

BAB I PENDAHULUAN

Bab pendahuluan ini berisi uraian tentang latar belakang, rumusan masalah, tujuan penelitian, dan manfaat penelitian. Pada latar belakang, diuraikan tentang regresi berdasarkan tipe variabel respon. Penelitian-penelitian sebelumnya yang berkaitan dengan regresi bertipe variabel respon diskrit menggunakan pendekatan regresi parametrik maupun regresi nonparametrik juga disebutkan. Selanjutnya, diuraikan juga tentang estimator lokal linier pada regresi nonparametrik, rumusan masalah dan tujuan penelitian. Pada bagian akhir Bab disebutkan manfaat yang diharapkan dalam penelitian ini.

1.1 Latar Belakang

Analisis regresi merupakan salah satu prosedur dalam statistika yang digunakan untuk menggambarkan hubungan fungsional antara variabel respon dengan variabel prediktor. Estimasi fungsi pada regresi dapat dilakukan melalui dua pendekatan yaitu pendekatan parametrik dan nonparametrik. Apabila bentuk fungsi regresi diketahui dan hubungan antara variabel respon dengan variabel prediktor diasumsikan mengikuti bentuk kurva tertentu antara lain linier, kuadratik, dan eksponensial maka yang diestimasi adalah parameter dalam fungsi regresi. Realitas di lapangan menunjukkan bahwa ada hubungan antara variabel respon dengan variabel prediktor yang tidak mengikuti pola tertentu atau tidak ada informasi apapun tentang bentuk dari fungsi regresi. Pendekatan yang cocok pada kondisi tersebut adalah pendekatan nonparametrik. Pendekatan regresi nonparametrik memiliki fleksibilitas yang tinggi karena fungsi regresi tidak dispesifikasikan tertentu bentuknya tetapi diasumsikan *smooth* sehingga diestimasi dengan menggunakan metode *smoothing* tertentu berdasarkan pola data (Eubank, 1999).

Berdasarkan tipe dari variabel respon, terdapat berbagai macam model regresi. Model regresi yang umum digunakan untuk menganalisis hubungan antar variabel respon diskrit dengan satu atau beberapa variabel prediktor dalam *Generalized Linear Model* (GLM) adalah model regresi poisson. Regresi poisson diterapkan pada kasus dimana variabel respon berdistribusi poisson. Menurut Ross (2004), distribusi poisson merupakan banyaknya sukses yang terjadi pada sampel yang besar dengan peluang yang kecil, seperti banyaknya penggunaan telepon per hari, banyaknya kesalahan ketik per halaman sebuah buku, dan lain sebagainya. Variabel respon (Y) dalam model regresi poisson harus memenuhi asumsi $\text{var}(Y) = E(Y)$ yang disebut *equi dispersion*. Pada kenyataannya, terkadang dalam model

regresi poisson terjadi pelanggaran asumsi tersebut, dimana $\text{var}(Y) > E(Y)$ yang disebut *over dispersion* atau $\text{var}(Y) < E(Y)$ yang disebut dengan *under dispersion*. Penggunaan regresi poisson yang tidak tepat karena adanya masalah *over dispersion* maupun *under dispersion*, dapat mengakibatkan kesalahan pada inferensi bagi parameter model (Payne *et al.*, 2017).

Untuk menghindari masalah *over dispersion* pada model regresi poisson, dalam model regresi parametrik telah berkembang pesat model regresi yang tidak mengasumsikan distribusi poisson pada variabel respon. Beberapa penelaitan tersebut antara lain: Famoye *et al.* (2004) menggunakan *generalized poisson* untuk diaplikasikan pada data kecelakaan; Chang (2005); Pradhan and Leung (2006); Ferrari *et al.* (2007); Greene (2008) dan Månsson (2012) mengembangkan regresi binomial negatif. Pada kasus birespon, Famoye (2010a) membahas tentang model *bivariate generalized poisson* dengan estimasi momen dan *maximum likelihood*; Famoye (2010b) dan Tohari *et al.* (2019) membahas tentang model *bivariate negative binomial regression* menggunakan *maximum likelihood*.

Model regresi dengan tipe variabel respon diskrit yang telah disebutkan di atas merupakan model regresi parametrik yang mengasumsikan hubungan antara variabel respon dengan variabel prediktor mengikuti bentuk tertentu seperti linier, kuadratik atau eksponensial. Realitas di lapangan menunjukkan bahwa ada hubungan antar variabel respon dengan variabel prediktor yang tidak menunjukkan pola tertentu sehingga pendekatan parametrik sulit diaplikasikan. Pada kondisi seperti ini, terdapat model regresi nonparametrik yang tidak mengasumsikan bentuk kurva regresi. Model regresi nonparametrik secara umum bertujuan meminimalkan asumsi dari bentuk kurva regresi dan membiarkan data mencari bentuk estimasi sesuai dengan pola yang dimiliki (Wang *et al.*, 2011).

Dalam regresi nonparametrik, beberapa metode *smoothing* yang sering digunakan antara lain metode Polinomial Lokal (Fan and Gijbels, 1996; Doksum *et al.*, 2000; Chamidah *et al.*, 2012; De Brabanter *et al.*, 2013; dan Chamidah *et al.*, 2019), metode Kernel (Cao and Lugosi, 2005; Tenreiro, 2013; Chamidah and Saifudin, 2013; Wang *et al.*, 2017), dan metode Spline (Lestari *et al.*, 2010; Lestari *et al.*, 2012; Wang *et al.*, 2011; Chamidah and Lestari, 2016; Lestari *et al.*, 2018, dan Lestari *et al.*, 2018). Teknik-teknik ini dikenal dengan nama *local fitting*, karena estimasi fungsi dilakukan secara lokal terhadap nilai-nilai di sekitar interval nilai tertentu. Estimator polinomial lokal merupakan suatu pendekatan baru yang populer dan mempunyai bentuk khusus yaitu jika derajat polinomial sama dengan nol

disebut estimator kernel dan derajat polinomial sama dengan satu disebut estimator lokal linier. Peneliti lain yang membahas tentang lokal linier adalah Fan and Gijbels (1992), Fan (1993), Chamidah and Rifada (2016), Jin and He (2016), Chamidah *et al.* (2018).

Model regresi nonparametrik dengan variabel respon bertipe diskrit belum banyak dikembangkan oleh para peneliti. Klimov *et al.* (2002) menggunakan estimator kernel pada regresi poisson, Young *et al.* (2006) menggunakan estimator spline pada regresi poisson untuk memodelkan hubungan antara status sosial ekonomi dengan angka kematian, Santos and Neves (2008) membahas tentang regresi poisson dengan estimator lokal linier. Demikian juga penelitian yang dilakukan oleh Shim and Hwang (2011) membahas regresi poisson dengan estimator kernel untuk klaim pemesanan mesin, dan Darnah *et al.* (2019) membahas tentang regresi poisson menggunakan estimator lokal linier. Sedangkan Astuti *et al.* (2013) menggunakan estimator polinomial lokal pada regresi generalized poisson untuk memodelkan angka kematian bayi berdasarkan usia.

Berdasarkan uraian di atas, sampai saat ini belum ada pengembangan teori estimasi untuk regresi binomial negatif dengan menggunakan pendekatan estimator lokal linier. Pendekatan lokal linier baik digunakan pada prinsip parsimoni model sehingga mudah diinterpretasikan. Selain itu, setiap fungsi yang *differensiable* dapat didekati secara lokal dengan suatu garis lurus (Loader, 1999). Estimator lokal linier merupakan kasus khusus dari estimator polinomial lokal. Teknik polinomial lokal digunakan karena dapat mengatasi secara lebih baik perubahan pola data dengan osilasi yang lebih cepat dan tajam dibandingkan dengan teknik lain, selain dapat menghasilkan estimasi kurva regresi, juga dapat menghasilkan estimasi dari turunan kurva regresi, memiliki bias pada daerah batas yang lebih baik dibandingkan dengan teknik kernel (Fan and Gijbels, 1996).

Pada penelitian ini akan dikaji secara teori tentang estimasi regresi binomial negatif menggunakan pendekatan lokal linier. Analisis regresi binomial negatif yang melibatkan satu variabel respon dengan satu atau lebih variabel prediktor disebut regresi binomial negatif unirespon, sedangkan regresi binomial negatif yang melibatkan dua variabel respon yang saling berkorelasi dengan satu atau lebih variabel prediktor disebut regresi binomial negatif birespon. Perluasan dari regresi binomial negatif birespon adalah regresi binomial negatif multirespon, yaitu regresi binomial negatif yang melibatkan lebih dari dua variabel respon yang saling berkorelasi dengan satu atau lebih variabel prediktor.

Dalam penelitian ini, selain aspek pengembangan teori juga dikembangkan aspek aplikasi yaitu dengan membuat program yang terkait dengan teori yang telah dikembangkan menggunakan software R. Program yang akan dikembangkan dapat diimplementasikan pada

data empiris. Implementasi program pada data empiris bertujuan untuk memodelkan data menggunakan regresi binomial negatif birespon dengan pendekatan nonparametrik berdasarkan estimator lokal linier.

Implementasi pada data empiris dilakukan pada bidang kesehatan. Kesehatan merupakan salah satu hal penting dalam investasi sumber daya manusia, sehingga peran pemerintah sangat tepat dalam upaya pembangunan di bidang kesehatan secara berkelanjutan. Pada tahun 2000, secara resmi digagas *Millenium Development Goals* (MDGs) yang bertujuan untuk mencapai kesejahteraan rakyat dan pembangunan masyarakat pada tahun 2015, salah satu agenda pencapaian MDGs adalah memerangi *human immunodeficiency virus* (HIV) dan *Acquired Immune Deficiency Syndrome* (AIDS) (Ngoya, 2015). Pada tahun 2017, wilayah Asia dan Pasifik dikenal sebagai rumah bagi sekitar 5,2 juta orang yang hidup dengan HIV, dan Indonesia menyumbang sekitar 12% dari jumlah tersebut (UNAIDS, 2018). Sampai dengan bulan Juni 2018, ada 433 (84,2%) dari 514 kabupaten/kota di 34 provinsi di Indonesia yang melaporkan HIV/AIDS. Kasus HIV yang dilaporkan pada Juni 2018 memiliki jumlah kumulatif 301.959 orang (pada tahun 2018, 47% orang dengan HIV/AIDS atau disebut dengan ODHA adalah 640.443 orang) dan paling sering ditemukan pada kelompok usia 25-49 tahun dan 20-24 tahun. Adapun provinsi dengan banyak kasus HIV tertinggi adalah DKI Jakarta (55.099), diikuti oleh Jawa Timur (43.399), Jawa Barat (31.293), Papua (30.699), dan Jawa Tengah (24.757) (Kemenkes, 2018b).

Selain masalah HIV dan AIDS yang menjadi fokus penyelesaian pemerintah Indonesia dan dunia secara global, pada akhir tahun 2019 muncul kasus *Corona Virus Disease* (COVID-19) yang ditetapkan sebagai pandemi oleh *World Health Organization* (WHO) pada 11 Maret 2020. Kasus pandemi COVID-19 memiliki potensi ancaman pada segala aspek kehidupan masyarakat, seperti kesehatan, pendidikan, sosial, ekonomi, dan psikologis. Pandemi COVID-19 di Indonesia juga belum mereda, sampai dengan 31 Juli 2020 banyak kasus positif COVID-19 telah menembus angka 108.376 kasus. Provinsi Jawa Timur merupakan provinsi penyumbang paling banyak kasus COVID-19 di Indonesia dengan 22.089 kasus dan kasus suspek sebanyak 6425 kasus (<http://infocovid19.jatimprov.go.id>).

Data banyak kasus HIV dan AIDS, serta data banyak kasus positif dan suspek COVID-19 merupakan data *count*, dimana penelitian-penelitian pada data *count* sebelumnya dilakukan dengan menggunakan asumsi distribusi poisson baik pada regresi parametrik maupun nonparametrik. Berdasarkan studi awal, diketahui bahwa data banyak kasus HIV dan AIDS di Jawa Timur tahun 2016 (Tohari *et al.*, 2020c), data banyak kasus HIV dan AIDS di Indonesia tahun 2017 (Tohari *et al.*, 2020a), serta data banyak kasus positif dan

suspek COVID-19 di Jawa Timur (Tohari *et al.*, 2020d) memiliki variansi yang tinggi, sehingga akan menimbulkan masalah *over dispersion* jika digunakan model regresi dengan asumsi distribusi poisson. Berdasarkan kondisi tersebut, maka dalam penelitian ini data banyak kasus HIV dan AIDS, data jumlah kasus positif dan suspek COVID-19 akan dianalisis menggunakan regresi binomial negatif birespon. Kelemahan penggunaan pendekatan parametrik adalah estimasi kurva regresi yang dihasilkan bersifat global, sedangkan estimator lokal linier akan menghasilkan kurva regresi secara lokal di masing-masing provinsi untuk data Indonesia, dan secara lokal di masing-masing kabupaten/kota untuk data Jawa Timur, sehingga estimator lokal linier merupakan pendekatan yang tepat dalam memodelkan data banyak kasus HIV dan AIDS di Jawa Timur, data banyak kasus HIV dan AIDS di Indonesia, dan data banyak kasus positif dan suspek COVID-19 di Jawa Timur.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang di atas, maka masalah dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana mengestimasi model regresi binomial negatif birespon menggunakan pendekatan estimator lokal linier?
2. Bagaimana membuat algoritma dan program untuk mengestimasi model regresi binomial negatif birespon menggunakan pendekatan estimator lokal linier?
3. Bagaimana implementasi model regresi binomial negatif birespon menggunakan pendekatan estimator lokal linier pada data banyak kasus HIV dan AIDS di Jawa Timur dan di Indonesia?
4. Bagaimana implementasi model regresi binomial negatif birespon menggunakan pendekatan estimator lokal linier pada data banyak kasus positif dan suspek COVID-19 di Jawa Timur?

1.3 Tujuan Penelitian

Berdasarkan rumusan permasalahan di atas, maka tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Mengestimasi model regresi Binomial Negatif Birespon menggunakan pendekatan estimator lokal linier.
2. Membuat algoritma dan program untuk mengestimasi model regresi binomial negatif birespon menggunakan pendekatan estimator lokal linier.

3. Mengimplementasikan model regresi binomial negatif birespon menggunakan pendekatan estimator lokal linier pada data banyak kasus HIV dan AIDS di Jawa Timur dan di Indonesia.
4. Mengimplementasikan model regresi binomial negatif birespon menggunakan pendekatan estimator lokal linier pada data banyak kasus positif dan suspek COVID-19 di Jawa Timur.

1.4 Manfaat Penelitian

Hasil dari penelitian ini diharapkan dapat memberikan manfaat, antara lain:

1. Memberikan kontribusi dalam pengembangan ilmu statistika yaitu estimasi model regresi binomial negatif birespon berdasarkan estimator lokal linier.
2. Algoritma dan program yang diperoleh dapat dimanfaatkan untuk aplikasi pada data empiris yaitu data HIV dan AIDS di Jawa Timur dan di Indonesia serta data COVID-19 di Jawa Timur berdasarkan estimator lokal linier.
3. Mengetahui seberapa besar faktor-faktor yang diduga mempengaruhi HIV dan AIDS di Jawa Timur dan di Indonesia secara lokal.
4. Mengetahui seberapa besar faktor-faktor yang diduga mempengaruhi kasus positif dan suspek COVID-19 di Jawa Timur secara lokal.