

# Pemodelan Regresi Nonparametrik Spline *Truncated* untuk Mengetahui Faktor-Faktor yang Mempengaruhi Penggunaan Kontrasepsi

**Ratna Rachmasari\*, Yayat Karyana**

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Islam Bandung, Indonesia.

\*ratnarachmasari8@gmail.com

**Abstract.** Population problem are the main problem experienced by all developing countries including Indonesia. One of the efforts made by the government to controlling the population growth rate is Keluarga Berencana (KB). One indicator that can be used is the Contraceptive Prevalence Rate (CPR), which is a number that shows the number of couples of childbearing age who are using contraceptives. The method used in this study is a truncated spline nonparametric regression analysis to model the proportion of CPR in Indonesia in 2015. Spline regression is a polynomial regression which is divided into several segments that form specific knot points and continuous. The best estimate of the nonparametric regression model for the truncated spline is obtained from selecting the optimal point knots by calculating the generalized cross validation (GCV). The results showed that the best truncated spline nonparametric regression model was to use a combination knots 1,2,3,3 with a GCV value of 79.20663 and all independent variables were not significant.

**Keywords:** CPR, GCV, Nonparametric Spline.

**Abstrak.** Masalah kependudukan merupakan masalah utama yang dialami oleh semua negara berkembang termasuk Indonesia. Salah satu upaya yang dilakukan pemerintah dalam mengatasi laju pertumbuhan penduduk adalah keluarga berencana (KB). Untuk mengevaluasi keberhasilan dari kebijakan program KB di Indonesia, salah satu indikator yang dapat digunakan adalah *Contraceptive Prevalence Rate (CPR)* yaitu angka yang menunjukkan jumlah pasangan usia subur yang menggunakan alat kontrasepsi. Metode yang digunakan dalam penelitian ini adalah analisis regresi nonparametrik spline *truncated* untuk memodelkan persentase CPR di Indonesia tahun 2015. Regresi spline *truncated* merupakan suatu regresi polinomial dimana terbagi menjadi beberapa segmen yang dibatasi titik-titik knot tertentu dan kontinu. Estimasi terbaik model regresi nonparametrik untuk truncated spline diperoleh dari pemilihan titik knot yang optimal dengan menghitung *generalized cross validation (GCV)*. Hasil penelitian menunjukkan bahwa model regresi nonparametrik spline truncated terbaik adalah menggunakan kombinasi knot 1,2,3,3 dengan nilai GCV 79.20663 dan semua variabel independen tidak signifikan.

**Kata Kunci:** CPR, GCV, Nonparametrik Spline.

## 1. Pendahuluan

Metode Statistika yang paling sering digunakan oleh para peneliti dalam berbagai bidang penelitian yaitu analisis regresi. Analisis ini digunakan untuk mengetahui pola hubungan antara satu atau lebih variabel dalam bentuk fungsional. Terdapat tiga jenis regresi yang dikembangkan oleh para peneliti yaitu regresi parametrik, regresi semiparametrik dan regresi nonparametrik. Model regresi nonparametrik sangat baik digunakan untuk menganalisis data yang polanya tidak diketahui, dimana data diharapkan mencari sendiri bentuk estimasi kurva regresinya tanpa dipengaruhi subyektifitas peneliti<sup>[2]</sup>.

Salah satu model regresi nonparametrik yaitu regresi spline *truncated*. Kurva regresi spline dapat mengatasi permasalahan pola pada data yang menunjukkan naik atau turun yang tajam dengan bantuan titik-titik knot<sup>[2]</sup> dimana banyaknya titik knot dan lokasi titik knot sangat mempengaruhi bentuk kurva regresi spline. Salah satu masalah yang dapat dianalisis menggunakan metode regresi nonparametric spline *truncated* berdasarkan penelitian sebelumnya adalah masalah kependudukan.

Masalah kependudukan merupakan masalah utama yang dialami oleh seluruh negara berkembang termasuk Indonesia. Menurut data Worldometers<sup>[5]</sup>, pada tahun 2019 Indonesia memiliki jumlah penduduk sebanyak 269 juta jiwa atau 3,49% dari total seluruh penduduk di dunia. Keluarga Berencana (KB) merupakan salah satu upaya pemerintah dalam mengendalikan jumlah penduduk Indonesia. Untuk mengevaluasi keberhasilan dari kebijakan program KB di Indonesia, salah satu ukuran yang dapat digunakan adalah *Contraceptive Prevalence Rate* (CPR) yaitu angka yang menunjukkan jumlah PUS yang sedang menggunakan alat kontrasepsi.

Oleh karena itu, keberhasilan program KB merupakan hal yang menarik untuk dianalisis dimana salah satu alat ukur yang dapat digunakan adalah rata-rata penggunaan kontrasepsi (CPR). Maka pada penelitian ini akan dilakukan analisis lebih lanjut untuk mengetahui keberhasilan program KB melalui indikator CPR berdasarkan faktor-faktor yang diduga berpengaruh dengan menggunakan regresi nonparametrik spline *truncated*.

Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah dalam penelitian ini sebagai berikut:

1. “Bagaimana karakteristik dari persentase CPR di Indonesia?”
2. “Bagaimana cara memodelkan faktor-faktor yang berpengaruh terhadap penggunaan kontrasepsi yang dilihat berdasarkan persentase CPR dengan menggunakan regresi nonparametric spline *truncated*?”

## 2. Landasan Teori

Regresi spline merupakan suatu regresi polynomial dimana terbagi menjadi beberapa segmen yang dibatasi titik-titik knot tertentu dan kontinu sehingga bersifat fleksibel dibandingkan polynomial biasa. Titik knot pada regresi spline merupakan titik perpaduan yang terjadi karena perubahan pola perilaku dari suatu fungsi pada selang yang berbeda. Kurva regresi spline dapat menyesuaikan diri secara efektif terhadap perubahan perilaku data, sehingga didapatkan hasil yang mendekati kebenaran<sup>[1]</sup>. Kurva regresi spline memiliki kemampuan yang sangat baik untuk menangani data yang perilakunya berubah-ubah pada sub-sub interval tertentu<sup>[3]</sup>. Model regresi spline mempunyai interpretasi statistik dan interpretasi visual yang sangat khusus dan sangat baik<sup>[4]</sup>.

Fungsi  $f(x_i)$  pada regresi nonparametrik merupakan fungsi regresi yang tidak diketahui bentuknya dan sifat pemulusnya, maka untuk mengestimasi  $f(x_i)$  tersebut dapat digunakan model regresi spline. Fungsi spline pada suatu fungsi  $f(x_i)$  dengan orde  $p$  dapat dinyatakan sebagai berikut:

$$f(x_i) = \beta_0 + \beta_1 x_i^1 + \dots + \beta_p x_i^p + \sum_{l=1}^r \beta_{(p+l)} (x_i - k_l)_+^p$$

$$f(x_i) = \sum_{j=0}^p \beta_j x_i^j + \sum_{l=1}^r \beta_{(p+l)} (x_i - k_l)_+^p$$

dengan  $r$  menyatakan banyaknya titik knot dan  $(x_i - k_l)_+^p$  menyatakan fungsi potongan

(truncated) yang dapat dijabarkan sebagai berikut:

$$(x_i - k_l)_+^p = \begin{cases} (x_i - k_l)^p, & x_i \geq k_l \\ 0, & x_i < k_l \end{cases}$$

Bentuk matematis dari fungsi spline pada persamaan, dapat dinyatakan bahwa spline adalah potongan-potongan polinom yang berbeda digabungkan bersama titik *knot*  $k_1, k_2, k_3, \dots, k_r$  untuk menjamin sifat kontinuitasnya.

Dalam regresi nonparametric dengan pendekatan spline *truncated*, hal penting yang berperan dalam mendapatkan estimator spline *truncated* adalah pemilihan titik *knot* yang optimal. Salah satu metode yang sering digunakan dalam memilih titik *knot* optimal adalah *Generalized Cross Validation* (GCV). Metode GCV juga memiliki kelebihan tidak memerlukan pengetahuan terhadap variansi populasi  $\sigma^2$  serta metode GCV invarians terhadap transformasi. Fungsi GCV untuk pemilihan titik *knot* optimal dapat ditunjukkan dalam persamaan dibawah.

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

Dengan  $MSE(k) = n^{-1} \sum_{i=1}^n (y_i - \hat{y})^2$ , n adalah jumlah data, I adalah matriks identitas, k adalah titik *knot*  $(k_1, k_2, k_3, \dots, k_n)$ , dan  $A(k) = x(x'x)^{-1}x'$ . Dalam menentukan model terbaik pada penelitian ini menggunakan nilai GCV, maka kriteria yang harus dipenuhi yaitu nilai GCV yang paling minimum.

**3. Hasil Penelitian dan Pembahasan**

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistika (BPS) berupa data survei penduduk antar sensus (SUPAS) 2015 dengan unit observasi yang digunakan adalah seluruh provinsi yang ada di Indonesia, yaitu sebanyak 34 provinsi.

**Karakteristik Persentase CPR**

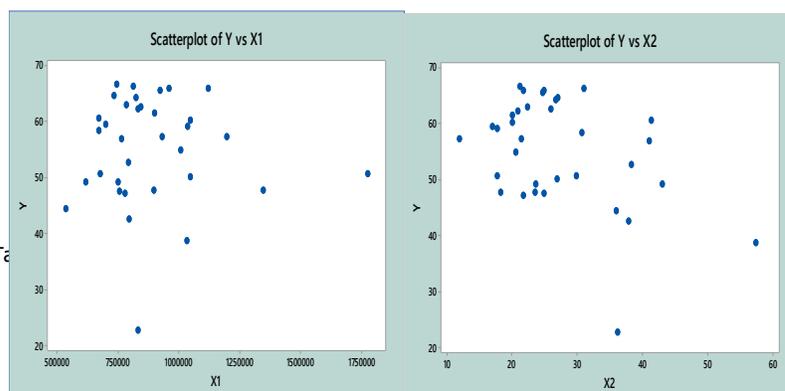
Karakteristik dari persentase CPR dan faktor-faktor yang diduga mempengaruhinya di Indonesia disajikan pada Tabel 1.

**Tabel 1.** Karakteristik Persentase CPR

Variabel	Y	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>
<b>Min</b>	22,67	533800	11,87	68,50	12,55
<b>Max</b>	66,51	1773400	57,33	93,01	50,44
<b>Mean</b>	54,96	883674	27,08	88,29	27,60

**Pola Hubungan Persentase CPR dengan Faktor yang diduga mempengaruhinya**

Pola hubungan yang terbentuk antara persentase CPR (Y) yang merupakan variabel dependen dengan variabel independen jumlah pendapatan (X<sub>1</sub>), IMR (X<sub>2</sub>), persentase penduduk usia 5 tahun keatas dengan pendidikan SLTP kebawah (X<sub>3</sub>) dan proporsi penduduk usia 5 tahun keatas yang mengakses internet (X<sub>4</sub>) divisualisasikan pada Gambar 1.



**Gambar 1.** *Scatterplot* Persentase CPR dengan Variabel Independen  $X_1, X_2, X_3$  dan  $X_4$

Gambar 1. Menunjukkan pola hubungan yang terbentuk antara persentase CPR dengan keempat variabel independen tidak mengikuti suatu pola tertentu. Karena tidak diketahui pola datanya maka estimasi model dilakukan dengan analisis regresi nonparametric.

**Pemilihan Titik Knot Optimum**

Pemilihan titik knot optimum dilakukan dengan cara mencari nilai GCV terendah. Untuk mendapatkan nilai GCV minimum dilakukan beberapa kemungkinan yaitu menggunakan satu titik knot, dua titik knot, tiga titik knot dan kombinasi titik knot. Nilai GCV minimum untuk setiap titik knotnya disajikan pada Tabel 2.

**Tabel 2 .** Nilai GCV Minimum

	<b>GCV</b>
1 Titik Knot	83.74395
2 Titik Knot	83.74395
3 Titik Knot	87.81671
Kombinasi Titik Knot	79.20663

Berdasarkan Tabel 2. Dapat diketahui bahwa pemodelan yang menghasilkan nilai GCV paling minimum merupakan pemodelan regresi nonparametric spline truncated dengan menggunakan kombinasi knot 1,2,3,3. Oleh karena itu dapat disimpulkan bahwa model terbaik yang akan dipilih adalah model regresi nonparametric spline truncated dengan menggunakan kombinasi knot 1,2,3,3. Dengan estimasi model regresi nonparametric sebagai berikut.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_{12}(x_1 - K_1) + \hat{\beta}_2 x_2 + \hat{\beta}_{21}(x_2 - K_2) + \hat{\beta}_{22}(x_2 - K_3) + \hat{\beta}_3 x_3 + \hat{\beta}_{31}(x_3 - K_4) + \hat{\beta}_{32}(x_3 - K_5) + \hat{\beta}_{33}(x_3 - K_6) + \hat{\beta}_4 x_4 + \hat{\beta}_{41}(x_4 - K_7) + \hat{\beta}_{42}(x_4 - K_8) + \hat{\beta}_{43}(x_4 - K_9)$$

### Pemodelan Regresi Nonparametrik Spline *Truncated* dengan Titik Knot Optimum

Berdasarkan kriteria pemilihan model terbaik diketahui bahwa nilai GCV paling minimum dihasilkan oleh model regresi nonparametric spline *truncated* dengan kombinasi titik knot 1,2,3,3. Dengan menggunakan metode *Maximum Likelihood Estimator* (MLE) diperoleh model regresi terbaik yang dihasilkan adalah sebagai berikut.

$$\hat{y} = 0,0000565 + 0,0000935x_1 - 0,000102(x_1 - 609693,9) - 0,215814x_2 - 0,213778(x_2 - 22,07504) - 0,160249(x_2 - 29,49689) + 0,0744153x_3 + 0,0470786(x_3 - 82,00845) + 0,0312852(x_3 - 86,01004) + 0,009351(x_3 - 90,01163) + 0,0418818x_4 + 0,00935(x_4 - 33,42924) - 0,00486(x_4 - 39,61534) - 0,00267(x_4 - 45,80144)$$

### Interpretasi Model Regresi Nonparametrik Spline *Truncated*

Berdasarkan penjelasan sebelumnya model dari regresi nonparametric spline *truncated* dapat diinterpretasikan sebagai berikut

1. Apabila variabel  $X_2, X_3$  dan  $X_4$  dianggap konstan, maka pengaruh jumlah pendapatan ( $X_1$ ) terhadap persentase CPR di Indonesia adalah sebagai berikut.

$$\hat{y} = 0,0000565 + 0,0000935x_1 - 0,000102(x_1 - 609693,9)$$

$$\hat{y} = \begin{cases} 0,0000565 + 0,0000935x_1 & , \quad x_1 < 609693,9 \\ 62,1888 - 0,0000085x_1 & , \quad x_1 \geq 609693,9 \end{cases}$$

Berdasarkan model tersebut dapat dijabarkan ketika jumlah pendapatan kurang dari 609693,9 rupiah maka jika jumlah pendapatan daerah naik sebesar satu satuan persentase CPR akan naik sebesar 0,0000935 persen. Ketika jumlah pendapatan lebih dari sama dengan 609693,9 rupiah maka jika jumlah pendapatan daerah naik sebesar satu satuan persentase CPR akan turun sebesar 0,0000085 persen.

2. Apabila variabel  $X_1, X_3$  dan  $X_4$  dianggap konstan, maka pengaruh IMR ( $X_2$ ) terhadap persentase CPR di Indonesia adalah sebagai berikut.

$$\hat{y} = 0,0000565 - 0,215814x_2 - 0,213778(x_2 - 22,07504) - 0,160249(x_2 - 29,49689)$$

$$\hat{y} = \begin{cases} 0,0000565 - 0,215814x_2 & , \quad x_2 < 22,07504 \\ 4,719214401 - 0,429592x_2 & , \quad 22,07504 \leq x_2 < 29,49689 \\ 9,446061527 - 0,589841x_2 & , \quad x_2 \geq 29,49689 \end{cases}$$

Berdasarkan model tersebut dapat dijabarkan ketika nilai IMR kurang dari 22,07504 maka jika nilai IMR naik sebesar satu satuan persentase CPR akan turun sebesar 0,215814 persen. Ketika nilai IMR lebih dari sama dengan 22,07504 dan kurang dari 29,49689 maka jika nilai IMR naik sebesar satu satuan persentase CPR akan turun sebesar 0,429592 persen. Ketika nilai IMR lebih dari sama dengan 29,49689 maka jika nilai IMR naik sebesar satu satuan persentase CPR akan turun sebesar 0,589841 persen.

3. Apabila variabel  $X_1, X_2$  dan  $X_4$  dianggap konstan, maka pengaruh persentase penduduk usia 5 tahun keatas dengan tingkat pendidikan SLTP kebawah ( $X_3$ ) terhadap persentase CPR di Indonesia adalah sebagai berikut.

$$\hat{y} = 0,0000565 + 0,0744153x_3 + 0,0470786(x_3 - 82,00845) + 0,0312852(x_3 - 86,01004) + 0,009351(x_3 - 90,01163)$$

$$\hat{y} = \begin{cases} 0,0000565 + 0,0744153x_3 & , \quad x_3 < 82,00845 \\ -1,318542114 + 0,1214939x_3 & , \quad 82,00845 \leq x_3 < 86,01004 \\ -4,009383418 + 0,1527791x_3 & , \quad 86,01004 \leq x_3 < 90,01163 \\ -4,851082170 + 0,1621301x_3 & , \quad x_3 \geq 90,01163 \end{cases}$$

Berdasarkan model tersebut dapat dijabarkan ketika persentase penduduk usia 5 tahun keatas dengan tingkat pendidikan SLTP kebawah kurang dari 82,00845 persen maka jika persentase penduduk usia 5 tahun keatas dengan tingkat pendidikan SLTP kebawah naik sebesar satu satuan persentase CPR akan naik sebesar 0,0744153 persen. Ketika persentase penduduk usia 5 tahun keatas dengan tingkat pendidikan SLTP kebawah lebih

dari sama dengan 82,00845 persen dan kurang dari 86,01004 persen maka jika persentase penduduk usia 5 tahun keatas dengan tingkat pendidikan SLTP kebawah naik sebesar satu satuan persentase CPR akan naik sebesar 0,1214939 persen. Ketika persentase penduduk usia 5 tahun keatas dengan tingkat pendidikan SLTP kebawah lebih dari sama dengan 86,01004 persen dan kurang dari 90,01163 persen maka jika persentase penduduk usia 5 tahun keatas dengan tingkat pendidikan SLTP kebawah naik sebesar satu satuan persentase CPR akan naik sebesar 0,1527791 persen. Ketika persentase penduduk usia 5 tahun keatas dengan tingkat pendidikan SLTP kebawah lebih dari sama dengan 90,01163 maka jika persentase penduduk usia 5 tahun keatas dengan tingkat pendidikan SLTP kebawah naik sebesar satu satuan persentase CPR akan naik sebesar 0,1621301 persen.

4. Apabila variabel  $X_1, X_2$  dan  $X_3$  dianggap konstan, maka pengaruh proporsi penduduk usia 5 tahun keatas yang mengakses internet ( $X_4$ ) terhadap persentase CPR di Indonesia adalah sebagai berikut.

$$\hat{y} = 0,0000565 + 0,0418818x_4 + 0,00935(x_4 - 33,42924) - 0,00486(x_4 - 39,61534) - 0,00267(x_4 - 45,80144)$$

$$\hat{y} = \begin{cases} 0,0000565 + 0,0418818x_4, & x_4 < 33,42924 \\ -0,312506894 + 0,051231800x_4, & 33,42924 \leq x_4 < 39,61534 \\ -0,119976342 + 0,0463718x_4, & 39,61534 \leq x_4 < 45,80144 \\ 0,002313503 + 0,0437018x_4, & x_4 \geq 45,80144 \end{cases}$$

Berdasarkan model tersebut dapat dijabarkan ketika proporsi penduduk usia 5 tahun keatas yang mengakses internet kurang dari 33,42924 maka jika ketika proporsi penduduk usia 5 tahun keatas yang mengakses internet naik sebesar satu satuan persentase CPR akan naik sebesar 0,0418818 satuan. Ketika ketika proporsi penduduk usia 5 tahun keatas yang mengakses internet lebih dari sama dengan 33,42924 dan kurang dari 39,61534 maka ketika proporsi penduduk usia 5 tahun keatas yang mengakses internet naik sebesar satu satuan persentase CPR akan naik sebesar 0,051231800 satuan. Ketika ketika proporsi penduduk usia 5 tahun keatas yang mengakses internet lebih dari sama dengan 39,61534 dan kurang dari 45,80144 maka jika ketika proporsi penduduk usia 5 tahun keatas yang mengakses internet naik sebesar satu satuan persentase CPR akan naik sebesar 0,0463718 satuan. Ketika ketika proporsi penduduk usia 5 tahun keatas yang mengakses internet lebih dari sama dengan 45,80144 maka jika ketika proporsi penduduk usia 5 tahun keatas yang mengakses internet naik sebesar satu satuan persentase CPR akan naik sebesar 0,0437018 satuan.

#### 4. Kesimpulan

Berdasarkan pembahasan dalam penelitian ini, peneliti menyimpulkan beberapa hasil penelitian sebagai berikut:

1. Rata-rata persentase Contraceptive Prevalence Rate (CPR) di Indonesia yaitu sebesar 54,96%. Nilai tersebut menunjukkan bahwa secara rata-rata persentase banyaknya pasangan usia subur (PUS) yang sedang menggunakan alat kontrasepsi adalah sebesar 54,96%. Dengan presentase CPR terendah yaitu sebesar 22,67% dan persentase CPR tertinggi yaitu sebesar 66,51%.
2. Model regresi nonparametrik spline truncated terbaik untuk pemodelan persentase Contraceptive Prevalence Rate (CPR) di Indonesia yaitu dengan menggunakan kombinasi knot 1,2,3,3 dengan nilai GCV sebesar 79,20663.

Model regresi nonparametrik spline *truncated* yang diperoleh adalah sebagai berikut.

$$\hat{y} = 0,0000565 + 0,0000935x_1 - 0,000102(x_1 - 609693,9) - 0,215814x_2 - 0,213778(x_2 - 22,07504) - 0,160249(x_2 - 29,49689) + 0,0744153x_3 + 0,0470786(x_3 - 82,00845) + 0,0312852(x_3 - 86,01004) + 0,009351(x_3 - 90,01163) + 0,0418818x_4 + 0,00935(x_4 - 33,42924) - 0,00486(x_4 - 39,61534) - 0,00267(x_4 - 45,80144)$$

## 5. Saran

Berdasarkan analisis dan pembahasan disampaikan beberapa saran sebagai berikut.

1. Untuk penelitian selanjutnya sebaiknya untuk menambahkan faktor-faktor lainnya yang diduga mempengaruhi persentase CPR dan juga dapat dilakukan untuk penambahan jumlah titik knot. Dimana pada penelitian ini titik knot dibatasi hingga tiga titik knot, untuk selanjutnya dapat ditambahkan menjadi empat atau lebih titik knot.
2. Untuk pemerintah disarankan untuk lebih memperhatikan secara rata mengenai program KB di seluruh provinsi di Indonesia, karena berdasarkan penelitian dapat dilihat perbedaan yang cukup signifikan antara provinsi dengan persentase terendah dan tertinggi.

## Daftar Pustaka

- [1] Eubank, R.L. (1988). *Spline Smoothing and Nonparametric Regression*. New York: Marcel Dekker, Inc.
- [2] Eubank, R.L. (1999). *Nonparametric Regression and Spline Smoothing Second Edition*. New York: Marcel Dekker, Inc.
- [3] Cox, D. D., & O'Sullivan, F. (1996). Penalized Type Estimator for Generalized Nonparametric Regression. *Journal of Multivariate Analysis*. Vol.56, No.10, hal.185-206.
- [4] Budiantara I.N. (2009). Spline dalam Regresi Nonparametrik dan Semiparametrik, Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang. Pidato Pengukuhan Guru Besar, Institut Teknologi Sepuluh Nopember, Surabaya.
- [5] *Indonesia Population*. (2019). Diambil dari Worldometers: <https://www.worldometers.info/world-population/indonesia-population/>