

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Statistik merupakan hasil data yang disajikan dalam bentuk tabel, grafik, dan sebagainya. Menurut Sudjana (2005) statistik adalah pengetahuan yang berhubungan dengan cara-cara pengumpulan data, pengolahan atau penganalisisannya dan penarikan kesimpulan berdasarkan kumpulan data yang dilakukan analisis. Adapun statistik sendiri terdapat dua jenis data yaitu kuantitatif dan kualitatif, dalam penelitian kali ini menggunakan data kuantitatif. Menurut Sudjana (2005) data kuantitatif data yang berbentuk bilangan, harganya berubah-ubah atau bersifat variabel.

Pada data kuantitatif dikelompokkan dalam dua bentuk yaitu data diskrit dan data kontinu, peneliti menggunakan data diskrit. Data diskrit yaitu data dalam bentuk angka (bilangan) yang diperoleh dengan cara membilang. Salah satunya mempunyai struktur lebih dari dua kategori (polikotomus), misalnya kejadian untuk mengukur kondisi mesin berupa tidak rusak, rusak, dan perlu diganti serta masih banyak lagi bentuk kejadian yang dapat dikategorikan lebih dari dua kategori.

Kejadian dengan lebih dua kategori dapat dinyatakan dengan model yang mengikuti distribusi multinomial. Ada beberapa asumsi dasar yang harus dipenuhi pada saat mengaplikasikan model yang berdistribusi multinomial pada gugus data tertentu. Beberapa asumsi itu salah satunya adalah bahwa variabel respon merupakan variabel acak yang saling bebas, dan peluang sukses dari suatu kejadian adalah konstan. Namun pada praktiknya tak jarang terjadi pelanggaran asumsi, di mana terdapat ketidakbebasan antar variabel acak. Ketidakbebasan antar variabel acak dimaknai sebagai adanya korelasi diantara variabel respon, hal tersebut merupakan suatu bukti indikasi bahwa terdapat masalah yang disebut overdispersi. Sebagaimana

yang diungkapkan oleh McCullagh dan Nelder (1989), masalah overdispersi akan sering dijumpai dalam analisis data diskrit, baik variabel respon yang berbentuk biner (*dikotomis*), cacahan, maupun *politokomis* seperti dalam model yang berdistribusi multinomial ini.

Masalah overdispersi yang terjadi pada pengamatan data diskrit ini dapat dijelaskan oleh dua hal, yaitu adanya korelasi antar variabel acak dan adanya keragaman dalam menentukan peluang pada variabel respon. Keduanya merupakan kejadian yang saling berhubungan, sehingga ketika terdapat korelasi antar variabel respon, maka akan terdapat juga keragaman pada peluang variabel respon dan begitupun sebaliknya. McCullagh dan Nelder (1989) menyatakan bahwa kedua kejadian tersebut dapat terjadi karena adanya pengelompokan (*clustering*) dalam populasi. Sedangkan Collet (1991) menyebutkan bahwa kejadian-kejadian tersebut muncul karena sejumlah unit percobaan diamati beberapa kali pada kondisi yang sama, sehingga akan diperoleh suatu peluang pada variabel respon yang berbeda dari satu percobaan ke percobaan yang lainnya. Munculnya masalah overdispersi dalam data diskrit juga akan berdampak pada nilai taksiran untuk variansnya. Ketika nilai penaksir varians ini digunakan dalam menghitung selang kepercayaan (*confident interval*), maka akan menghasilkan rata-rata yang terlalu kecil sehingga menyebabkan selang kepercayaan yang terlalu pendek. Selain itu juga, apabila nilai penaksir parameter ini digunakan untuk pengujian hipotesis maka hasilnya akan cenderung menerima hipotesis nol.

Berdasarkan uraian di atas, maka perlu dicari suatu metode statistika yang tepat untuk dapat digunakan dalam menentukan taksiran peluang bagi parameter pada data yang mengikuti distribusi multinomial, dimana pada data yang akan di analisis terdapat masalah overdispersi di dalamnya.

Ada beberapa macam model distribusi yang telah diusulkan oleh para penulis untuk mengatasi masalah overdispersi pada data multinomial, distribusi Dirichlet multinomial yaitu gabungan dari dua buah distribusi multinomial Dirichlet dan diusulkan oleh Mosimann (1962), distribusi *random-clumped* multinomial merupakan campuran dari dua buah distribusi diskrit yaitu distribusi binomial dan multinomial diusulkan oleh Morel & Nagaraj (1993) dan Neerchal & Morel (1998).

Dari beberapa macam model distribusi yang dapat digunakan untuk mengatasi masalah overdispersi, salah satunya adalah dengan menggunakan distribusi *random-clumped* multinomial. Distribusi *random-clumped* multinomial ini yang akan penulis bahas dan digunakan pada data tingkat kepuasan pengguna PDAM di Kelurahan Antapani Wetan tahun 2019.

1.2 Tujuan dan Manfaat Penelitian

Tujuan umum yang ingin dicapai dalam penelitian untuk memberikan informasi baru mengenai overdispersi pada data yang mengikuti distribusi multinomial. Berdasarkan latar belakang masalah yang telah diuraikan dalam penulisan skripsi terdapat juga tujuan khusus yaitu:

1. Mendeteksi masalah overdispersi pada data berdistribusi multinomial dengan menggunakan statistik uji Z.
2. Mengatasi masalah overdispersi pada data berdistribusi multinomial dengan menggunakan distribusi *random-clumped* multinomial.

Adapun manfaat penelitian secara praktis yaitu menambah ilmu baru yang belum dibahas pada saat perkuliahan, dan manfaat secara teoritis dapat mengembangkan masalah overdispersi melalui metode distribusi *random-clumped* multinomial.

1.3 Sistematika Penulisan

Untuk mempermudah pemahaman mengenai pembahasan skripsi ini, maka berikut ini akan dipaparkan secara garis besar tentang sistematika penulisan skripsi ini. Pada Bab I dijelaskan tentang pendahuluan yang memuat latar belakang masalah, tujuan dan manfaat penelitian, serta sistematika penulisan. Pada Bab II berisikan tentang uraian tinjauan pustaka yang merupakan teori-teori yang berkaitan dengan permasalahan yang dibahas dalam Bab IV. Pada Bab III akan diuraikan mengenai data dan sumber data yang digunakan, serta tahapan penelitian. Bab IV berisikan hasil-hasil perhitungan berdasarkan metode yang digunakan, kemudian dari hasil analisis yang diperoleh ditarik kesimpulan yang diuraikan pada Bab V.

