

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Secara garis besar penyajian data pada analisis statistika ada dua cara, yaitu dengan menggunakan tabel dan grafik. Biasanya peragaan data dalam bentuk grafik lebih disenangi dibandingkan dengan penyajian dalam bentuk tabel atau dalam bentuk narasi, karena lebih menarik, lebih informatif, dan lebih mudah dipahami. Pada analisis statistika peubah ganda, salah satu grafik yang biasa digunakan adalah biplot.

Analisis biplot merupakan suatu alat statistika yang menyajikan posisi relatif  $n$  objek dengan  $p$  peubah secara simultan dalam dua dimensi. Menggunakan analisis biplot, data berpeubah ganda dapat direpresentasikan dalam ruang berdimensi kecil. Analisis biplot dapat memberikan informasi tentang objek dan peubah dalam satu gambar, diantaranya adalah hubungan antar peubah, kemiripan relatif antar objek pengamatan, hubungan relatif antara objek dan peubah, dan keragaman peubah. Perbedaan hasil analisis biplot dengan kenyataan yang sebenarnya dapat saja terjadi. Hal ini mungkin karena pada dasarnya representasi data dengan analisis biplot merupakan pendekatan terhadap data yang berdimensi besar ke dalam dimensi dua, sehingga ada informasi yang hilang. Namun ini bukanlah masalah serius selama informasi yang penting dapat diperoleh.

Analisis biplot diusulkan pertama kali oleh Gabriel dengan didasarkan pada Penguraian Nilai Singular (PNS), PNS memerlukan matriks data. Warsito (2009) mengemukakan jika dalam suatu matriks data terdapat data *outlier* maka perhitungan terhadap matriks tersebut tidak memberikan hasil yang mencerminkan data sebenarnya. Dampak keberadaan data *outlier* diduga akan mengganggu dalam proses analisis data. Pengamatan data *outlier* mungkin berpengaruh sekali pada struktur matriks *covariance*, karenanya perlu dicari sebuah alternatif biplot yang kekar (*robust*) terhadap pengaruh data *outlier*, misalnya dengan menggunakan matriks *covariance* yang *robust*.

Metode pendugaan *robust* yang sudah digunakan adalah, *Minimum Volume Ellipsoid* (MVE), *Minimum Covariance Determinant* (MCD), *M-estimator*, dan *S-Estimator*. Penduga *robust* MCD merupakan rata-rata dan *covariance* dari sebagian pengamatan yang meminimumkan determinan matriks *covariance*. Metode MCD bertujuan untuk menemukan *h-observations* atau subsampel dari data dimana matriks *covariance*-nya memiliki determinan terkecil.

Pada penelitian-penelitian sebelumnya yang dikemukakan oleh Budyandra (2010), metode *Minimum Covariance Determinant* dibandingkan dengan metode pendugaan *robust* yang lain memiliki kelebihan yaitu efisiensi statistiknya lebih baik dibandingkan MVE (Butler dkk, 1993 dalam Rousseeuw, 1999) dan metode MCD memiliki *convergence rate* yang lebih cepat dibanding MVE (Davies, 1992). Walaupun memiliki banyak kelebihan, ternyata metode MCD jarang digunakan, karena metode ini cukup sulit dalam penghitungannya serta membutuhkan waktu penghitungan yang lebih lama bila jumlah pengamatannya besar. Untuk mengatasi keterbatasan ini, maka Rousseeuw dan van Driessen (1999) mengemukakan suatu algoritma baru untuk metode MCD yang dinamakan algoritma *Fast-MCD*, algoritma *fast-MCD* selanjutnya disebut metode *fast-MCD*. Penaksir *robust* dengan metode *fast-MCD* tidak terpengaruh oleh data *outlier*. (Hubert and van Driessen, 2002). Metode *fast-MCD* memiliki sifat statistik yang baik karena tergolong penaksir *robust* dengan *breakdown point* tinggi karena memenuhi batas nilai maksimum *breakdown* 50 persen (Hubert dkk, 2008). Metode *fast-MCD* selain memenuhi sifat statistik *high breakdown point*, juga efisien dalam komputasinya dan yang paling penting adalah efektif dalam mengatasi pengamatan yang mengandung data *outlier*.

Metode *fast-MCD* belum banyak dipakai dalam mengatasi data *outlier*, padahal metode ini banyak sekali kelebihannya, sehingga dalam skripsi ini penulis akan membahas tentang analisis *robust* biplot dengan menggunakan metode *fast-MCD* pada data *outlier*. *Robust* biplot dilakukan dengan menggunakan *first two component* yaitu dengan menggunakan *eigen value* terbesar pertama dan terbesar kedua untuk menentukan matriks kiri ( $\mathbf{U}$ ) dan matriks kanan ( $\mathbf{A}$ ), di mana *eigen value* diperoleh dari matriks *covariance robust* yang diperoleh dengan menggunakan metode *fast Minimum Covariance Determinant*.

## 1.2. Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah yang akan dibahas dalam penelitian ini adalah: Apakah terdapat perbedaan hasil analisis biplot klasik dan informasi yang diperoleh dari analisis *robust* biplot menggunakan *fast-Minimum Covariance Determinant (fast-MCD)*?

## 1.3. Batasan Masalah

Batasan masalah yang digunakan dalam penelitian ini adalah:

1. Pendeteksian data *outlier* dilakukan dengan menggunakan jarak mahalanobis
2. Analisis biplot yang digunakan merupakan analisis biplot dua dimensi
3. Analisis biplot klasik menggunakan  $\alpha = \frac{1}{2}$ .

## 1.4. Tujuan

Berdasarkan rumusan masalah, tujuan dari penelitian ini adalah membandingkan hasil analisis biplot klasik dan *robust* biplot dengan menggunakan metode pendugaan *robust fast-Minimum Covariance Determinant (fast-MCD)* pada data *outlier* berdasarkan gambar biplot yang diperoleh dan pemeriksaan kesesuaian biplot.

## 1.5. Manfaat

Penyusunan skripsi ini diharapkan dapat memberi masukan tentang analisis biplot pada data *outlier* dan dapat menjelaskan analisis *robust* biplot metode pendugaan *fast Minimum Covariance Determinant (fast-MCD)*.

