

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam berbagai penelitian, dapat dijumpai adanya permasalahan mengenai hubungan ketergantungan antara peubah respon dengan peubah prediktor. Analisis mengenai hubungan ketergantungan tersebut dapat dimodelkan dengan analisis regresi. Analisis regresi linier mengasumsikan bahwa peubah respon merupakan peubah kontinyu yang mengikuti sebaran normal. Namun pada aplikasinya sering dijumpai peubah respon berupa data diskrit (*count*) dan tidak mengikuti sebaran normal. Menurut Agresti (2002), untuk mengatasi hal tersebut terdapat pengembangan dalam model linier klasik yaitu *Generalized Linear Model* (GLM). GLM mengasumsikan peubah respon mengikuti sebaran keluarga eksponensial. Adapun sebaran yang termasuk dalam keluarga eksponensial adalah sebaran *Poisson*, Binomial, Eksponensial dan *Gamma*.

Salah satu model regresi yang dapat digunakan untuk memodelkan hubungan antara peubah respon yang berupa data diskrit dengan peubah prediktor berupa data diskrit, kontinyu, kategorik atau campuran adalah model regresi *Poisson*. Regresi *Poisson* merupakan analisis regresi bagian dari *Generalized Linear Model* (GLM) yang digunakan untuk data diskrit atau *count*. Menurut Ismail dan Jemain (2005), asumsi penting pada analisis regresi *Poisson* adalah nilai ragam harus sama dengan nilai rata-rata yang disebut dengan *equidispersion*. Akan tetapi, dalam penerapannya sering dijumpai adanya *overdispersion* dan *underdispersion*. *Overdispersion* merupakan suatu kejadian di mana nilai ragam peubah respon lebih besar daripada nilai rata-rata peubah respon tersebut. Sebaliknya, jika nilai ragam lebih kecil dari nilai rata-rata maka dikatakan *underdispersion* (Famoye *et al*, 2004). Adapun tidak terpenuhi kondisi *equidispersion* pada regresi *Poisson* dapat menyebabkan model yang terbentuk menghasilkan penduga parameter yang bias, nilai penduga kesalahan baku yang *underestimates*, dan kesalahan dalam pendugaan parameter. Penanganan *overdispersion* dan *underdispersion* pada regresi *Poisson* dapat menggunakan regresi *Generalized Poisson* (GP), regresi *Negative Binomial* (NB), regresi *Zero Inflated Poisson* (ZIP),

regresi *Zero Inflated Generalized Poisson* (ZIGP), dan regresi *Zero Inflated Negative Binomial* (ZINB).

Seringkali pada peubah respon data jenis diskrit atau *count* ditemukan amatan yang bernilai nol dengan proporsi nilai nol yang besar (*zero inflation*). Regresi *Generalized Poisson* (GP) dan regresi *Negative Binomial* (NB) dapat mengatasi masalah *overdispersion* dan *underdispersion* namun masih belum mengatasi masalah nilai nol berlebih (*zero inflation*) pada peubah respon. Menurut Famoye dan Singh (2006), dalam penelitian dapat dijumpai kondisi di mana terlalu banyak nol pada peubah respon, yaitu lebih dari 50 persen. Besarnya proporsi data yang bernilai nol dapat berakibat pada ketepatan (presisi) dari inferensia. Selain itu, regresi *Poisson* juga menjadi tidak tepat lagi memodelkan data yang sebenarnya.

Regresi *Zero Inflated Poisson* (ZIP) merupakan analisis regresi untuk data *count* di mana pada peubah respon banyak terdapat amatan bernilai nol (*zero inflation*). Lambert (1992) pertama kali memperkenalkan model regresi *Zero Inflated Poisson* (ZIP) untuk pemodelan data dengan mengaplikasikan model tersebut pada data yang dikumpulkan dari sebuah studi *Quality Control*, di mana peubah respon yang digunakan adalah banyaknya produk cacat dalam unit sampel (Famoye dan Singh, 2006). Ridout *et al* (1998) mengaplikasikan model ZIP pada data biologi yaitu data perkembangbiakan tunas apel akan tetapi model ini tidak cocok karena adanya *overdispersion*, sehingga pada akhirnya beralih ke model *Zero Inflated Negative Binomial* (ZINB). Jika pada suatu pemodelan data *count* banyak terdapat amatan yang bernilai nol pada peubah respon dan terjadi *overdispersion* atau *underdispersion* maka dapat diatasi dengan menggunakan regresi *Zero Inflated Generalized Poisson* (ZIGP) dan *Zero Inflated Negative Binomial* (ZINB).

1.2 Rumusan Masalah

Model manakah yang lebih baik antara model regresi *Zero Inflated Generalized Poisson* (ZIGP) dan *Zero Inflated Negative Binomial* (ZINB) untuk data *overdispersion*.

1.3 Tujuan Penelitian

Penelitian ini dilakukan dengan tujuan:

1. Membentuk model regresi *Zero Inflated Generalized Poisson* (ZIGP).
2. Membentuk model regresi *Zero Inflated Negative Binomial* (ZINB).
3. Membandingkan model regresi *Zero Inflated Generalized Poisson* (ZIGP) dan *Zero Inflated Negative Binomial* (ZINB) untuk mengetahui model yang lebih sesuai digunakan pada data *overdispersion* dengan menggunakan nilai *Akaike Information Criterion* (AIC).

1.4 Batasan Masalah

Data yang digunakan adalah tiga data sekunder yang *overdispersion* dan *zero inflation* pada peubah respon.

1.5 Manfaat Penelitian

Manfaat dari penulisan skripsi ini adalah dapat mengetahui model yang lebih baik antara model regresi *Zero Inflated Generalized Poisson* (ZIGP) atau *Zero Inflated Negative Binomial* (ZINB) jika digunakan pada data *overdispersion*.