

BAB 1

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).

Berbagai eksperimen, seringkali data penelitian yang dimiliki berupa data cacahan. Model yang dapat digunakan untuk data cacahan diantaranya adalah model regresi Poisson. Regresi Poisson termasuk kedalam *Generalized Linear Model* (GLM) (Kusuma, Komalasari, Hadijati, 2013). Pada model regresi ini, peubah respon berupa data cacahan yang mengikuti distribusi Poisson. Distribusi Poisson sering digunakan untuk kejadian kejadian yang jarang terjadi dengan data berupa cacahan yang mempunyai nilai non negatif (Kismiantini, 2008).

Model regresi Poisson mensyaratkan equidispersi yaitu harus memenuhi asumsi bahwa nilai varinasi dari variabel respon harus sama dengan nilai rata-ratanya (Cahyandari, 2014). Sementara dalam penelitian regresi poisson biasanya terjadi pelanggaran terhadap asumsi tersebut. Dimana nilai variansi lebih kecil daripada rata-

ratanya yang disebut dengan *underdispersi* dan ketika nilai variansinya lebih besar dari rata-ratanya yang disebut dengan *Overdispersi*. Apabila data dengan kondisi overdispersi tetap dianalisis menggunakan analisis regresi Poisson, maka akan ada informasi yang hilang akibat tidak termodelkannya parameter disperse dalam model regresi yang terbentuk. Parameter dispersi adalah parameter yang muncul akibat tidak terjadinya kondisi ekuidispersi (Keswari, Sumarjaya, Suciptawati, 2014).

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). 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 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 (Stasinopoulos 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. 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.

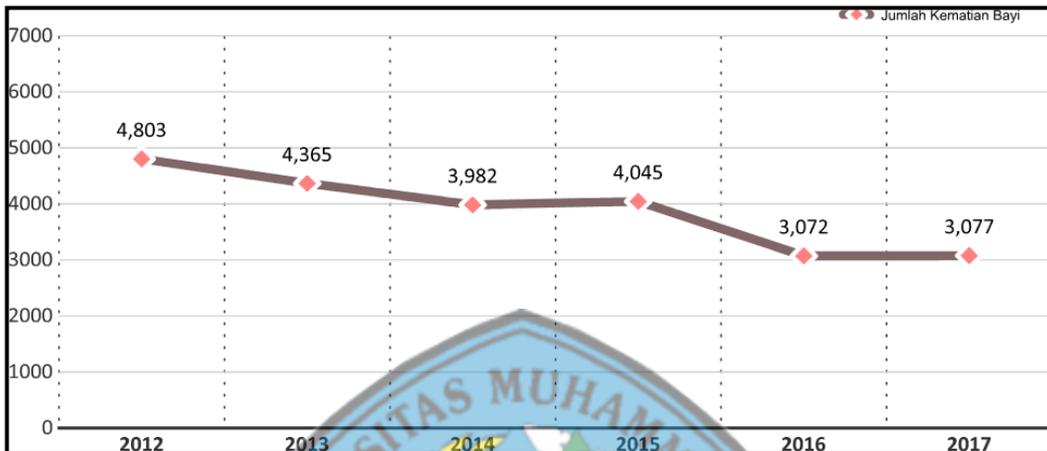
Nur, I.M (2016) menggunakan Zero Inflated Poisson pada angka kematian bayi (AKB) di Provinsi Jawa Tengah Tahun 2014. Dalam penelitiannya di peroleh model regresi Zero Inflated Poisson dimana menghasilkan variabel yang signifikan berpengaruh dimana menyatakan jumlah sarana kesehatan pada tiap kabupaten/kota, rasio ketersediaan bidan desa pada tiap kabupaten/kota dan menyatakan presentase persalinan ditolong tenaga kesehatan.

Ramadhan, M.N (2019) menggunakan regresi Poisson Inverse Gaussian pada jumlah penyakit kusta di Provinsi Sulawesi Tenggara Tahun 2017. Dalam penelitiannya model regresi Poisson Invers Gaussian dan faktor-faktor yang berpengaruh secara signifikan terhadap jumlah penyakit kusta di Provinsi Sulawesi Tenggara yaitu presentase kepadatan penduduk dan presentase penggunaan air bersih.

Pada penelitian sebelumnya juga yang mengambil kasus kematian pada bayi dengan beberapa judul yaitu pemodelan jumlah kematian bayi di Provinsi Sulawesi Selatan pada tahun 2015, jumlah kematian bayi dan faktor-faktor yang memepengaruhi jumlah kematian bayi menurut kecamatan di Kabupaten Bengkulu Selatan pada tahun 2013, dan pemodelan jumlah kematian ibu dan jumlah kematian pada bayi di Provinsi Jawa Timur pada tahun 2012.

Angka Kematian Bayi (AKB) atau *Infant Mortality Rate (IMR)* merupakan indikator yang sangat sensitif terhadap upaya pelayanan kesehatan terutama yang berhubungan dengan bayi baru lahir perinatal dan neonatal. AKB menggambarkan besarnya risiko kematian bayi (<1 tahun) dalam 1.000 kelahiran hidup. Berdasarkan kesepakatan international AKB merupakan indikator yang menggunakan konsep *rate*, meskipun dalam kenyataannya hanya *ratio*. Berdasarkan publikasi BPS, AKB Provinsi Jawa Barat sejak tahun 2007 sampai dengan 2012 cenderung mengalami penurunan. AKB berhasil diturunkan sebesar 9 poin (range 39 – 30/1.000 kelahiran hidup). Untuk AKB ditahun 2012, BPS melakukan publikasi berdasarkan SDKI 2012, di mana Provinsi Jawa Barat mempunyai AKB sebesar 30/1.000 kelahiran hidup. Berdasarkan

pencatatan dan pelaporan, di Provinsi Jawa Barat tahun 2017 terdapat 3.077 bayi meninggal meningkat 5 orang dibanding tahun 2016 yang tercatat 3.072 kematian bayi.



Sumber : Tabel Profil Kesehatan Provinsi Jawa Barat Tahun 2017

Gambar 1.1 Angka Kematia Bayi di Provinsi Jawa Barat Tahun 2012-2017

Berdasarkan pada grafik Proporsi Kematian Bayi pada tahun 2017 sebesar 3,4/1000 kelahiran hidup, menurun 0,53 poin dibanding tahun 2016 sebesar 3,93/1000 kelahiran hidup. Dari kematian bayi sebesar 3,4/1.000 kelahiran hidup, terdapat angka kematian neonatal (bayi berumur 0-28 hari) sebesar 3,1/1.000 kelahiran hidup atau 84,63 % kematian bayi berasal dari bayi usia 0-28 hari, dengan demikian disarankan dalam penanganan AKB lebih difokuskan pada Bayi Baru Lahir. Angka Kematian Bayi sebesar 3,4/1000 kelahiran hidup, sudah melampaui target MDGs yang pada tahun 2015 harus sudah mencapai 17/1.000 kelahiran hidup.

Pada penelitian ini yang akan diteliti yaitu jumlah kematian bayi di Provinsi Jawa Barat. Dimana jumlah kematian bayi merupakan salah satu indikator penting bagi pemerintah dalam mengevaluasi dibidang kesehatan. Mengetahui faktor-faktor yang

mempengaruhi, tentu saja pemerintah memiliki gambaran langkah apa saja yang harus dilakukan dalam menekan angka kematian bayi.

Data kematian bayi merupakan data cacahan sehingga dalam pemodelannya bisa menggunakan regresi Poisson. Data kematian bayi 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 penelitian yang akan dilakukan maka rumusan masalahnya adalah:

1. Bagaimana gambaran umum angka kematian bayi di Provinsi Jawa Barat?
2. Bagaimana bentuk pemodelan angka kematian bayi di Provinsi Jawa Barat menggunakan Regresi *Poisson Inverse Gaussian*?
3. Faktor-faktor apa saja yang berpengaruh secara signifikan terhadap jumlah kematian bayi di Provinsi Jawa Barat?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah penelitian maka tujuan yang ingin dicapai dari penelitian yang dilakukan adalah:

1. Untuk mengetahui gambaran umum angka kematian bayi di Provinsi Jawa Barat.

2. Untuk mengetahui pemodelan angka kematian bayi di Provinsi Jawa Barat menggunakan Regresi Poisson Inverse Gaussian.
3. Untuk mengetahui faktor-faktor yang berpengaruh secara signifikan terhadap angka kematian bayi di Provinsi Jawa Barat.

1.4 Manfaat Penelitian

Berdasarkan tujuan penelitian yang hendak dicapai, maka penelitian ini diharapkan mempunyai manfaat sebagai berikut:

1. Manfaat Teoritis

Hasil penelitian ini dapat digunakan sebagai referensni pengetahuan tentang metode statistika yang digunakan untuk mengetahui suatu hubungan dan pengaruh variabel independen terhadap variael dependen khususnya yang berkaitan dengan metode Regresi Poisson Inverse Gaussian (PIG) untuk pemodelan angka kemtian pada bayi (AKB).

2. Manfaat Praktis

- a. Bagi Penulis

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

- b. Bagi Pembaca

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

- c. Bagi Universitas

Menambah literasi kepustakaan guna menambah wawasan keilmuan dibidang statistika.

d. Bagi Pemerintah

Menjadi salah satu rujukan bagi pemerintah dalam menekan/mencegah kematian bayi khususnya di Provinsi Jawa Barat.

1.5 Batasan Penelitian

Berdasarkan penelitian ini, pembatasan masalah sangat diperlukan demi menjamin kebebasan dalam mengambil kesimpulan yang diperoleh agar tidak terjadi penyimpangan dari tujuan semula. Berdasarkan pada latar belakang masalah dan kajian-kajian pendukung lain maka pembahasan akan difokuskan pada analisis Regresi Poisson Inverse Gaussian (PIG) untuk mengetahui pemodelan terbaik angka kematian bayi di Jawa Barat dengan menggunakan metode *Backward* dengan variabel prediktor yang signifikan berdasarkan nilai AIC terkecil serta faktor-faktor yang mempengaruhi angka kematian bayi di Provinsi Jawa Barat dengan unit pengamatan yang dikaji dalam penelitian ini adalah unit kabupaten/kota Provinsi Jawa Barat. *Software* yang digunakan dalam penelitian ini adalah R-Studio program.