

BAB I

PENDAHULUAN

1.1 Latar Belakang

Data cacahan (count data) adalah data yang menggambarkan sejumlah kejadian yang terjadi pada suatu kurun waktu tertentu. Suatu data cacahan hanya dapat berupa bilangan bulat positif karena suatu kejadian tidak mungkin terjadi dalam sejumlah bilangan negatif. Pemodelan data cacahan banyak digunakan dalam berbagai ilmu seperti sosiologi, kedokteran, aktuaria dan demografi. Pemodelan data cacahan tidak dapat dilakukan dengan menggunakan regresi OLS (Ordinary Least Square), karena pemodelan data cacahan akan melanggar dua asumsi yang disyaratkan dalam regresi OLS yaitu, error mengikuti distribusi normal (normalitas) dan memiliki sifat homokedastisitas (varians konstan). Distribusi bersyarat dari data cacahan cenderung menceng kanan dengan tidak ada kejadian yang di bawah nol dan jumlah kejadian dengan frekuensi tinggi hanya sedikit sehingga melanggar asumsi kenormalan. Pelanggaran asumsi varians konstan terjadi karena dalam data cacahan sering terjadi peningkatan varians bersyarat sebagai akibat peningkatan nilai prediktor. Inilah yang disebut sebagai heterokedastisitas yang dapat mengakibatkan standar error dan uji signifikansi yang bias bila menerapkan regresi OLS (Coxe, West dan Aiken, 2009).

Pengembangan dalam pemodelan data cacahan memunculkan pemodelan data cacahan dengan Generalized Linear Models (GLMs). GLMs merupakan generalisasi dari model regresi normal klasik atau regresi OLS dari asumsi yang

ketat dan menyediakan metode analisis bagi data tidak normal (DeJong dan Heller, 2008). Regresi Poisson adalah salah satu anggota keluarga dari GLMs yang berasal dari distribusi poisson. Distribusi poisson merupakan distribusi diskrit dengan nilai variabel random berupa bilangan bulat positif sehingga menjadi pilihan yang baik untuk pemodelan data cacahan. Distribusi poisson hanya ditentukan oleh satu parameter yang mendefinisikan baik mean maupun varians dari distribusi tersebut, sehingga dalam regresi Poisson terdapat asumsi yang harus terpenuhi yaitu *mean* dan varians variabel respon harus sama (*equidispersion*). Namun dalam kenyataannya sering terjadi pelanggaran asumsi tersebut dimana varians lebih kecil dari *mean* (*underdispersion*) atau varians lebih besardari *mean* (*overdispersion*). Pada kebanyakan data *count* terkadang ditemukan kasus overdispersi (Consul dan Famoye, 1992).

Kasus overdispersi bila diabaikan bisa mengakibatkan terjadinya *underestimate* pada estimasi standar error, sehingga dapat mengakibatkan kesalahan pada pengambilan keputusan beberapa uji hipotesis, misalnya suatu variabel prediktor berpengaruh signifikan ketika pada kenyataan tidak berpengaruh signifikan (Hilbe, 2007). Dalam mengatasi kasus overdispersi, dibentuk beberapa pemodelan yang merupakan perpaduan antara distribusi Poisson dengan beberapa distribusi baik diskrit maupun kontinu (*mixed poisson distribution*). *Mixed poisson distribution* tersebut merupakan solusi alternatif untuk kasus overdispersi, tetapi hanya beberapa distribusi yang sering digunakan dalam penelitian dikarenakan penghitungannya yang rumit. Salah satunya adalah distribusi Poisson Invers Gaussian (PIG) yang merupakan *mixed poisson distribution* dengan random efek

yang memiliki distribusi Invers Gaussian. Distribusi ini pertama kali diperkenalkan oleh Holla pada tahun 1966 (Karlis dan Nikoloulopoulos, 2005). Distribusi PIG sendiri merupakan bentuk dari distribusi *Siche* (*SI*) dengan dua parameter. *SI* disebut-sebut sebagai model yang lebih baik dari model binomial negatif, terutama untuk data yang overdispersi yang tinggi dan cenderung menceng kanan (*highly skewed to the right*). Namun penghitungannya lebih rumit karena memiliki tiga parameter pada fungsi kepadatan peluangnya. Sebagai bentuk dari distribusi *SI* adalah distribusi PIG yang digunakan dalam memodelkan data cacahan yang menceng kanan serta memiliki ekor yang sedikit lebih panjang. Akan tetapi, distribusi PIG memiliki bentuk fungsi likelihood yang *close form* dan penghitungannya lebih mudah sehingga banyak penelitian yang melibatkan data cacahan banyak yang menggunakan model ini (Stasinopoulus dan Rigby, 2007).

Willmot (1987) menunjukkan potensi dari pemodelan dengan regresi poisson inverse gaussian sebagai alternatif dari regresi binomial negatif pada data klaim asuransi mobil. Enam set data klaim asuransi mobil disajikan dengan karakteristik hampir 80 persen data yang mengandung nol, dan menghasilkan kesimpulan bahwa pemodelan dengan regresi PIG merupakan model yang lebih baik dibandingkan model regresi binomial negatif. Penelitian lainnya yang menggunakan model regresi PIG adalah penelitian dari Shoukri, Asyali dan Vandorp (2004) yang menggunakan model tersebut pada data jumlah kasus penyakit mastitis pada sapi perah di Ontario Canada dan menghasilkan kesimpulan bahwa regresi PIG merupakan model yang lebih baik dari model regresi binomial negatif. Selanjutnya dalam beberapa penelitian keselamatan di jalan raya seperti

pemodelan data kecelakaan dan penelitian asuransi motor, model regresi PIG sering digunakan sebagai alternatif dari model regresi binomial negatif (Denis, Ghitany dan Gupta, 2014). Pemodelan dengan regresi PIG juga digunakan pada penelitian Zha, Lord dan Zou (2014) pada kasus jumlah kecelakaan motor yang terjadi di dua tempat berbeda yaitu di Texas dan Washington. Dalam penelitian tersebut data jumlah kecelakaan motor menceng kanan dengan ekor yang sedikit panjang serta 37 persen data mengandung nilai nol. Dengan nilai *Akaike Information Criterion* (AIC) dan *Bayesian Information Criterion* (BIC) yang lebih kecil, menunjukkan bahwa pemodelan dengan regresi PIG lebih baik untuk memodelkan kasus jumlah kecelakaan motor di dua tempat tersebut.

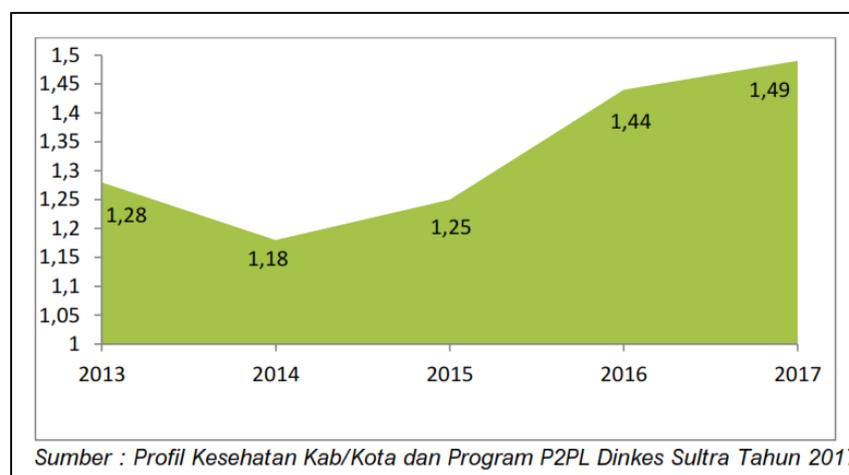
Widiari (2016) menggunakan model regresi Poisson Inverse Gaussian pada data jumlah kasus baru HIV di Jawa Timur Tahun 2013 dan menghasilkan variabel yang signifikan berpengaruh adalah persentase penduduk dengan pendidikan terakhir SLTA, persentase PUS yang menggunakan kondom, dan rasio fasilitas kesehatan. Kemudian Herindrawati, Latra dan Puhadi (2017) juga melakukan penelitian dengan menggunakan regresi Poisson Inverse Gaussian pada data kasus baru penyakit HIV di Provinsi Jawa Tengah Tahun 2015. Dalam penelitiannya diperoleh model regresi Poisson Invers Gaussian dan faktor-faktor yang berpengaruh secara signifikan terhadap kasus baru penyakit HIV di Provinsi Jawa Tengah yaitu persentase PUS yang menggunakan kondom, rasio fasilitas kesehatan, persentase daerah perkotaan, dan persentase penduduk usia 25-34 tahun.

Pada penelitian sebelumnya juga yang mengambil kasus penyakit kusta dengan beberapa judul yaitu faktor-faktor yang berhubungan dengan derajat

kecacatan pasien *morbus Hansen* (Kusta) di kota Bandar Lampung, faktor-faktor yang berhubungan dengan kejadian kusta di wilayah kerja puskesmas kabunan kabupaten pemalang tahun 2014, analisis faktor risiko tingkat kecacatan pada penderita kusta di puskesmas Padas kabupaten Ngawi pada tahun 2010, sebaran kasus penyakit kusta berdasarkan wilayah kecamatan dikota Bekasi tahun 2006-2008.

Selain itu penyakit Kusta dikenal juga sebagai penyakit Lepra atau penyakit Hansen, disebabkan oleh bakteri *Mycobacterium leprae*. Bakteri ini mengalami proses pembelahan cukup lama antara 2-3 minggu. Daya tahan hidup kuman kusta mencapai 9 hari di luar tubuh manusia. Kuman kusta memiliki masa inkubasi 2-5 tahun bahkan lebih.

Pada Tahun 2017, jumlah kasus kusta baru di Sulawesi Tenggara sebesar 327 kasus dengan angka penemuan kasus baru (*New Case Detection Rate/NCDR*) per 100.000 penduduk sebesar 12,40, lebih rendah dari tahun 2016 (342 kasus baru / *NCDR*; 13,20).



Gambar 1.1 Angka Prevalensi Kusta Per 10.000 Penduduk Provinsi Sulawesi Tenggara Tahun 2013-2017

Berdasarkan grafik di atas bahwa pada tahun 2013-2017, angka prevalensi kusta Sulawesi Tenggara sempat turun mencapai angka terendah pada tahun 2014, tetapi kembali naik pada tiga tahun terakhir, bila merujuk pada target nasional yang ditetapkan mengenai prevalensi kusta sebesar kurang dari 1/10.000 penduduk (lebih dari 10/100.000), maka prevalensi kusta Sulawesi Tenggara sebesar 1.49/10.000 penduduk makin menjauh dari target, setelah hampir mencapai target pada tahun 2014. Tampaknya upaya penanggulangan kusta masih membutuhkan upaya yang lebih keras beberapa tahun ke depan agar angka prevalensi kusta Sulawesi Tenggara bisa mencapai target nasional.

Pada penelitian ini yang akan diteliti yaitu jumlah penyakit kusta di Sulawesi Tenggara. Dimana jumlah penyakit kusta merupakan salah satu indikator penting bagi pemerintah dalam mengevaluasi dibidang kesehatan. Dengan mengetahui faktor-faktor yang mempengaruhi, tentu saja pemerintah memiliki gambaran langkah apa saja yang harus dilakukan dalam menekan jumlah penyakit kusta. Mengingat dalam tahun terakhir yaitu Pada Tahun 2017, jumlah kasus kusta di Sulawesi Tenggara sebesar 327 kasus dengan angka penemuan kasus per 100.000 penduduk sebesar 12,40, lebih rendah dari tahun 2016 (342 kasus baru / NCDR; 13,20). Berdasarkan prevalensi, kasus kusta di Sulawesi Tenggara menunjukkan kurva menaik pada rentang tahun 2013-2017, dengan prevalensi tertinggi terjadi pada tahun 2017 sebesar 1,49.

Data penyakit kusta merupakan data cacahan sehingga dalam pemodelannya bisa menggunakan regresi Poisson. Data penyakit kusta juga berpotensi mengalami *overdispersi* sehingga dalam penanganannya diperlukan sebuah model regresi yang dapat dilakukan untuk data yang mengalami *overdispersi*. Berdasarkan penelitian yang telah dilakukan oleh beberapa peneliti, maka model regresi yang akan diterapkan pada penelitian ini yaitu model regresi Poisson Inverse Gaussian.

1.2 Rumusan Masalah

Berdasarkan latar belakang dapat diambil perumusan masalah sebagai berikut.

1. Bagaimana pola penyebaran jumlah penyakit kusta di Sulawesi Tenggara menggunakan Peta Tematik?
2. Bagaimana hasil analisis dari pemodelan penyakit kusta di Provinsi Sulawesi Tenggara Menggunakan Metode Regresi Poisson Inverse Gaussian?

1.3 Tujuan Penelitian

Adapun tujuan penelitian ini adalah

1. Mengetahui pola penyebaran jumlah penyakit kusta di Sulawesi Tenggara menggunakan Peta Tematik.
2. Mengetahui hasil analisis dari pemodelan penyakit kusta di Provinsi Sulawesi Tenggara Menggunakan Metode Regresi Poisson Inverse Gaussian.

1.4 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut:

1. Bagi Penulis

Sebagai kontribusi dalam mengembangkan wawasan dalam bidang Statistika, terkhusus pada Regresi Poisson Inverse Gaussian.

2. Bagi Pembaca

Menambah wawasan serta dapat melakukan penelitian lebih lanjut untuk menemukan penyelesaian yang tepat.

3. Bagi Pemerintah

Menjadi salah satu rujukan bagi pemerintah dalam menekan/mencegah penyebaran penyakit kusta.

1.5 Batasan Penelitian

Adapun batasan penelitian ini adalah sebagai berikut:

1. Unit pengamatan yang dikaji dalam penelitian ini adalah unit kabupaten/kota Provinsi Sulawesi Tenggara.
2. Pada penelitian ini digunakan model Regresi Poisson Inverse Gaussian untuk mengetahui model penyakit kusta di Sulawesi Tenggara serta faktor-faktor yang mempengaruhi jumlah penyakit kusta di Sulawesi Tenggara