

# BAB I

## PENDAHULUAN

Bab ini memuat latar belakang, rumusan masalah, batasan penelitian, tujuan penelitian, metode penelitian, dan sistematika penulisan.

### 1.1. Latar Belakang

Analisis regresi berasal dari penelitian yang dilakukan oleh seorang ilmuwan Inggris, Sir Francis Galton pada abad ke-19. Pada penelitiannya Galton mempelajari hubungan antara tinggi orangtua dan tinggi anak-anak mereka dan menemukan bahwa anak-anak dari orangtua yang lebih tinggi cenderung lebih tinggi juga. Penemuan ini kemudian disebut sebagai "regresi ke arah rata-rata" dan mengarah pada pengembangan konsep regresi modern saat ini. Konsep regresi baru menjadi lebih sistematis dan diperluas oleh sejumlah ilmuwan selama abad ke-20, termasuk Karl Pearson, R.A. Fisher, Jerzy Neyman, dan Egon Pearson. Saat ini, analisis regresi digunakan luas di berbagai bidang seperti ilmu sosial, ekonomi, ilmu biologi, dan teknik.

Dalam konteks regresi terdapat satu peubah tertentu yang menjadi fokus untuk dipahami atau dimodelkan (peubah terikat), dan sekelompok peubah lain yang dapat digunakan untuk memprediksi atau memodelkan peubah terikat tersebut (peubah bebas). Biasanya, analisis regresi digunakan untuk memodelkan hubungan antara peubah terikat  $Y$  dengan satu atau lebih peubah bebas  $X$ , melakukan prediksi pada peubah terikat (peramalan), atau pengujian hipotesis. Namun, model ini memiliki kekurangan yaitu tidak cocok digunakan jika hubungan antara peubah  $X$  dan  $Y$  tidak linear. Selain itu, regresi linear mengasumsikan peubah terikatnya untuk berdistribusi normal yang mana hanya terpenuhi jika peubahnya kontinu. Artinya model ini juga tidak cocok digunakan saat peubah terikatnya berupa diskrit.

Regresi non-linear dapat digunakan untuk memodelkan hubungan antara peubah bebas dan terikat yang tidak linear. Namun, model ini lebih sulit untuk diinterpretasikan, berisiko mengalami overfitting karena lebih kompleks dari model

linear dan kerap kali hanya dapat diterapkan pada jenis peubah terikat dengan distribusi tertentu.

Pada tahun 1972, John Nelder dan Robert Weddeburn merumuskan teknik pemodelan statistik yang diberi nama *Generalized Linear Models (GLM)*. GLM ini merupakan istilah yang membawahi banyak model lainnya, termasuk Model Regresi Linear, Regresi Logistik, dan Regresi Poisson. Tidak seperti Model Regresi Linear, peubah terikatnya tidak harus berdistribusi normal namun mengikuti distribusi *exponential family* (contohnya normal, binomial, poisson, atau gamma). GLM memungkinkan untuk membangun hubungan linear antara peubah terikat dan peubah bebas, meskipun hubungan yang mendasarinya non linear. Hal ini menjadi mungkin dengan menggunakan sebuah *link function*, yang mana menghubungkan peubah terikat dengan model linear.

Pada model regresi linear biasa peubah terikatnya bersifat kontinu dan berdistribusi normal, dan peubah bebasnya bisa berupa diskrit atau kontinu. Sedangkan pada regresi logistik biner peubah terikatnya diasumsikan berdistribusi binomial dengan satu kali percobaan dan peluang sukses  $E(Y) = \pi$ . Regresi Logistik Biner memodelkan bagaimana peluang sukses untuk peubah terikat  $Y$  bergantung pada peubah bebasnya. Peubah bebasnya bisa berupa diskrit atau kontinu.

Regresi Poisson juga merupakan kasus khusus dari GLM di mana peubah terikatnya diasumsikan berdistribusi Poisson dengan mean  $\lambda$ . Regresi Poisson memodelkan nilai *mean* dari peubah terikat diskrit  $Y$  bergantung pada peubah bebasnya (kontinu atau diskrit). Model ini biasanya baik digunakan saat peubah terikatnya merupakan jumlah kejadian, seperti jumlah panggilan yang diterima *customer service* dalam satu jam atau jumlah mobil yang melalui persimpangan dalam satu hari. Tidak seperti distribusi binomial yang menghitung jumlah sukses dalam sejumlah percobaan, distribusi Poisson tidak terbatas pada jumlah sukses. Ketika semua peubah bebasnya berupa diskrit, model regresi Poisson bisa menjadi model loglinear. Menurut Agresti (2019) Model loglinear adalah *generalized linear models (GLMs)* untuk data cacah. Model ini memodelkan hubungan dan pola interaksi di antara peubah kategorik [1]–[3].

*Mixture Model* merupakan pemodelan peubah acak laten yang bertujuan untuk memulihkan kelompok yang tersembunyi dari data hasil pengamatan. Dengan *mixture model* suatu populasi yang tadinya diasumsikan merupakan satu kelompok dan bisa jadi merupakan beberapa kelompok berbeda. Peubah laten merupakan istilah yang digunakan peneliti untuk merujuk peubah dalam model yang tidak dapat diukur atau diamati secara langsung, tapi diperoleh dari peubah-peubah lain yang dapat diamati atau diukur secara langsung [3].

Dalam statistika, penggunaan peubah laten telah banyak digunakan untuk memodelkan data. Misalnya, *exploratory* dan *confirmatory factor analysis*, *latent curve models*, *item response theory*, dan *latent class analysis*. Penggunaan metode bergantung pada bagaimana analisis data yang dilakukan (prior atau posterior), jenis data (kontinu, kategorik, atau *hybrid*), atau indikator dari peubah laten tersebut [4].

Beberapa penelitian sebelumnya yang membahas model poisson loglinear di antaranya penelitian yang dilakukan oleh Ismail dan Jemain [5] yang meneliti perbandingan antara model poisson loglinear dan model binomial negatif sebagai model regresi yang dapat menangani overdispersi dan dapat menghasilkan estimasi parameter yang lebih akurat.

Penelitian Adiatma dkk. [6] membahas mengenai regresi poisson bivariat dan mengaplikasikannya pada studi kasus jumlah kematian ibu dan anak di Sulawesi Selatan. Hasil penelitian menunjukkan pada model terbaik, hanya peubah persentase bayi diberi ASI eksklusif yang berpengaruh signifikan terhadap jumlah kematian ibu. Sedangkan, peubah bebas yang berpengaruh signifikan terhadap jumlah kematian bayi adalah persentase ibu hamil yang melaksanakan K1, persentase ibu hamil yang melaksanakan K4, persentase penanganan komplikasi kebidanan, dan persentase bayi diberi ASI eksklusif.

Kemudian A. van Strien dkk. [7] mengimplementasikan model poisson loglinear untuk menganalisis data pemantauan burung. Penelitian ini bertujuan untuk mengidentifikasi faktor-faktor yang mempengaruhi populasi burung dan pola distribusinya. Hasil penelitian menunjukkan bahwa faktor habitat dan waktu pengamatan memiliki pengaruh yang signifikan pada populasi burung, sementara

faktor cuaca tidak signifikan. Metode analisis regresi Poisson log-linear juga digunakan untuk memodelkan pola distribusi burung di seluruh wilayah studi, dan menunjukkan bahwa burung cenderung berkumpul di area dengan habitat yang sesuai dengan kebutuhan mereka.

Penelitian skripsi ini akan mengkaji penggabungan antara Model Poisson Loglinear dengan *Mixture Model*. Model Poisson Loglinear mensyaratkan bahwa peubah terikatnya  $(y_1, y_2, \dots, y_n)$  memiliki nilai integer. Setiap  $y_i$  dimodelkan sebagai peubah acak bebas Poisson ( $\lambda_i$ ), di mana  $\lambda_i$  adalah kombinasi linear dari peubah bebas yang sesuai dengan pengamatan ke- $i$ . Dari model ini kemudian akan dibangun model campuran dengan memberikan parameter bobot sebagai parameter peubah latennya, kemudian bersama dengan parameter regresinya akan diestimasi bersama-sama nilai parameternya.

## 1.2. Rumusan Masalah

1. Bagaimana langkah-langkah mengestimasi parameter bobot dan parameter regresi pada Model Poisson Loglinear Campuran.
2. Bagaimana menentukan model yang sesuai untuk data prevalensi balita *stunting* di Provinsi Jawa Barat dan Bali tahun 2021 pada Model Poisson Loglinear Campuran.

## 1.3. Batasan Masalah

Untuk memastikan tidak terjadi penyimpangan dari topik yang dibahas, maka perlu ditetapkan batasan masalah dalam penelitian ini sebagai berikut:

1. Metode yang digunakan untuk mengestimasi parameter bobot dan parameter regresi pada Model Poisson Loglinear Campuran pada penelitian ini adalah Maksimum Likelihood dengan iterasi Algoritma Ekspektasi-Maksimisasi (EM).
2. Komponen campuran yang ditetapkan berjumlah dua.

#### **1.4. Tujuan dan Manfaat Penelitian**

Tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Mengestimasi parameter bobot dan parameter regresi pada Model Poisson Loglinear Campuran dengan Algoritma Ekspektasi-Maksimisasi (EM).
2. Menentukan model yang sesuai untuk data prevalensi balita *stunting* di Provinsi Jawa Barat dan Bali tahun 2021 pada Model Poisson Loglinear Campuran.

#### **1.5. Metode Penelitian**

Dalam penelitian ini, digunakan metode berupa studi pustaka yang di mana peneliti mengkaji dan memahami topik yang dibahas dalam penelitian melalui buku-buku dan jurnal-jurnal terkait. Topik penelitian yang dikaji kemudian diterapkan ke dalam studi kasus setelah itu dilakukan analisis statistik.

Analisis statistik yang dilakukan yaitu analisis deskriptif dan inferensi. Analisis deskriptif bertujuan untuk mendeskripsikan karakteristik dari data, seperti ukuran pemusatan dan ukuran penyebarannya. Sedangkan, analisis inferensi dilakukan untuk membuat pendugaan dari data yang diteliti. Teknik ini bertujuan untuk mengetahui apakah hasil penelitian dapat digeneralisasi dan sejauh mana hasil tersebut merepresentasikan populasi secara keseluruhan. Pada penelitian ini hasil analisis statistik tersebut diperoleh dengan alat bantu berupa *software R* dan excel.

#### **1.6. Sistematika Penulisan**

Skripsi ini tersusun dalam lima bab, di mana setiap bab dibagi menjadi beberapa subbab sesuai dengan sistematika penulisan yang digunakan.

##### **BAB I: PENDAHULUAN**

Bab ini memuat pendahuluan yang meliputi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian, dan sistematika penulisan.

## BAB II: LANDASAN TEORI

Bab ini memuat teori-teori terkait topik pembahasan dalam skripsi. Secara umum bab ini meliputi semua yang berkaitan dengan data kategorik, distribusi peluang pada data kategorik, metode estimasi, kriteria informasi Akaike konsisten, uji Kolmogorov-Smirnov, dan uji Wald.

## BAB III: MODEL POISSON LOGLINEAR CAMPURAN

Bab ini memuat pengantar kajian utama yang dibahas dalam skripsi, termasuk pembahasan mengenai Model Poisson Loglinear, Model Campuran, dan Model Poisson Loglinear Campuran.

## BAB IV: STUDI KASUS DAN ANALISIS

Bab ini memuat langkah-langkah analisis data, pengujian data, serta hasil analisis yang diperoleh.

## BAB V: KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dan saran untuk penelitian yang lebih baik.

