

**ESTIMASI SELANG KEPERCAYAAN NILAI UJIAN NASIONAL BERBASIS
KOMPETENSI BERDASARKAN MODEL REGRESI SEMIPARAMETRIK
MULTIRESPON *TRUNCATED SPLINE***

Lilik Hidayati¹, Nur Chamidah², I Nyoman Budiantara³

^{1,2}Departement of Mathematics, Airlangga University, Indonesia

³Department of Statistics, Sepuluh Nopember Institute of Technology, Indonesia

e-mail: nur-c@fst.unair.ac.id

DOI: 10.14710/me dstat.13.1.92-103

Article Info:

Received: 7 March 2020

Accepted: 25 June 2020

Available Online: 26 June 2020

Keywords:

Confidence Interval

Estimation, UNBK Scores,

Multi-Response

Semiparametric, Truncated

Spline.

Abstract: Confidence interval estimation is important in statistical inference for the parameters of the regression model, but the theory of confidence interval estimation for multi-response semiparametric regression model parameters based on the truncated spline estimator has not been examined. In this study, we estimate the confidence interval of the multi-response semiparametric regression model based on the truncated spline estimator by using pivotal quantity method with the central limit theorem approach. This confidence interval theory is applied to data of competency-based national exam (UNBK) scores in West Nusa Tenggara Province where its UNBK in the lowest position among other provinces in Indonesia. The method used for estimating parameters is weighted least square. The best model is determined based on the Generalized Cross Validation (GCV) minimum value. Based on the estimated 95% confidence interval of parameters of the multi-response truncated spline semiparametric regression model, the results showed that the insignificant factors affecting the UNBK scores were gender and parental education duration while the report card of scores and USBK scores had a positive effect on the UNBK scores but only the UNBK scores of mathematics that report card of scores factor has a negative effect on it.

1. PENDAHULUAN

Estimasi selang kepercayaan merupakan salah satu bagian dari inferensi statistik yang sangat penting karena dengan estimasi selang kepercayaan tersebut dapat diketahui apakah nilai suatu parameter signifikan atau tidak terutama parameter dalam suatu model regresi. Dalam model regresi yang melibatkan lebih dari satu variabel prediktor ada tiga pendekatan yang dapat digunakan yaitu model regresi parametrik, regresi nonparametrik dan regresi semiparametrik. Apabila dalam analisis regresi fungsi regresi pola datanya

diasumsikan mengikuti pola tertentu (antara lain linier, kuadratik, kubik dan eksponensial) berdasarkan *scatter plot*nya maka digunakan pendekatan model regresi parametrik sedangkan regresi nonparametrik tidak mengasumsikan bentuk pola tertentu dari fungsi regresinya, sehingga sangat fleksibel dan obyektif (Budiantara, 2009). Realitanya ada pola hubungan antara variabel respon dan variabel prediktor yang memiliki pola tertentu sedangkan dengan variabel prediktor lainnya tidak memiliki pola tertentu sehingga model regresi semiparametrik yang lebih sesuai untuk memodelkan pola data seperti itu.

Beberapa peneliti yang telah meneliti terkait model regresi semiparametrik unirespon antara lain: Yang *et al.* (2016) menggunakan estimator penalized spline, Prahutama dan Suparti (2017) menggunakan pendekatan polinomial lokal, Utami *et al.* (2018) menggunakan estimator kernel, dan Ramadan *et al.* (2019) menggunakan estimator truncated spline sedangkan untuk model regresi semiparametrik multirespon antara lain: Wibowo, *et al.* (2012) dan Chamidah dan Lestari (2019) menggunakan estimator polinomial lokal, Chamidah dan Rifada (2016) menggunakan estimator linier lokal, Chamidah, *et al.* (2018) menggunakan estimator *least square spline* dan Hidayati *et al.* (2019) menggunakan estimator *truncated spline*. Namun penelitian-penelitian tersebut masih dalam batas mengestimasi titik belum membahas estimasi selang kepercayaan. Hidayati *et al.* (2020) sudah membahas terkait estimasi selang kepercayaan parameter model regresi semiparametrik multirespon berdasarkan estimator truncated spline dan melakukan studi simulasi untuk menunjukkan untuk kasus homoskedastik dan heteroskedastik estimator yang menggunakan *weighted least square* memiliki akurasi yang lebih baik daripada *unweighted least square*.

Data Puspendik Balitbang Kemendikbud (2017) menyatakan bahwa rata-rata nilai Ujian Nasional Berbasis Komputer (UNBK) Sekolah Menengah Kejuruan Negeri (SMKN) di Provinsi Nusa Tenggara Barat (NTB) secara nasional dari tahun ke tahun menempati posisi paling bawah dibandingkan provinsi lainnya di Indonesia yaitu tahun 2015 berada di posisi 29, tahun 2016 berada di posisi 34, dan pada tahun 2017 posisi 34 dari 34 provinsi yang ada di Indonesia. Berdasarkan fakta tersebut, pada tulisan ini diteliti faktor-faktor yang mempengaruhi nilai UNBK Sekolah Menengah Kejuruan Negeri (SMKN) di Provinsi NTB dengan pendekatan estimasi selang kepercayaan parameter model regresi semiparametrik multirespon. Variabel respon yang terlibat dalam model adalah nilai UNBK Matematika, Bahasa Indonesia, Bahasa Inggris, dan kompetensi keahlian sedangkan variabel prediktornya Jenis Kelamin dan Nilai Rapor (sebagai komponen parametrik) sedangkan nilai Ujian Sekolah Berbasis Kompetensi (USBK) dan lama pendidikan orang tua (sebagai komponen nonparametrik). Estimasi selang kepercayaan parameter dari masing-masing variabel prediktor yang diduga mempengaruhi nilai UNBK, digunakan untuk mengetahui apakah variabel-variabel prediktor tersebut berpengaruh secara signifikan atau tidak. Sehingga bermanfaat bagi pemerintah provinsi NTB untuk menentukan upaya-upaya perbaikan dan program pembinaan di satuan pendidikan.

2. TINJAUAN PUSTAKA

2.1 Model Regresi Semiparametrik Multirespon

Diberikan data berpasangan $(y_i^{(k)}, x_i, t_i)$, $i = 1, 2, \dots, n$; $k = 1, 2, \dots, r$ yang mengikuti model regresi semiparametrik (Chamidah, *et al.* 2018):

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\alpha} + \mathbf{f}(t_i) + \boldsymbol{\varepsilon}_i \quad (1)$$

dengan variabel respon $\mathbf{y}_i = (y_{1i}, y_{2i}, \dots, y_{ri})^T$, $\boldsymbol{\alpha} = (\alpha_0^{(1)} \quad \alpha_1^{(1)} \quad \alpha_0^{(2)} \quad \alpha_1^{(2)} \quad \dots \quad \alpha_0^{(r)} \quad \alpha_1^{(r)})^T$, $\mathbf{X}_i = (1, x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(r)})$ adalah variabel prediktor sebagai komponen parametrik dan $\mathbf{f}(t_i) = (f_1(t_i) \quad f_2(t_i) \quad \dots \quad f_r(t_i))^T$ adalah fungsi regresi yang diestimasi dengan pendekatan regresi nonparametrik, $\boldsymbol{\varepsilon}_i = (\varepsilon_{1i} \quad \varepsilon_{2i} \quad \dots \quad \varepsilon_{ri})^T$ adalah *error random* dengan mean $\mathbf{0}$ dan variansi $\boldsymbol{\Sigma}_i$.

Model dalam persamaan (1) yang telah dibahas dalam Chamidah, *et al.* (2018) hanya melibatkan satu variabel prediktor pada komponen nonparametriknya, sedangkan dalam Hidayati *et al.* (2020) telah membahas estimasi model regresi semiparametrik multirespon yang melibatkan lebih dari satu variabel prediktor (multiprediktor) pada komponen nonparametriknya menggunakan estimator *truncated spline*.

Diberikan pasangan observasi $(x_1^{(r)}, \dots, x_p^{(r)}, t_1^{(r)}, \dots, t_q^{(r)}, y^{(r)})$ dengan $r=1,2,3,\dots,R$ yang mengikuti model semiparametrik

$$y_i^{(r)} = \beta_0^{(r)} + \sum_{j=1}^p \beta_j x_{ji}^{(r)} + \sum_{t=1}^q \left[\gamma_t^{(r)} t_{ii}^{(r)} + \sum_{d=1}^{S_t} \gamma_{t+d}^{(r)} (t_{ii}^{(r)} - K_{td}^{(r)})_+ \right] + \varepsilon_i^{(r)}, \quad i=1,2,\dots,n \quad (2)$$

Untuk n observasi persamaan (2) dapat dinyatakan dalam bentuk matriks berikut ini

$$\underline{y} = C\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, W) \quad (3)$$

$$C = \begin{pmatrix} x & t \end{pmatrix} \quad \boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{pmatrix}$$

$$\underline{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(R)} \end{bmatrix} \quad C = \begin{bmatrix} C^{(1)} & 0 & \dots & 0 \\ 0 & C^{(2)} & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & C^{(R)} \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}^{(1)} \\ \boldsymbol{\varepsilon}^{(2)} \\ \vdots \\ \boldsymbol{\varepsilon}^{(R)} \end{bmatrix}$$

$$\boldsymbol{\theta}^{(r)} = \begin{bmatrix} \beta_0^{(r)} & \beta_1^{(r)} & \beta_2^{(r)} & \dots & \beta_p^{(r)} & \gamma_1^{(r)} & \gamma_2^{(r)} & \dots & \gamma_q^{(r)} \end{bmatrix}$$

Vektor *error* $\boldsymbol{\varepsilon}^{(1)}, \boldsymbol{\varepsilon}^{(2)}, \dots, \boldsymbol{\varepsilon}^{(R)}$ antar respon saling berkorelasi sehingga untuk mengestimasi parameter $\boldsymbol{\theta}$ digunakan metode *Weighted Least Square* (WLS) dengan pembobot adalah invers dari matriks variansi-kovariansi *error* (W^{-1}). Struktur matriks variansi-kovariansi *error* ditunjukkan sebagai berikut:

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T) = E \begin{bmatrix} E(\boldsymbol{\varepsilon}^{(1)}\boldsymbol{\varepsilon}^{(1)T}) & E(\boldsymbol{\varepsilon}^{(1)}\boldsymbol{\varepsilon}^{(2)T}) & \dots & E(\boldsymbol{\varepsilon}^{(1)}\boldsymbol{\varepsilon}^{(R)T}) \\ E(\boldsymbol{\varepsilon}^{(2)}\boldsymbol{\varepsilon}^{(1)T}) & E(\boldsymbol{\varepsilon}^{(2)}\boldsymbol{\varepsilon}^{(2)T}) & \dots & E(\boldsymbol{\varepsilon}^{(2)}\boldsymbol{\varepsilon}^{(R)T}) \\ \vdots & \vdots & \ddots & \vdots \\ E(\boldsymbol{\varepsilon}^{(R)}\boldsymbol{\varepsilon}^{(1)T}) & E(\boldsymbol{\varepsilon}^{(R)}\boldsymbol{\varepsilon}^{(2)T}) & \dots & E(\boldsymbol{\varepsilon}^{(R)}\boldsymbol{\varepsilon}^{(R)T}) \end{bmatrix} = \begin{bmatrix} W_{11} & W_{12} & \dots & W_{1R} \\ W_{21} & W_{22} & \dots & W_{2R} \\ \vdots & \vdots & \ddots & \vdots \\ W_{R1} & W_{R2} & \dots & W_{RR} \end{bmatrix} = W$$

Berdasarkan metode WLS diperoleh estimasi untuk parameter $\boldsymbol{\theta}$

$$\hat{\boldsymbol{\theta}} = (C^T W^{-1} C)^{-1} C^T W^{-1} \underline{y} \quad (4)$$

Untuk mendapatkan model terbaik dapat dilakukan dengan pemilihan titik knot optimal berdasarkan metode GCV sebagai berikut:

$$GCV = \frac{\text{Mean Square Error}(MSE)}{\left((nR)^{-1} \text{trace}(I - H)\right)^2} \quad (5)$$

dengan $MSE = \frac{1}{nR - R^*} (\underline{y} - \hat{\underline{y}})^T (\underline{y} - \hat{\underline{y}})$, $R^* = \left(R + Rp + qR + R \sum_{\ell=1}^q S_{\ell} \right)$ dan

$$H = (C^T W^{-1} C)^{-1} C^T W^{-1}$$

2.2 Selang Kepercayaan Parameter Model Regresi Semiparametrik Multirespon

Selang kepercayaan $(1 - \alpha)100\%$ untuk parameter model regresi semiparametrik multirespon berdasarkan *truncated spline* θ_v , $v = 1, 2, \dots, R^*$, $R^* = \left(R + Rp + qR + R \sum_{\ell=1}^q S_{\ell} \right)$

dengan σ^2 tidak diketahui dapat digunakan *pivotal quantity* sebagai berikut (Hidayati *et al* 2020):

$$T_v(x_1, x_2, \dots, x_p, t_1, t_2, \dots, t_q, \underline{y}) = \frac{\hat{\theta}_v - \theta_v}{\sqrt{\text{Var}(\hat{\theta}_v)}} \quad (6)$$

$$= \frac{\hat{\theta}_v - \theta_v}{\sqrt{(C^T W^{-1} C)^{-1}_{vv}}}$$

Dengan pendekatan *Central Limit Teorem* (CLT) untuk $n \rightarrow \infty$ diperoleh

$$Z = \frac{\hat{\theta}_v - \theta_v}{\sqrt{(C^T W^{-1} C)^{-1}_{vv}}}, n \rightarrow \infty, \text{ dengan } Z \sim N(0,1). \quad (7)$$

Jika tingkat signifikansi sebesar α maka selang kepercayaan $(1 - \alpha)100\%$ untuk parameter $\hat{\theta}$ pada persamaan (3) adalah sebagai berikut:

$$P\left(\hat{\theta}_v - Z_{\left(\frac{\alpha}{2}\right)} \sqrt{(C^T W^{-1} C)^{-1}_{vv}} \leq \theta_v \leq \hat{\theta}_v + Z_{\left(\frac{\alpha}{2}\right)} \sqrt{(C^T W^{-1} C)^{-1}_{vv}}\right) = 1 - \alpha$$

$(C^T W^{-1} C)^{-1}_{vv}$ menunjukkan elemen dari baris ke v kolom ke- v dari matriks $(C^T W^{-1} C)^{-1}$

2.3. Nilai Ujian Nasional Berbasis Komputer (UNBK)

Ujian Nasional Berbasis Komputer (UNBK) pertama kali diselenggarakan tahun 2014, dan terbatas pada beberapa sekolah yang sudah memenuhi kriteria. UNBK diselenggarakan pada sekolah yang sudah siap baik dari segi infrastruktur, SDM maupun peserta. Penilaian pencapaian standar kompetensi lulusan pada mata pelajaran tertentu secara nasional, pemerintah melakukan evaluasi yang merupakan hasil belajar dan dilakukan secara nasional. Hasil ujian nasional digunakan sebagai pertimbangan untuk pemetaan mutu program pada satuan pendidikan, sebagai dasar seleksi masuk jenjang pendidikan berikutnya dan pembinaan dan pemberian bantuan kepada satuan pendidikan dalam upayanya untuk meningkatkan mutu pendidikan.

Menurut Mahmud (1989), faktor-faktor yang mempengaruhi prestasi belajar yaitu faktor internal dan faktor eksternal siswa. Prestasi belajar yang dimaksud dalam penelitian ini adalah nilai Ujian Nasional Berbasis Komputer (UNBK). Faktor internal adalah faktor yang berasal dari dalam diri siswa sedangkan faktor ekstern adalah faktor yang berasal dari luar individu siswa. Faktor internal dari penelitian ini adalah jenis kelamin, sedangkan faktor eksternal adalah pendidikan orang tua, nilai rapor, nilai Ujian Sekolah Berbasis Komputer (USBK). Faktor internal pada dasarnya motivasi adalah dorongan pada diri seseorang untuk mengarahkannya dalam bertindak laku. Jenis kelamin sering menjadi penentu karena berdasarkan kebiasaan siswa yang berjenis kelamin perempuan dianggap lebih rajin dalam belajar, sehingga faktor jenis kelamin ini dianggap penting dalam penelitian ini (Sardiman, 2011).

Selanjutnya adalah faktor eksternal, merupakan interaksi siswa terjadi selama proses belajar dalam hal ini salah satunya adalah pendidikan orang tua, nilai rapor dan nilai ujian sekolah (Slameto, 2010). Pendidikan orang tua dianggap penting karena diprediksikan orang tua yang pendidikannya lebih tinggi akan bisa mendidik anaknya lebih baik sehingga berakibat nilai UNBK menjadi lebih baik. Begitu juga nilai rapor karena nilai rapor ini merupakan nilai yang berproses dari nilai ulangan harian, nilai tengah semester dan nilai ujian akhir semester, sehingga proses belajar yang panjang dapat tergambar pada nilai rapor. Nilai USBK dianggap akan mempengaruhi nilai UNBK, karena soal-soal yang ada pada USBK merupakan soal-soal yang standar kompetensinya sesuai dengan soal UNBK, hanya saja soal ini dikeluarkan dari pemerintah daerah, sedangkan soal UNBK dari pemerintah pusat.

3. METODE PENELITIAN

3.1. Data dan Sumber Data

Sumber data yang digunakan adalah data UNBK SMKN pada jurusan Teknik Komputer Jaringan (TKJ) di Provinsi Nusa Tenggara Barat pada tahun 2017. Variabel yang digunakan terdiri dari 4 variabel respon yaitu nilai UNBK Matematika (y_1), nilai UNBK Bahasa Indonesia (y_2), nilai UNBK Bahasa Inggris (y_3), dan nilai UNBK Kompetensi (y_4). Variabel prediktor yaitu Jenis Kelamin (x_1) dan Nilai Rapor (x_2) sebagai komponen parametrik, Lama Pendidikan Orang Tua (t_1) dan Nilai USBN (t_2) sebagai komponen nonparametrik.

3.2 Langkah-langkah Penelitian

Langkah-langkah estimasi selang kepercayaan parameter model regresi semiparametrik multirespon dan multiprediktor pada komponen nonparametriknya yang diterapkan pada data UNBK Provinsi Nusa Tenggara Barat dilakukan sebagai berikut:

1. melakukan analisis deskriptif pada tiap variabel respon dan variabel prediktor.
2. membuat *scatter plot* antara variabel respon dengan variabel prediktor untuk mengetahui perilaku data.
3. menguji signifikansi korelasi antar respon menggunakan korelasi Pearson.
4. membuat *script* program R antara lain untuk penentuan titik knot optimal berdasarkan metode GCV, mengestimasi titik dan selang kepercayaan parameter model sesuai dengan teori yang telah dihasilkan.
5. menerapkan hasil dari langkah (4) untuk mengestimasi nilai UNBK SMKN di Provinsi Nusa Tenggara Barat dengan pendekatan model regresi semiparametrik multirespon berdasarkan estimator *truncated spline*.

6. menganalisis dan menginterpretasikan hasil, selanjutnya mengambil kesimpulan dengan menentukan variabel prediktor yang berpengaruh signifikan terhadap variabel respon.

4. HASIL DAN PEMBAHASAN

Tabel 1 memberikan gambaran umum dari karakteristik untuk masing-masing variabel. Variabel respon yaitu nilai UNBK Matematika (y_1), nilai UNBK Bahasa Indonesia (y_2), nilai UNBK Bahasa Inggris (y_3), dan nilai UNBK Kompetensi Keahlian (y_4). Variabel prediktor yaitu Jenis Kelamin (x_1); Nilai Rapor Matematika (x_{21}), Nilai Rapor Bahasa Indonesia (x_{22}), Nilai Rapor Bahasa Inggris (x_{23}), dan Nilai Rapor Kompetensi Keahlian (x_{24}); Pendidikan Orang Tua (t_1); dan Nilai USBK Matematika (t_{21}), nilai USBK Bahasa Indonesia (t_{22}), nilai USBK Bahasa Inggris (t_{23}), dan nilai USBK Kompetensi Keahlian (t_{24}).

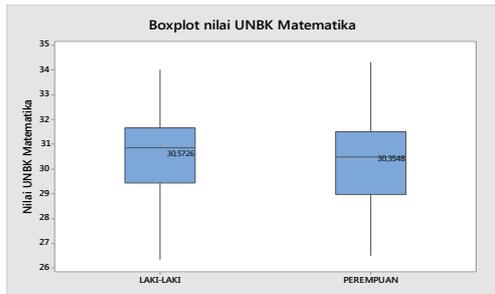
Tabel 1. Statistika Deskriptif Variabel Respon dan Variabel Prediktor

Variabel	Minimum	Maksimum	Rata-Rata	Std Deviasi
Nilai Rapor Matematika	74,00	87,00	81,0571	4,10610
Nilai Rapor B. Indonesia	79,00	92,00	86,0357	3,91715
Nilai Rapor B. Inggris	75,00	90,00	82,6071	4,49892
Nilai Rapor Komp. Keahlian	75,00	90,00	82,2500	4,44814
Lama Pendidikan Orang Tua	3,00	18,00	7,8000	3,55876
Nilai USBK Matematika	68,00	77,00	72,786	2,79277
Nilai USBK B. Indonesia	76,00	90,00	83,0857	4,28732
Nilai USBK B. Inggris	75,00	94,00	85,0357	5,15572
Nilai USBK Komp. Keahlian	75,00	90,00	82,5429	4,02398
Nilai UNBK Matematika	26,34	34,32	30,4559	1,69089
Nilai UNBK B. Indonesia	54,33	68,99	61,7682	2,79034
Nilai UNBK B. Inggris	28,30	41,20	34,7407	2,84713
Nilai UNBK Komp. Keahlian	57,30	67,80	63,9564	2,07722

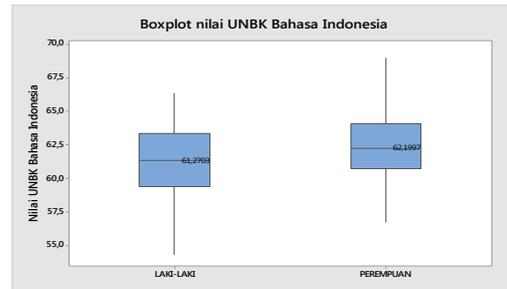
Berdasarkan Tabel 1 diketahui bahwa nilai UNBK Bahasa Indonesia mempunyai rata-rata paling tinggi yaitu 61,76 sedangkan rata-rata nilai UNBK yang paling rendah adalah nilai UNBK Matematika yaitu 30,45. Nilai rapor Bahasa Indonesia mempunyai rata-rata paling tinggi yaitu 86,03 sedangkan rata-rata nilai rapor yang paling rendah adalah nilai rapor Matematika yaitu 81,05. Nilai USBK Bahasa Inggris mempunyai rata-rata paling tinggi yaitu 85,03 sedangkan rata-rata nilai USBK yang paling rendah adalah nilai USBK Matematika yaitu 72,78. Rata-rata pendidikan orang tua adalah 7,8 tahun, berarti para orang tua siswa berpendidikan Sekolah Menengah Pertama (SMP).

Selanjutnya dilakukan identifikasi variabel-variabel yang mempengaruhi nilai Ujian Nasional Berbasis Komputer (UNBK) dilakukan dengan *scatter plot* dan *boxplot*. Nilai UNBK diduga dipengaruhi oleh jenis kelamin, lama pendidikan orang tua, nilai rapor dan nilai Ujian Sekolah Berbasis Komputer (USBK) untuk mata pelajaran Matematika, Bahasa Indonesia, Bahasa Inggris, Kompetensi Keahlian. Untuk variabel jenis kelamin merupakan variabel dummy yang merupakan komponen parametrik. Berikut ini nilai UNBK pada mata pelajaran Matematika, Bahasa Indonesia, Bahasa Inggris, Kompetensi Keahlian berdasarkan jenis kelamin disajikan dalam *boxplot* seperti Gambar 1a, 1b, 1c, dan 1d.

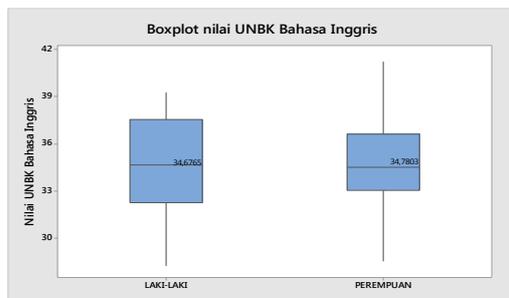
Berdasarkan *Boxplot* pada Gambar 1a, 1b, 1c dan 1d terlihat untuk rata-rata nilai UNBK mata pelajaran Matematika dan Kompetensi Keahlian murid laki-laki lebih dari murid perempuan, namun untuk UNBK mata pelajaran Bahasa Indonesia dan Bahasa Inggris murid perempuan lebih dari murid laki-laki.



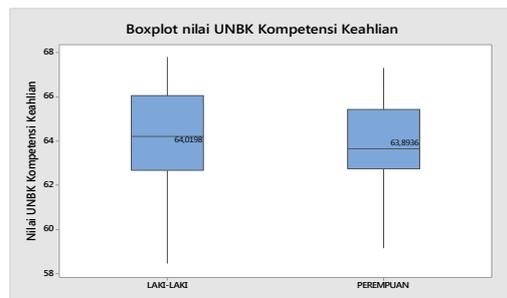
Gambar 1a *Boxplot* antara Nilai UNBK Matematika Berdasarkan Jenis Kelamin



Gambar 1b *Boxplot* antara Nilai UNBK Bahasa Indonesia Berdasarkan Jenis Kelamin

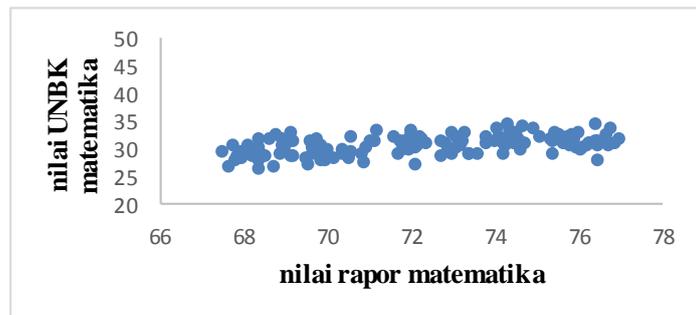


Gambar 1c *Boxplot* antara Nilai UNBK Bahasa Inggris berdasarkan Jenis Kelamin



Gambar 1d *Boxplot* antara Nilai UNBK Kompetensi Keahlian berdasarkan Jenis Kelamin

Identifikasi variabel prediktor nilai rapor yang diwakili oleh rapor Matematika digambarkan dalam *scatter plot* Gambar 2.



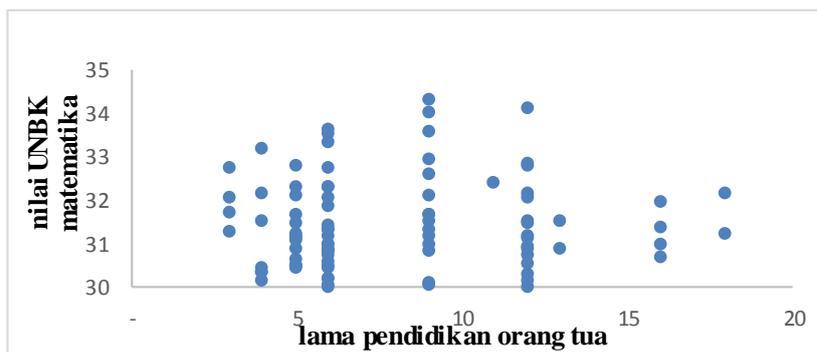
Gambar 2 *Scatter Plot* antara Nilai UNBK Matematika dengan Nilai Rapor Matematika

Berdasarkan *scatter plot* pada Gambar 2 terlihat bahwa variabel nilai rapor merupakan variabel yang membentuk pola tertentu yaitu liner sehingga variabel nilai ini merupakan komponen parametrik dalam model regresi semiparametrik multirespon *truncated spline*. Identifikasi variabel prediktor lama pendidikan orang tua terhadap nilai UNBK pada mata pelajaran Matematika digambarkan dalam *scatter plot* Gambar 3.

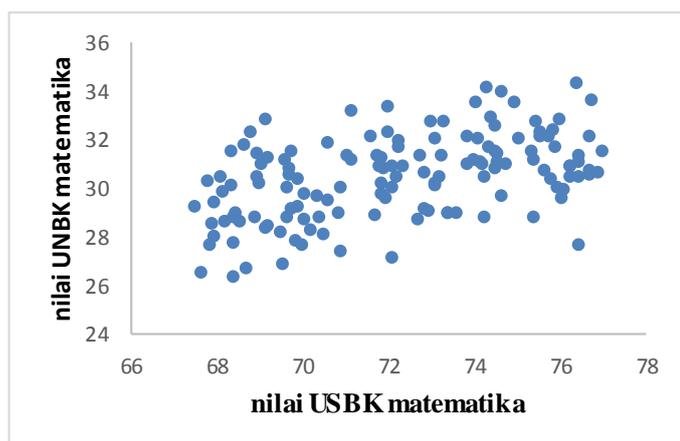
Variabel lama pendidikan orang tua pada Gambar 3 terhadap nilai UNBK Matematika terlihat tidak membentuk pola tertentu, sehingga variabel lama pendidikan orang tua merupakan komponen nonparametrik pada model regresi semiparametrik multirespon *truncated spline*. Identifikasi variabel prediktor nilai USBK terhadap nilai UNBK pada mata pelajaran Matematika digambarkan dalam *scatter plot* Gambar 4.

Berdasarkan pada Gambar 4 bahwa hubungan antara nilai USBK dengan UNBK Matematika terlihat tidak membentuk pola tertentu, sehingga variabel USBK merupakan komponen nonparametrik pada model regresi semiparametrik multirespon *truncated spline*.

Selanjutnya menguji ada tidaknya hubungan antar variabel respon maka dilakukan uji korelasi *Pearson* antar variabel respon dengan hasil tampak pada Tabel 2.



Gambar 3. Scatter Plot antara Nilai UNBK Matematika dengan Lama Pendidikan Orang Tua.



Gambar 4. Scatter Plot antara Nilai UNBK dengan Nilai USBK Matematika

Tabel 2. Hasil Analisis Uji Korelasi *Pearson* Antar Variabel Respon

Variabel Respon	$y^{(1)}$	$y^{(2)}$	$y^{(3)}$	$y^{(4)}$
$y^{(1)}$	1	0,531 (0.000)	0,486(0.000)	0,520(0.000)
$y^{(2)}$	0,531(0.000)	1	0,373(0.000)	0,435(0.000)
$y^{(3)}$	0,486(0.000)	0,373(0.000)	1	0,424(0.000)
$y^{(4)}$	0,520(0.000)	0,435(0.000)	0,424(0.000)	1

Berdasarkan Tabel 2 diperoleh nilai *p-value* semuanya $0,000 < 0,05$ sehingga disimpulkan bahwa terdapat korelasi yang signifikan antar variabel respon. Selanjutnya untuk mendapatkan estimasi model terbaik ditentukan titik knot optimal berdasarkan metode GCV. Dari hasil analisis diperoleh nilai GCV minimum sebesar 2,720 dan nilai MSE sebesar 2,843 dengan titik knot optimal seperti pada Tabel 3. Berdasarkan Tabel 3 pada nilai UNBK Matematika terdapat 3 titik knot optimal pada nilai 10; 71 dan 74. Nilai UNBK Bahasa Indonesia terdapat 4 titik knot optimal pada nilai 7; 11,66; 80,66 dan 85,33. Nilai UNBK Bahasa Inggris terdapat 4 titik knot optimal pada nilai 7; 11,6; 81,33 dan 87,66. Nilai UNBK Kompetensi Keahlian terdapat 3 titik knot optimal pada nilai 7; 11,66 dan 82,5.

Tabel 3. Titik Knot Optimal Untuk Variabel Prediktor Lama Pendidikan Orang Tua, dan Nilai USBK pada Masing-Masing Variabel Respon

Respon 1	Respon 2	Respon 3	Respon 4
K ₁₁ =10	K ₂₁ =7	K ₃₁ =7	K ₄₁ =7
K ₁₂ =71	K ₂₂ =11,66	K ₃₂ =11,66	K ₄₂ =11,66
K ₁₃ =74	K ₂₃ =80,66	K ₃₃ =81,33	K ₄₃ =82,50
	K ₂₄ =85,33	K ₃₄ =87,66	

Estimasi terbaik model semiparametrik multirespon *truncated* spline untuk masing-masing respon ditunjukkan sebagai berikut:

$$\begin{aligned}
 y^{(1)} &= -0.097 + 0.192x_1^{(1)} - 0.084^{(1)}x_2^{(1)} - 0.014^{(1)}t_1^{(1)} + 0.077^{(1)}(t_1^{(1)} - 10^{(1)})_+ \\
 &\quad + 0.526^{(1)}t_2^{(1)} - 0.250^{(1)}(t_2^{(1)} - 71^{(1)})_+ - 0.301^{(1)}(t_2^{(1)} - 74^{(1)})_+ + \varepsilon^{(1)} \\
 y^{(2)} &= -0.026^{(2)} + 0.185^{(2)}x_1^{(2)} + 0.423^{(2)}x_2^{(2)} - 0.029^{(2)}t_1^{(2)} + 0.117^{(2)}(t_1^{(2)} - 7^{(2)})_+ \\
 &\quad - 0.136^{(2)}(t_1^{(2)} - 11.66^{(2)})_+ + 0.311^{(2)}t_2^{(2)} - 0.081^{(2)}(t_2^{(2)} - 80.66^{(2)})_+ \\
 &\quad - 0.250^{(2)}(t_2^{(2)} - 85.33^{(2)})_+ + \varepsilon^{(2)} \\
 y^{(3)} &= -0.032^{(3)} + 0.323^{(3)}x_1^{(3)} + 0.304^{(3)}x_2^{(3)} - 0.097^{(3)}t_1^{(3)} + 0.189^{(3)}(t_1^{(3)} - 7^{(3)})_+ \\
 &\quad - 0.182^{(3)}(t_1^{(3)} - 10.66^{(3)})_+ + 0.098^{(3)}t_2^{(3)} + 0.689^{(3)}(t_2^{(3)} - 81.33^{(3)})_+ \\
 &\quad - 1.503^{(3)}(t_2^{(3)} - 87.66^{(3)})_+ + \varepsilon^{(3)} \\
 y^{(4)} &= 2.084^{(4)} - 0.002^{(4)}x_1^{(4)} + 0.287^{(4)}x_2^{(4)} + 0.038^{(4)}t_1^{(4)} - 0.001^{(4)}(t_1^{(4)} - 7^{(4)})_+ \\
 &\quad - 0.056^{(4)}(t_1^{(4)} - 11.66^{(4)})_+ + 0.480^{(4)}t_2^{(4)} - 0.961^{(4)}(t_2^{(4)} - 82.50^{(4)})_+ + \varepsilon^{(4)}
 \end{aligned}$$

Berdasarkan estimasi titik pada model terbaik, selanjutnya diestimasi selang kepercayaan parameter model semiparametrik untuk mengetahui variabel prediktor yang diduga mempengaruhi nilai UNBK berpengaruh secara signifikan atau tidak. Hasil estimasi titik dan estimasi selang kepercayaan parameter model regresi semiparametrik multirespon *truncated* spline secara lengkap ditunjukkan pada Tabel 4.

Berdasarkan hasil selang kepercayaan 95% parameter model pada Tabel 4, variabel prediktor yang tidak berpengaruh signifikan terhadap nilai UNBK adalah jenis kelamin dan lama pendidikan orang tua karena selang kepercayaan parameternya diperoleh rentang dari nilai negatif sampai nilai positif yg memuat nilai nol, sedangkan yang berpengaruh signifikan positif adalah nilai rapor dan nilai ujian sekolah (USBK) yang memiliki rentang dengan batas bawah sampai batas atas bernilai positif. Setiap peningkatan nilai rapor satu satuan akan meningkatkan nilai UNBK Bahasa Indonesia, Bahasa Inggris dan Kompetensi keahlian masing- masing sebesar 0,423; 0,304 dan 0,287, namun hanya pada nilai UNBK mata pelajaran Matematika saja nilai rapor berpengaruh negatif yaitu sebesar -0,084. Pengaruh nilai ujian sekolah (USBK) secara umum semuanya berpengaruh positif, yaitu semakin tinggi nilai USBK semakin tinggi nilai UNBK-nya.

Tabel 4 Estimasi Selang Kepercayaan 95% Parameter Model

Parameter	Estimator ($\hat{\theta}$)	$\sqrt{\text{Var}(\hat{\theta})}$	Z-value	Batas bawah	Batas Atas	Sig.
$\beta_0^{(1)}$	-0,097	0,104	-0,934	-0,301	0,107	Tidak Signifikan
$\beta_1^{(1)}$	0,192	0,246	0,779	-0,292	0,676	Tidak Signifikan

$\tilde{\beta}_2^{(1)}$	-0,084	0,010	-8,693	-0,104	-0,065	Signifikan
$\tilde{\chi}_{11}^{(1)}$	-0,014	0,052	-0,274	-0,117	0,088	Tidak Signifikan
$\tilde{\chi}_{12}^{(1)}$	0,077	0,106	0,726	-0,131	0,285	Tidak Signifikan
$\tilde{\chi}_{21}^{(1)}$	0,526	0,013	41,577	0,501	0,551	Signifikan
$\tilde{\chi}_{22}^{(1)}$	-0,250	0,045	-5,578	-0,338	-0,162	Signifikan
$\tilde{\chi}_{23}^{(1)}$	-0,301	0,09	-3,360	-0,477	-0,125	Signifikan
$\tilde{\beta}_0^{(2)}$	-0,026	0,029	-0,898	-0,084	0,031	Tidak Signifikan
$\tilde{\beta}_1^{(2)}$	0,185	0,371	0,498	-0,544	0,913	Tidak Signifikan
$\tilde{\beta}_2^{(2)}$	0,423	0,016	27,116	0,392	0,454	Signifikan
$\tilde{\chi}_{12}^{(2)}$	-0,029	0,12	-0,241	-0,264	0,206	Tidak Signifikan
$\tilde{\chi}_{12}^{(2)}$	0,117	0,145	0,811	-0,167	0,402	Tidak Signifikan
$\tilde{\chi}_{13}^{(2)}$	-0,136	0,149	-0,913	-0,429	0,157	Tidak Signifikan
$\tilde{\chi}_{21}^{(2)}$	0,311	0,019	16,376	0,274	0,349	Signifikan
$\tilde{\chi}_{22}^{(2)}$	-0,081	0,055	-1,476	-0,19	0,027	Tidak Signifikan
$\tilde{\chi}_{23}^{(2)}$	-0,250	0,102	-2,448	-0,45	-0,049	Signifikan
$\tilde{\beta}_0^{(3)}$	-0,032	0,009	-3,680	-0,049	-0,015	Signifikan
$\tilde{\beta}_1^{(3)}$	0,323	0,265	1,217	-0,198	0,844	Tidak Signifikan
$\tilde{\beta}_2^{(3)}$	0,304	0,015	20,481	0,275	0,334	Signifikan
$\tilde{\chi}_{12}^{(3)}$	-0,097	0,099	-0,973	-0,291	0,098	Tidak Signifikan
$\tilde{\chi}_{12}^{(3)}$	0,189	0,133	1,421	-0,072	0,450	Tidak Signifikan
$\tilde{\chi}_{13}^{(3)}$	-0,182	0,127	-1,428	-0,432	0,068	Tidak Signifikan
$\tilde{\chi}_{21}^{(3)}$	0,098	0,016	5,997	0,066	0,131	Signifikan
$\tilde{\chi}_{22}^{(3)}$	0,689	0,039	17,471	0,611	0,766	Signifikan
$\tilde{\chi}_{23}^{(3)}$	-1,503	0,070	-21,391	-1,641	-1,365	Signifikan
$\tilde{\beta}_0^{(4)}$	2,084	2,011	1,037	-1,866	6,035	Tidak Signifikan
$\tilde{\beta}_1^{(4)}$	-0,002	0,223	-0,009	-0,441	0,437	Tidak Signifikan
$\tilde{\beta}_2^{(4)}$	0,287	0,009	30,687	0,269	0,305	Signifikan
$\tilde{\chi}_{11}^{(4)}$	0,038	0,076	0,498	-0,111	0,186	Tidak Signifikan
$\tilde{\chi}_{12}^{(4)}$	-0,001	0,095	-0,013	-0,187	0,185	Tidak Signifikan
$\tilde{\chi}_{21}^{(4)}$	-0,056	0,094	-0,599	-0,242	0,129	Tidak Signifikan
$\tilde{\chi}_{22}^{(4)}$	0,480	0,023	21,104	0,436	0,525	Signifikan
$\tilde{\chi}_{23}^{(4)}$	-0,961	0,039	-24,615	-1,038	-0,884	Signifikan

Untuk melihat kesesuaian model, selanjutnya dilakukan uji asumsi error model yaitu:

a. Pengujian Normal Multivariat Error

Asumsi yang harus dipenuhi dalam model regresi semiparametrik multirespon adalah dari *error* berdistribusi normal multivariat. Pengujian dilakukan dengan hipotesis:

Hipotesis H_0 : data berdistribusi normal multivariat

H_1 : data tidak berdistribusi normal multivariat

Statistik uji Henze-Zikler (HZ)

$$HZ = \frac{1}{m} \sum_{i=1}^p \sum_{j=1}^m e^{-\frac{p}{2} D_{ij}} - 2(1 + \beta)^{-\frac{p}{2}} \sum_{i=1}^p e^{-\frac{\beta^2}{2+1(\beta^2)} D_{ij}} + m(1 + 2\beta^2)^{-\frac{p}{2}}$$

$i = 1, 2, 3, \dots, p; \quad j = 1, 2, 3, \dots, m$

Kriteria Berdasarkan Svantesson dan Wallace (2010) nilai statistik uji $HZ > p\text{-value}$ maka H_0 diterima yaitu data berdistribusi normal multivariat. *Package MVTests* di OSS-R dan *function mvn* digunakan untuk memperoleh nilai $p\text{-value}=0,9130649 > 0,05$, maka dapat disimpulkan *error* berdistribusi normal multivariat.

b. Uji Mean dari Error

Setelah uji normalitas multivariat terpenuhi selanjutnya adalah asumsi mean dari error yaitu nol dengan hipotesis sebagai berikut:

Hipotesis $H_0 : \bar{\varepsilon}^{(r)} = 0 \quad r = 1, 2, 3, 4$

$H_1 : \bar{\varepsilon}^{(r)} \neq 0 \quad r = 1, 2, 3, 4$

Statistik uji $t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$

Kriteria H_0 ditolak jika nilai $p\text{-value} < 0,05$. Berdasarkan hasil pengujian mean *error* untuk masing-masing respon diperoleh nilai $p\text{-value}$ masing-masing sebesar 0,9844; 0,9796; 0,9872 dan 0,9832. Dengan nilai $p\text{-value}$ yang semuanya $> 0,05$ maka keputusan H_0 diterima sehingga diperoleh kesimpulan bahwa asumsi mean *error* sama dengan nol terpenuhi.

5. KESIMPULAN

Berdasarkan hasil estimasi selang kepercayaan 95% parameter model regresi semiparametrik multirespon berdasarkan estimator *truncated spline* pada nilai UNBK di Provinsi NTB diperoleh rentang dari nilai negatif sampai nilai positif yang memuat nilai yang memuat nol sehingga disimpulkan bahwa jenis kelamin dan lama pendidikan orang tua tidak berpengaruh signifikan terhadap nilai UNBK. Hal ini berarti bahwa baik murid laki-laki maupun murid perempuan mempunyai kesempatan yang sama untuk memperoleh nilai UNBK untuk semua mata pelajaran dan tidak terpengaruh dengan lama pendidikan orang tuanya. Secara umum variabel nilai rapor dan nilai ujian sekolah berada pada rentang batas bawah dan batas atas positif sehingga disimpulkan semakin besar nilai rapor dan ujian sekolah semakin besar pula nilai UNBK yang bersesuaian dengan mata pelajarannya. Hal ini berarti kalau seorang murid bagus dalam pembelajaran sehari-hari yang tercermin dalam nilai rapor dan nilai ujian sekolah maka nilai UNBK-nya juga akan bagus.

DAFTAR PUSTAKA

- Arikunto, S. 2001. *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi. Aksara
- Budiantara, I. N. 2009. *Spline dalam Regresi Nonparametrik dan Semiparametrik, sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang*. Pidato Pengukuhan Untuk Jabatan Guru Besar dalam Bidang Ilmu: Matematika Statistika dan Probabilitas, Jurusan Statistika FMIPA ITS, Surabaya
- Chamidah, N. dan Rifada, M. 2016. Local Linier Estimator In Bi-Response Semiparametric Regression Model For Estimating Median Growth Charts of

- Children. *Far East Journal of Mathematical Sciences*, Vol. 99, No. 8, hal 1233-1244.
- Chamidah, N., Kurniawan, K., Zaman, B., dan Muniroh, L. 2018. Least Square-Spline Estimator in Multi-Response Semiparametric Regression Model for Estimating Median Growth Charts of Children in East Java, Indonesia. *Far East Journal of Mathematical Sciences*, Vol.107, No. 2, hal 295-307.
- Chamidah, N dan Lestari, B. 2019. Estimation of Covariance Matrix using Multi-Response Local Polynomial Estimator for Designing Children Growth Charts: a Theoretically Discussion. *Journal of Physics: Conference Series*, Vol. 1397 012072
- Hidayati, L., Chamidah, N., dan Budiantara, I. N. 2019. Truncated Spline Estimator in Multiresponse Semiparametric Regression Model For Computer Based National Exam In West Nusa Tenggara. *IOP Conf. Series: Material Science and Engineering*, Vol. 546 052029 doi: 10.1088/1757-899X/546/5/052029
- Hidayati, L., Chamidah, N., dan Budiantara, I. N. 2020. Confidence Interval of Multiresponse Semiparametric Regression Model Parameters Using Truncated Spline. *International Journal of Academic Applied Research*, Vol. 4, Issue 1 hal 14-18.
- Mahmud, D. 1989. *Psikologi Pendidikan*. Jakarta: Depdikbud Dirjen
- Prahatama, A. dan Suparti. 2017. Pemodelan Harga Cabai di Kota Semarang terhadap Harga Inflasi Menggunakan Regresi Semiparametrik Polinomial Lokal, *Jurnal Statistika*, Vol. 5, No. 1, hal 1-7.
- Puspendik Balitbang Kemendikbud. 2017. *Panduan Pemanfaatan Hasil Ujian Nasional Tahun Pelajaran 2016/2017*. Jakarta: BSNP
- Ramadan, W., Chamidah, N., Zaman, B., Muniroh, L., dan Lestari, B. 2019. Standard Growth Chart of Weight for Height to Determine Wasting Nutritional Status in East Java Based on Semiparametric Least Square Spline Estimator. *IOP Conf. Series: Materials Science and Engineering*, Vol. 546 052063 doi:10.1088/1757-899X/546/5/052063
- Sardiman. 2011. *Interaksi dan Motivasi Belajar Mengajar*. Jakarta: PT. RajaGrafindo Persada.
- Slameto. 2010. *Belajar dan Faktor-faktor yang Mempengaruhinya*. Jakarta: PT Rineka Cipta
- Svantesson T., dan Wallace J. 2010. *Tests for Assessing Multivariate Normality and The Covariance Structure of Mimo Data*. National Science Foundation under Wireless Initiatives Grant CCR-9979452 and Information Technology Research Grant CCR-0081476
- Syaodih, S.N. 2007, *Metode Penelitian Pendidikan*. Bandung; Remaja Rosda Karya.
- Utami, T.W., Nur, I.M., dan Endah, S. 2018. Pemodelan Produksi Padi Menggunakan Regresi Semiparametrik Kernel. *Jurnal Statistika*, Vol. 6, No. 2, hal. 160-165.
- Wibowo, W., Haryatmi, S., dan Budiantara, I. N. 2012. On Multiresponse Semiparametric Regression Model. *Journal of Mathematics and Statistics*, Vol. 8, No. 4: 489-499.
- Yang. J. dan Yang. H. 2016. A Robust Penalized Estimation for Identification in Semiparametric Additive Models. *Statistics and Probability Letters*, Vol. 110, hal 268-277