

Aplikasi *Software R* pada Analisis Angka Harapan Hidup di Jawa Timur Menggunakan Estimator *P-Spline*

Helmina Andriani¹, Wahyu Wibowo², Santi Puteri Rahayu³, Muhammad Rijal Alfian⁴

Politeknik Medica Farma Husada Mataram¹

Institut Teknologi Sepuluh Nopember^{2,3}, Universitas Teknologi Mataram⁴

ena.andriani@gmail.com¹, rijal.alfian@gmail.com⁴

Abstrak – Derajat kesehatan yang tinggi dapat digunakan sebagai indikator keberhasilan program kesehatan dan program pembangunan sosial ekonomi yang secara tak langsung dapat meningkatkan angka harapan hidup (AHH). Secara keseluruhan AHH Jawa Timur mengalami peningkatan, namun masih terdapat beberapa kabupaten/kota yang memiliki AHH relatif rendah, seperti di Probolinggo yang hanya mencapai 61,78 tahun atau Sampang 64,39 tahun. Sementara Kota Blitar mencapai 72,99 tahun. Secara teori, salah satu faktor yang mempengaruhi perubahan data AHH adalah PDRB perkapita. Semakin tinggi pendapatan seseorang maka akan semakin tinggi pula kemampuan seseorang untuk membayar berbagai pungutan yang ditetapkan pemerintah, misalnya biaya kesehatan. Pengaruh PDRB perkapita terhadap AHH di Jawa Timur dapat diestimasi menggunakan estimator *p-spline*. Metode ini merupakan gabungan dari dua macam pendekatan spline yaitu regresi spline dan *smoothing spline*. *Software* yang cukup populer dalam pengolahan data adalah *software R*, yaitu perangkat lunak pengolahan data yang lebih *powerfull* dibandingkan dengan *software* statistik lainnya, terutama dalam hal manipulasi, pemodelan, *machine learning* dan simulasi data. Tujuan penelitian ini adalah menganalisis data AHH di Jawa Timur menggunakan estimator *p-spline* dengan bantuan *software R*. Langkah penelitian meliputi pendekatan model AHH dengan regresi nonparametrik spline, menentukan jumlah knot, menghitung nilai GCV untuk jumlah knot yang diperoleh, menentukan lokasi knot, menghitung parameter pemulus, mendapatkan estimator *p-spline*, mendapatkan estimasi model dan interpretasi hasil. Pada penelitian ini ditunjukkan *syntax R* yang digunakan dalam pengolahan data AHH. Model regresi nonparametrik menggunakan estimator *p-spline* pada analisis data AHH di Jawa Timur menunjukkan hasil bahwa pada interval tertentu: (a) kenaikan PDRB perkapita sebesar satu satuan maka AHH akan cenderung bertambah, maksimal sebesar 12,632 tahun, (b) kenaikan PDRB perkapita sebesar satu satuan maka AHH akan cenderung berkurang, maksimal sebesar 10,903 tahun.

Kata kunci: angka harapan hidup; *penalized spline*; R

Abstract – High health status can be used as an indicator of the success of health programs and socioeconomic development programs that can indirectly increase life expectancy (AHH). Overall, AHH in East Java has increased, but there are still some districts / cities that have relatively low AHH, such as in Probolinggo which only reaches 61,78 years or Sampang 64,39 years. While the City of Blitar reaches 72,99 years. In theory, one of the factors affecting changes in AHH data is PDRB per capita. The higher one's income, the higher one's ability to pay various levies set by the government, for example health costs. The effect of PDRB per capita on AHH in East Java can be estimated using the *p-spline* estimator. This method is a combination of two kinds of spline approaches namely spline regression and *smoothing spline*. *Software* that is quite popular in data processing is *R software*, which is data processing software that is more powerful than other statistical software, especially in terms of manipulation, modeling, machine learning and data simulation. The purpose of this study is to analyze AHH data in East Java using the *p-spline* estimator with the help of *R software*. The research steps include the AHH model approach with spline nonparametric regression, determine the number of knots, calculate the GCV value for the number of knots obtained, determine the location of the knots, calculate smoothing parameters, get the *p-spline* estimator, get the model estimation and interpret the results. This research shows that *syntax R* is used in AHH data processing. Nonparametric regression model using *p-spline* estimator in AHH data analysis in East Java shows the results that at certain intervals: (a) an increase in PDRB per capita by one unit then AHH will tend to increase, a maximum of 12,632 years, (b) an increase in PDRB per capita then AHH units will tend to decrease, a maximum of 10.903 years.

Keyword : life expectancy; *penalized spline*; R;

1. Latar Belakang

Derajat kesehatan menjadi salah satu pilar penentu kualitas hidup manusia selain pendidikan. Oleh karena itu, diperlukan kepedulian tinggi dari pemerintah dan seluruh masyarakat untuk senantiasa peduli pada peningkatan derajat kesehatan. Menurut Henrik L. Blum, peningkatan derajat kesehatan masyarakat dapat diukur dari tingkat mortalitas dan morbiditas penduduk. Derajat kesehatan dipengaruhi oleh empat faktor penentu, yaitu faktor lingkungan (45%), perilaku kesehatan (30%), pelayanan kesehatan (20%), dan kependudukan/keturunan (5%). Oleh karena itu, analisis mengenai derajat kesehatan penduduk dapat dilihat melalui empat aspek tersebut [2]. Derajat kesehatan yang tinggi dapat digunakan sebagai indikator keberhasilan program kesehatan dan program pembangunan sosial ekonomi yang secara tak langsung dapat meningkatkan angka harapan hidup [11].

Secara definisi, angka harapan hidup (AHH) adalah perkiraan rata-rata lamanya hidup yang akan dicapai oleh sekelompok penduduk dari sejak lahir. Data AHH diperoleh melalui survey yang dilakukan oleh Badan Pusat Statistik (BPS). Penduduk Indonesia diperkirakan akan mencapai 273,65 juta jiwa pada tahun 2025. Pada tahun yang sama, AHH nasional diperkirakan mencapai 73,7 tahun. Pada tahun 2012, berdasarkan data BPS Provinsi Jawa Timur, AHH Provinsi Jawa Timur mencapai 70,09 tahun. Angka ini meningkat dibandingkan tahun 2011, yaitu sebesar 69,81 tahun [6]. Peningkatan AHH Provinsi Jawa Timur kembali terjadi pada tahun 2013 yaitu mencapai 70,19 tahun. Meskipun secara keseluruhan AHH Jawa Timur mengalami peningkatan, namun masih terdapat beberapa kabupaten/kota dengan AHH yang relatif rendah, seperti di Probolinggo yang hanya mencapai 61,78 tahun atau Sampang 64,39 tahun. Sementara daerah lain seperti Kota Blitar mencapai 72,99 tahun [7].

Secara teori, salah satu faktor yang mempengaruhi perubahan data AHH di Jawa Timur adalah PDRB perkapita. Semakin tinggi pendapatan seseorang maka akan semakin tinggi pula kemampuan seseorang untuk membayar berbagai pungutan yang ditetapkan pemerintah, misalnya biaya kesehatan [1]. PDRB perkapita diperoleh dari pembagian PDRB terhadap jumlah penduduk [3].

Pengaruh PDRB perkapita terhadap AHH di Jawa Timur dapat diestimasi menggunakan estimator *p-spline* atau *penalized spline regression*. Metode ini

merupakan gabungan dari dua macam pendekatan spline yaitu regresi spline dan *smoothing spline* [8]. Spline merupakan salah satu jenis *piecewise polynomial*, yaitu polinomial yang memiliki sifat tersegmen. Sifat tersegmen ini memberikan fleksibilitas lebih dari polinomial biasa, sehingga memungkinkan untuk menyesuaikan diri secara lebih efektif terhadap karakteristik lokal suatu fungsi atau data [4].

Pada praktiknya, dalam melakukan analisis data perlu dibantu oleh beberapa software pengolahan data, diantaranya adalah SPSS, Minitab, SAS, Spuls, Open Stats, SalStat dan R [12]. R adalah suatu *software* terintegrasi yang memiliki fasilitas untuk memanipulasi data, perhitungan, dan penampilan grafik. Kemampuan R dalam mengolah dan menganalisis data cukup baik sehingga cukup banyak dipakai oleh masyarakat [5].

Berdasarkan uraian di atas, akan dilakukan analisis terhadap data AHH di Jawa Timur menggunakan estimator *p-spline* dengan bantuan *software* R.

2. Kajian Pustaka

a. *Penalized Spline Regression (P-Spline)*

Model regresi nonparametrik merupakan salah satu pendekatan yang dapat digunakan untuk mengestimasi fungsi regresi dalam analisis regresi [13]. Analisis regresi merupakan model matematika yang digunakan untuk meneliti hubungan variabel dependen dengan satu atau lebih variabel independen [14]. Estimasi fungsi regresi dapat dilakukan dengan beberapa metode, diantaranya adalah metode kernel, spline dan ekspansi deret Wavelet dan Fourier [8]. Metode yang populer digunakan adalah spline [4] karena mampu menggambarkan perubahan pola perilaku dari fungsi pada sub-sub interval tertentu dan dapat digunakan untuk mengatasi pola data yang mengalami kenaikan atau penurunan tajam dengan bantuan titik-titik knot, serta kurva yang dihasilkan relatif mulus [9].

Titik knot merupakan titik perpaduan bersama dimana terdapat perubahan pola perilaku fungsi atau kurva. Estimator spline terbaik diperoleh dengan menggunakan titik knot optimal. Titik knot optimal dapat diperoleh dengan menggunakan metode *generalized cross validation* (GCV) [11].

Terdapat dua komponen penting dalam mengestimasi *p-spline*, yaitu pemilihan parameter penghalus (*smoothing parameter*) dan pemilihan jumlah knot dan lokasinya [15]. Nilai parameter penghalus ditentukan

berdasarkan kriteria GCV, sedangkan penentuan jumlah titik knot dan lokasi knot yang optimum ditentukan dengan *fixed selection method*.

b. Software R

R adalah suatu sistem untuk analisis data yang termasuk kelompok *software* statistik *open source* yang tidak memerlukan lisensi atau gratis, yang dikenal dengan *freeware* [12]. Menurut Sarvina [10] dalam penelitiannya menyebutkan bahwa *software* R adalah perangkat lunak pengolahan data yang lebih *powerfull* dibandingkan dengan *software* statistik lainnya, terutama dalam hal manipulasi, pemodelan, *machine learning* dan simulasi data. *Software* ini memiliki perangkat dan *library* yang banyak dan berkembang cepat. Bahkan jika ingin menggunakan paket dengan fungsi yang belum tersedia maka pengguna dapat mengembangkan sendiri dengan cara menurunkan dari paket yang sudah ada atau mengembangkan dari awal. Dengan kata lain, pengguna dapat mengembangkan sesuai dengan ide dan permasalahan yang akan diatasi, karena *software* R merupakan perangkat pengolah data yang menggunakan bahasa pemrograman.

3. Metode Penelitian

Data pada penelitian ini diambil berdasarkan hasil Survey Sosial Ekonomi Nasional (SUSENAS) Jawa Timur tahun 2013. Unit observasi dalam penelitian ini adalah 38 kabupaten/kota di Jawa Timur. Variabel penelitian meliputi variabel dependen (y) yaitu AHH dan variabel independen (x) yaitu PDRB perkapita. Secara sistematis, tahapan analisis data AHH di Provinsi Jawa Timur menggunakan estimator *p-spline* adalah sebagai berikut:

1. Memodelkan AHH di Jawa Timur menggunakan regresi nonparametrik dengan pendekatan spline
2. Menentukan jumlah knot yang digunakan berdasarkan *fixed selection method*
3. Menghitung nilai GCV untuk jumlah knot yang diperoleh
4. Memilih jumlah knot optimum berdasarkan nilai GCV minimum
5. Menentukan lokasi knot
6. Menghitung parameter pemulus yang optimal berdasarkan jumlah knot optimal menggunakan kriteria GCV

7. Mendapatkan estimator spline terpenalti menggunakan metode *penalized least square* (PLS)
8. Mendapatkan estimasi model nonparametrik
9. Melakukan interpretasi model dan menarik kesimpulan.

4. Implementasi Sistem dan Hasil

Pendekatan model regresi untuk data AHH Jawa Timur dilakukan dengan menggunakan model spline linear yaitu:

$$y_i = \beta_0 + \beta_1 x_i + \sum_{k=1}^K u_k (x_i - \kappa_k)_+ + \varepsilon_i \quad (4.1)$$

Proses berikutnya adalah menghitung nilai GCV untuk masing-masing jumlah knot yang ditentukan berdasarkan *fixed selection method*, yaitu:

$$\kappa = \min\left(\frac{1}{4} \times n, 35\right) = 9,5 \approx 10$$

artinya, jumlah knot yang digunakan dalam penelitian ini adalah 1, 2, 3, ..., 10. Berdasarkan kriteria GCV minimum, diperoleh jumlah knot optimum sebanyak 8 buah knot yang masing-masing terletak pada titik 5,492; 6,078; 6,340; 6,905; 7,728; 9,213; 10,650; 16,868. Nilai-nilai tersebut diperoleh dengan pengolahan data menggunakan *software* R, syntax yang digunakan dapat dilihat pada Gambar 4.1.

Selanjutnya dihitung nilai parameter pemulus (λ) optimum, karena jika nilai λ terlalu kecil akan memberikan bentuk kurva regresi yang kasar sedangkan jika nilai λ terlalu besar akan memberikan bentuk kurva regresi yang sangat halus. Untuk itu, perlu dilakukan perhitungan nilai λ optimal. Output hasil pengolahan data diperoleh λ optimal adalah 0,00507944. Visualisasi nilai λ dapat dilihat pada Gambar 4.2 (a) dan (b).

Dengan demikian, estimasi model regresi nonparametrik berdasarkan estimator *p-spline* adalah

$$\begin{aligned} \hat{y} = & 66,102 + 0,343x - 2,102(x - 5,492)_+ \\ & + 14,391(x - 6,078)_+ - 23,535(x - 6,340)_+ \\ & + 14,803(x - 6,905)_+ - 1,706(x - 7,728)_+ \\ & - 1,448(x - 9,213)_+ - 0,768(x - 10,650)_+ \\ & + 0,018(x - 16,868)_+ \end{aligned} \quad (4.2)$$

```

# Membangkitkan data
data <- matrix(scan("D://data thesis 38 kab.kota.txt"),byrow=TRUE,38,2)
t <- data[,1]
y <- data[,2]
n <- length(y)

#Mendapatkan knot dan derajat yang optimum
GCVmin<- Inf
GCVmax<- 0
j <- 0
Kmax <- 10
pmax <- 1
KpGCV <- matrix(c(rep(0,(Kmax*pmax*3))), (Kmax*pmax), 3)
for (k in 1:Kmax)
{
  for (p in 1:pmax)
  {
    knot <- matrix(c(rep(0,(3*k))),3,k)
    tunique <- unique(t)
    for (k in 1:k) knot[k] <- quantile(tunique,(k+1)/(k+2))
    x1t <- -1
    for (d in 1:p)
    {
      x1t<-cbind(x1t,t^d)
    }
    x1knot<- (pmax(0,t-knot[1,1]))^p
    if (k>1)
    {
      for (nk in 2 : k)
      {
        x1knot<-cbind(x1knot,(pmax(0,t-knot[1,nk]))^p)
      }
    }
    X <- cbind(x1t,x1knot)
    A <- X%*(ginv(t(X)%*X))%*t(X)
    In <- diag(n)
    IA <- In-A
    GCV <- (1/n)*t(IA%*y)%*(IA%*y)/((1/n)*(sum(diag(IA))))
    if (GCV<=GCVmin)
    {
      GCVmin<- GCV
      Kopt <- k
      popt <- p
    }
    if (GCV>=GCVmax)
    {
      GCVmax <- GCV
      K.GCVmax <- k
      p.GCVmax <- p
    }
    j <- j+1
    KpGCV[j,] <- c(k,p,GCV)
  }
}
}

```

Gambar 4.1. Syntax R untuk analisis AHH menggunakan estimator *p-spline*

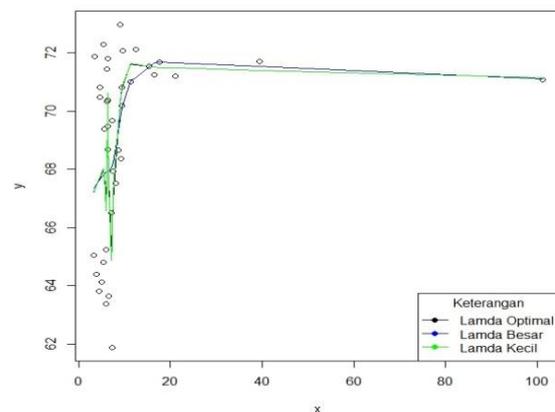
$$\hat{y} = \begin{cases} 0,343x + 66,102 & ; x < 5,492 \\ -1,759x + 77,646 & ; 5,492 \leq x < 6,078 \\ 12,632x + 1,722 & ; 6,078 \leq x < 6,340 \\ -10,903x + 75,010 & ; 6,340 \leq x < 6,905 \\ 3,900x + 46,083 & ; 6,905 \leq x < 7,728 \\ 2,194x + 30,340 & ; 7,728 \leq x < 9,213 \\ 0,746x + 27,937 & ; 9,213 \leq x < 10,650 \\ -0,022x + 33,713 & ; 10,650 \leq x < 16,868 \\ -0,004 + 39,186 & ; x \geq 16,868 \end{cases} \quad (4.3)$$

Persamaan (4.3) di atas dapat diinterpretasikan sebagai berikut

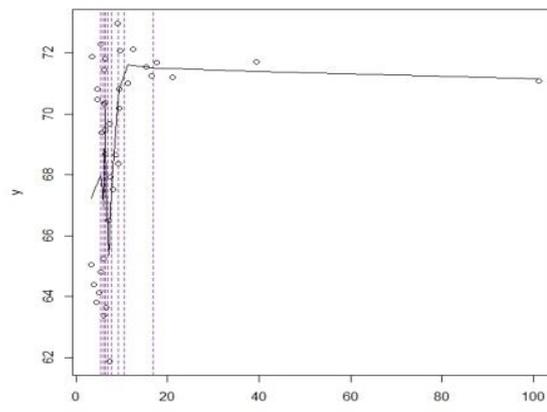
- Untuk nilai PDRB perkapita kurang dari 5,492, setiap kenaikan PDRB perkapita sebesar satu satuan maka AHH akan cenderung bertambah sebesar 0,343 tahun.
- Untuk nilai PDRB perkapita antara 5,492 dan 6,078, setiap kenaikan PDRB perkapita sebesar satu satuan maka AHH akan cenderung berkurang sebesar 1,759 tahun.
- Untuk nilai PDRB perkapita antara 6,078 dan 6,340, setiap kenaikan PDRB perkapita sebesar satu satuan maka AHH akan cenderung bertambah sebesar 12,632 tahun.
- Untuk nilai PDRB perkapita antara 6,340 dan 6,905, setiap kenaikan PDRB perkapita sebesar satu satuan maka AHH akan cenderung berkurang sebesar 10,903 tahun.
- Untuk nilai PDRB perkapita antara 6,905 dan 7,728, setiap kenaikan PDRB perkapita sebesar satu satuan maka AHH akan cenderung bertambah sebesar 3,9 tahun.
- Untuk nilai PDRB perkapita antara 7,728 dan 9,213, setiap kenaikan PDRB perkapita sebesar satu satuan maka AHH akan cenderung bertambah sebesar 2,194 tahun.
- Untuk nilai PDRB perkapita antara 9,213 dan 10,650, setiap kenaikan PDRB perkapita sebesar satu satuan maka AHH akan cenderung bertambah sebesar 0,746 tahun.
- Untuk nilai PDRB perkapita antara 10,650 dan 16,868, setiap kenaikan PDRB perkapita sebesar satu satuan maka AHH akan cenderung berkurang sebesar 0,022 tahun.

- Untuk nilai PDRB perkapita lebih dari 16,868, setiap kenaikan PDRB perkapita sebesar satu satuan maka AHH akan cenderung berkurang sebesar 0,004 tahun.

Seberapa besar proporsi keragaman (variasi) AHH yang dapat dijelaskan oleh PDRB perkapita diukur menggunakan nilai koefisien determinasi. Nilai yang diperoleh adalah 36,76%, artinya 36,76% variasi dalam data AHH yang mampu dijelaskan oleh PDRB perkapita.



(a)



(b)

Gambar 4.2 Grafik regresi nonparametrik dengan 8 titik knot berdasarkan
(a) Nilai λ besar, λ optimum, dan λ kecil
(b) Nilai λ optimum dan lokasi knot

5. Kesimpulan

Model regresi nonparametrik menggunakan estimator *p-spline* pada analisis data AHH di Jawa Timur menunjukkan hasil bahwa pada interval tertentu:

- a. kenaikan PDRB perkapita sebesar satu satuan maka AHH akan cenderung bertambah, maksimal sebesar 12,632 tahun
- b. kenaikan PDRB perkapita sebesar satu satuan maka AHH akan cenderung berkurang, maksimal sebesar 10,903 tahun.

6. Pustaka

- [1] Arsyad, L. (1999). *Ekonomi Pembangunan*. Edisi Keempat. Yogyakarta: BP STIE YKPN.
- [2] Badan Pusat Statistik. (2014). *Indeks Pembangunan Manusia 2013*. Surabaya: Badan Pusat Statistik,
- [3] ----- (2015). *Sistem Informasi Rujukan Statistik*. <http://sirusa.bps.go.id/sirusa/index.php/indikator/74>.
- [4] Budiantara, I.N., Suryadi, F., Otok, B.W., dan Guritno, S. (2006). Pemodelan B-Spline dan MARS pada Nilai Ujian Masuk Terhadap IPK Mahasiswa Jurusan Disain Komunikasi Visual UK. Petra Surabaya. *Jurnal Teknik Industri*, Vol. 8, No. 1, hal. 1-13.
- [5] Darnius, O. (2002). *Aplikasi Software R dalam Analisis Regresi*. USU Digital Library. <https://s3.amazonaws.com/academia.edu/documents/49977582/matematika-Darnius.pdf>
- [6] Dinas Kesehatan Provinsi Jawa Timur (2013). *Profil Kesehatan Provinsi Jawa Timur Tahun 2012*. Surabaya: Dinas Kesehatan Provinsi Jawa Timur.
- [7] Direktorat Jenderal Perimbangan Keuangan (2012). *Tinjauan Ekonomi dan Keuangan Daerah Provinsi Jawa Timur*. Surabaya: Direktorat Jenderal Perimbangan Keuangan.
- [8] Djuraidah, A. dan Aunuddin (2006). Pendugaan Regresi Spline Terpenalti dengan Pendekatan Model Linear Campuran. *Statistika*, Vol. 6, No. 1, hal. 47-54.
- [9] Fathurahman, M. (2011). Estimasi Parameter Model Regresi Spline. *Jurnal Eksponensial*. Vol. 2, No. 1, hal. 53-58.
- [10] Sarvina, Y. (2017). Pemanfaatan Software Open Source R untuk Penelitian Agroklimat. *Informatika Pertanian*. Vol. 26. No. 1. Hlm: 23 - 30
- [11] Sugiantari, A.P. dan Budiantara, I.N. (2013). Analisis Faktor-Faktor yang Mempengaruhi Angka Harapan Hidup di Jawa Timur Menggunakan Regresi Semiparametrik Spline. *Jurnal Sains dan Seni Pomits*, Vol. 2, No. 1, hal. 37-41.
- [12] Suhartono. (2008). Analisis Data Statistik dengan R. Buku Ajar Program Studi Statistika, Jurusan Statistika, Institut Teknologi Sepuluh Nopember, Surabaya.
- [13] Tripena, A. (2011). Penentuan Model Regresi Spline Terbaik. *Prosiding Sewindu Statistika*. Universitas Diponegoro, Semarang, hal. 92-102.
- [14] Wang, Y. (2011). *Smoothing Splines Methods and Applications*. New York: CRC Press.
- [15] Yao, F. dan Lee, T.C.M. (2008). On Knot Placement for Penalized Spline Regression. *Journal of the Korean Statistical Society*, No. 37, hal. 259-267.

Aplikasi *Software R* pada Analisis Angka Harapan Hidup di Jawa Timur Menggunakan Estimator *P-Spline*