik knot kombinasi untuk setiap variabel prediktor.
- 4) Mengembangkan model estimator campuran untuk setiap variabel prediktor yang disesuaikan berdasarkan karakteristik pola hubungannya terhadap variabel respon.

DAFTAR PUSTAKA

- [1] Budiantara, I.N. (2009). *Spline Dalam Regresi Nonparametrik dan Semiparametrik: Pidato Pengukuhan untuk Jabatan Guru Besar Institut Teknologi Sepuluh Nopember*. Surabaya: ITS Press.
- [2] Eubank, R.L. (1999). *Nonparametric Regression and Spline Smoothing*. New York: Marcel Dekker.
- [3] Budiantara, I.N. (2005). *Model Keluarga Spline Polinomial Truncated Dalam Regresi Semiparametrik*. Semarang: Universitas Diponegoro.
- [4] Merdekawati, I.P. (2013). Pemodelan Regresi Spline Truncated Multivariabel pada Faktor-Faktor yang Mempengaruhi Kemiskinan di Kabupaten/Kota Provinsi Jawa Tengah. *Jurnal Sains dan Seni Pomits*. 2. D19-D24.
- [5] Saputro, D.R.S., Demu, K.R., dan Widyaningsih, P. (2018). Nonparametric Truncated Spline Regression Model on Data of Human Development Index (HDI) in Indonesia. *IOP Conf. Series: Journal of Physics*. 1188.
- [6] Rahim, F., Budiantara, I.N., dan Permatasari, E.O. (2019). Spline Truncated Nonparametric Regression Modeling for Maternal Mortality Rate in East Java. *INFERENCE*. 2.

- [7] Dani, A.T.R., Ni'matuzzahroh, L., Ratnasari, V., dan Budiantara, I.N. (2021). Pemodelan Regresi Nonparametrik Spline Truncated pada Data Longitudinal. *INFERENSI*. 4(1). 47-55.
- [8] Tupen, S.N. & Budiantara, I.N. (2011). Uji Hipotesis dalam regresi Nonparametrik Spline. *Prosiding Seminar Nasional Statistika ke-8*, Universitas Diponegoro, Semarang. 184-199.
- [9] Dani, A.T.R., Adrianingsih, N.Y., dan Ainurrochmah, A. (2020). Pengujian Hipotesis Simultan Model Regresi Nonparametrik Spline Truncated dalam Pemodelan Kasus Ekonomi. *JAMBURA Journal of Probability and Statistics*. 1(2). 98-106.
- [10] Dani, A.T.R. & Ni'matuzzahroh, L. (2021). Pemodelan Persentase Penduduk Miskin Kabupaten/Kota di Provinsi Jawa Barat dengan Pendekatan Regresi Nonparametrik Spline Truncated. *J Statistika*. 14(1). 24-29.
- [11] Draper, N. & Smith, H. (1992). *Analisis Regresi Terapan*. Jakarta: PT Gramedia Pustaka.
- [12] Budiantara, I.N. (2019). *Regresi Nonparametrik Spline Truncated*. Surabaya: ITS Press.
- [13] Wahba, G. (1990). *Spline Models for Observational Data*. Pennsylvania: SIAM.