

BAB I

PENDAHULUAN

1.1 Latar Belakang

Analisis regresi merupakan salah satu analisis dalam statistika yang dipergunakan untuk menaksir pola hubungan antara variabel prediktor atau variabel bebas X dengan variabel respon atau variabel terikat atau variabel dependen Y. Hasil dari analisis regresi ini akan diperoleh suatu persamaan yang disebut dengan persamaan regresi. Terdapat dua pendekatan dalam analisis regresi, yaitu pendekatan regresi parametrik dan pendekatan regresi nonparametrik.

Pendekatan regresi parametrik digunakan jika bentuk kurva regresi diketahui, diperlukan pengetahuan masa lalu tentang karakteristik data yang akan diselidiki, dan data bersifat normal. Jika pola hubungan data membentuk pola linear, maka digunakan pendekatan regresi parametrik linear, jika pola hubungan data kuadrat maka digunakan pendekatan regresi kuadratik. Berbeda dengan pendekatan regresi parametrik, dalam regresi nonparametrik bentuk kurva regresi tidak diketahui, dan data cenderung tidak normal. Kurva regresi nonparametrik hanya diasumsikan *smooth* (mulus) dalam arti termuat di dalam satu ruang fungsi tertentu. Data diharapkan mencari sendiri bentuk taksirannya, tanpa dipengaruhi oleh faktor subyektifitas. Dengan demikian, pendekatan regresi nonparametrik memiliki fleksibilitas tinggi (Budiantara, 2005).

Selain untuk mengetahui pola hubungan antara variabel respon dan variabel prediktor, analisis regresi juga dapat dilakukan untuk peramalan (Wikipedia,2011). Penelitian dengan menggunakan pendekatan regresi sering dilakukan untuk menyelidiki hubungan antara satu atau banyak variabel prediktor dengan satu atau banyak variabel respon, dalam analisis regresi, penyelidikan tentang hubungan antara satu variabel prediktor dengan satu variabel respon disebut analisis regresi univariabel, sedangkan penyelidikan hubungan tentang satu atau lebih variabel prediktor dengan satu atau lebih variabel respon maka analisis regresi seperti ini dinamakan analisis regresi multivariabel (Budiantara, 2005).

Keadaan data yang tidak diketahui fungsi regresinya seringkali membuat para peneliti memutuskan untuk menyelesaikan penelitiannya atau menganalisis datanya kedalam analisis regresi nonparametrik. Regresi nonparametrik yang diawali dengan penelitian tentang histogram dan Kernel, selanjutnya memunculkan banyak model pendekatan yang dikembangkan. Semua model pendekatan regresi nonparametrik ini mempunyai cara dan latar belakang masing-masing dalam memperoleh taksiran kurva regresi f . Berkaitan dengan pengtaksiran tersebut banyak penaksir pada model regresi nonparametrik yang dapat digunakan untuk mengidentifikasi kurva regresi yang tidak diketahui serta membentuknya kedalam suatu model persamaan yang interpretatif dan kemudian lebih jauh lagi dari hasil tersebut dapat ditarik kesimpulan untuk suatu peramalan (Eubank,1988:3, Hardle, 1990:4).

Fungsi polinomial, seringkali digunakan untuk memodelkan data yang tidak diketahui bentuk kurva regresinya. Namun, terkadang kita menemukan ketidakcocokan suatu data jika dimodelkan ke dalam fungsi polinomial orde rendah, begitu pula dengan polinomial sederhana yang ditingkatkan ordenya, secara substansi ternyata tidak menjadikan data lebih baik. Masalah ini dapat diselesaikan ketika fungsi tersebut berperilaku lain pada suatu bagian tertentu yang berbeda dalam interval tertentu. Pendekatan yang biasa dilakukan untuk menyelesaikan permasalahan tersebut adalah dengan membagi suatu interval kedalam segmen-segmen dan mencocokkan setiap segmen kedalam kurva yang tepat. Fungsi spline merupakan fungsi yang digunakan untuk menyajikan data dengan permasalahan seperti tersebut di atas kedalam suatu bentuk atau model yang lebih baik.

Spline merupakan fungsi polinomial terpotong dalam orde k , yang mana di dalam fungsi tersebut terdapat titik-titik penghubung yang disebut dengan titik *knot*. Titik knot dapat diartikan sebagai suatu titik fokus dalam fungsi spline sehingga kurva yang dibentuk dapat terbagi pada titik tersebut. Orde dalam fungsi spline menunjukkan ketinggian derajat polinom fungsi tersebut. Titik knot dan orde inilah yang kemudian akan dipakai untuk menentukan model regresi splinenya. Spline dapat menggambarkan perubahan pola perilaku dan fungsi pada sub interval, mengatasi pola data yang menunjukkan naik atau turun yang tajam dengan bantuan titik-titik knot serta kurva yang dihasilkan oleh regresi spline ini relatif mulus (Eubank, 1988).

Fungsi spline dapat berupa spline linear, spline kuadratik, spline kubik, maupun spline derajat k , dan fungsi spline dengan jenis spline kubik, merupakan fungsi spline yang seringkali digunakan untuk penelitian. Spline merupakan jumlahan dari fungsi polinomial dengan suatu fungsi (truncated) (Hardle, 1990) dan fungsi spline dengan satu variabel prediktor didefinisikan sebagai berikut:

Regresi nonparametrik yang kurva regresinya hanya diasumsikan termuat dalam suatu ruang fungsi tertentu, jika didekati dengan fungsi spline dengan satu variabel prediktor didefinisikan sebagai berikut:

$$f(x) = \sum_{j=0}^q \alpha_j X^j + \sum_{k=1}^m \beta_k (X - K_k)_+^q$$

dimana,

$$(X - K_k)_+^q = \begin{cases} (X - K_k)_+^q & ; X \geq K_j \\ 0 & ; X < K_j \end{cases}, \text{ merupakan fungsi } \textit{truncated},$$

α_j : koefisien pengganda nilai X^j terhadap variabel respon y .

X^j : variabel prediktor ke- j

β_k : koefisien pada variabel X^j *truncated* knot ke- k pada spline ber-orde m

K_k : knot ke- k pada variabel X^j .

m : banyaknya knot dalam variabel prediktor ke- j

Spline mempunyai kelemahan pada saat orde tinggi, knot yang banyak dan knot yang terlalu dekat, yang mana salah satu dari ketiga hal tersebut

menyebabkan terbentuknya matriks dalam perhitungan yang hampir singular, sehingga persamaan tidak dapat diselesaikan.

Penggunaan spline difokuskan pada adanya perilaku atau pola data yang pada daerah tertentu mempunyai karakteristik yang berbeda dengan daerah yang lain. Spline, berdasarkan banyaknya variabel prediktor dan variabel respon dibagi kedalam dua bagian, yaitu spline univariabel dan spline multivariabel. Pada spline univariabel variabel prediktor dan variabel respon masing-masing hanya ada satu variabel saja, sedangkan dalam spline multivariabel baik variabel prediktor maupun variabel respon dapat berjumlah lebih dari satu.

Tiga kriteria yang harus diperhatikan dalam membentuk model regresi Spline yaitu:

1. menentukan orde untuk model
2. menentukan banyaknya knot
3. menentukan lokasi penempatan knot.

Orde untuk model dapat ditentukan berdasarkan pola umum yang terjadi pada data, sedangkan banyaknya knot dan lokasi knot ditentukan berdasarkan perubahan pola di daerah tertentu pada kurva. Karena keakuratannya dalam memodelkan data, jenis spline yang seringkali digunakan untuk penelitian adalah spline kubik. Nilai GCV (*Generalized Cross Validation*) merupakan nilai yang selanjutnya dicari setelah knot dan orde fungsi ditentukan, nilai GCV optimum inilah yang kemudian akan dipakai untuk menentukan model yang paling interpretatif untuk memodelkan data.

Pembahasan mengenai regresi spline ini membuat penulis tertarik dan memutuskan untuk mengkajinya lebih lanjut, serta mengaplikasikan modelnya pada sekumpulan data nilai UN, sehingga penulis mengangkat bahasan tersebut kedalam tugas akhir yang kemudian diberi judul “Aplikasi Metode Regresi Spline dalam Memodelkan Nilai UN” .

1.2 Rumusan Masalah

Dari uraian latar belakang diperoleh rumusan masalah yang diangkat adalah:

1. Bagaimana cara memperoleh model regresi spline univariabel dan spline multivariabel?
2. Model regresi spline apa yang paling cocok untuk memodelkan nilai UN SMPN 1 Lembang?

1.3 Batasan Masalah

Masalah yang dibahas dalam penulisan ini dibatasi pada cara memperoleh model regresi spline univariabel dan model regresi spline multivariabel, penerapan regresi spline pada data primer nilai rata-rata UN SMP N 1 Lembang, orde yang digunakan untuk memodelkan nilai rata-rata UN siswa SMP ini hanya dibatasi sampai orde ketiga, sedangkan untuk knotnya dibatasi sampai dua knot, hal tersebut ditujukan untuk mempermudah perhitungan dan untuk kesederhanaan model.

1.4 Tujuan Penelitian

Tujuan dari penulisan tugas akhir ini adalah:

1. Memperoleh model regresi spline univariabel dan multivariabel.
2. Memilih model yang paling cocok untuk data nilai UN dan menentukan faktor apa saja yang berpengaruh terhadap nilai rata-rata UN SMP.

1.5 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah sebagai sarana berpikir kritis dalam melihat fenomena yang berkembang di masyarakat dan mengkaitkannya dengan ilmu statistika yang telah dipelajari. Selain itu, dengan mempelajarinya diperoleh pendalaman mengenai analisis regresi dengan asumsi model nonparametrik, dalam hal ini adalah model regresi spline univariabel, model regresi spline multivariabel, penerapan model terhadap data-data yang berkaitan. Setelah diperoleh model yang paling tepat untuk data pengaturan strategi pengambilan keputusan khususnya penentuan faktor yang mempengaruhi nilai UN SMP menjadi lebih mudah.