

ABSTRAK

Penggunaan analisis regresi untuk suatu model $Y = X\beta + \varepsilon$, dengan p parameter, diberlakukan asumsi terhadap faktor galat ε , yaitu $\varepsilon \sim N(0, I\sigma^2)$. Dalam prakteknya penyimpangan terhadap asumsi - asumsi tersebut seringkali terjadi. Jenis penyimpangan tersebut seperti terjadinya sebaran faktor galat yang normal, tetapi mengandung pengamatan yang merupakan pencilan (*outlier*). Dengan kondisi demikian ini dikatakan bahwa parameter - parameter dalam model regresi yang diestimasi dengan kuadrat terkecil biasa (OLS) bersifat tidak tegar terhadap adanya outlier. Penolakan begitu saja terhadap outlier adalah merupakan prosedur yang kurang bijaksana. Oleh karena itu diperlukan metode estimasi parameter yang bersifat tegar untuk mengatasi adanya outlier. Dalam penulisan tugas akhir ini akan dibahas mengenai sisaan yang distudentkan secara eksternal (*externally studentized residual*), untuk mengidentifikasi suatu titik data merupakan outlier atau bukan dan dibahas pula mengenai estimasi parameter yang bersifat tegar yaitu estimator M, sebagai alternatif pemecahan adanya titik data yang merupakan outlier.

BAB I

PENDAHULUAN

Sering kali kita dihadapkan adanya perubahan yang akan mempengaruhi metode kuadrat terkecil tersebut tidak terpengaruh terhadap adanya outlier, artinya nilai objek yang lainnya. Pendekatan analisis regresi merupakan analisis yang estimasinya dapat dipengaruhi secara kuat oleh adanya satu titik data ekstrem yang digunakan untuk mempelajari bentuk hubungan ketergantungan antara satu merupakan outlier, peubah tak bebas dengan satu atau lebih peubah bebas, yang di dalamnya terdapat asumsi-asumsi pokok yang mendasari.

Prosedur umum untuk menduga pola hubungan yang dimaksud adalah dengan mencocokkan suatu bentuk fungsional, sedemikian rupa sehingga terhadap adanya outlier. Oleh karena outlier tidak selalu berarti pengamatan ini komponen kesalahan model regresi

$$Y = X\beta + \varepsilon, \text{ dapat diminimumkan.}$$

Salah satu bentuk pendugaan / estimasi untuk memperoleh estimasi parameter untuk model di atas adalah dengan metode Kuadrat Terkecil Biasa / Ordinary Least Square (**OLS**). Estimasi parameter untuk n pengamatan dengan metode kuadrat terkecil pada model regresi di atas, dimana ada p parameter, diberlakukan asumsi tertentu terhadap faktor galat ε , yaitu $\varepsilon \sim N(0, I\sigma^2)$.

Dalam prakteknya penyimpangan terhadap asumsi-asumsi tersebut seringkali terjadi, seperti adanya faktor galat normal, tetapi mengandung pengamatan yang merupakan pencilan (*outlier*). Penyimpangan ini dapat diamati melalui perilaku galat atau sisaannya. Sisaan yang merupakan outlier adalah sisaan yang nilai mutlaknya jauh lebih besar dari sisaan-sisaan yang lainnya.

Namun adanya data yang merupakan outlier dapat memberikan

pengaruh terhadap hasil analisis regresi. Seperti terjadinya koefisien regresi yang seharusnya signifikan menjadi tidak signifikan. Dengan demikian dikatakan bahwa parameter-parameter dalam model regresi linier yang diestimasi dengan metode kuadrat terkecil bersifat tidak tegar terhadap adanya outlier, artinya nilai estimasinya dapat dipengaruhi secara kuat oleh adanya satu titik data ekstrim yang merupakan outlier.

Identifikasi outlier dan melihat peranannya terhadap nilai estimasi, merupakan tahapan diagnosis yang perlu ditempuh terutama jika estimasi modelnya dilakukan dengan metode kuadrat terkecil yang dikenal cukup peka terhadap adanya outlier. Oleh karena outlier tidak selalu berarti pengamatan ini keliru, maka pengikutsertaan atau penyisihan outlier dalam penyusunan model bukanlah masalah yang sederhana, tetapi membutuhkan penjelasan yang lebih mendalam.

Prosedur analisis statistik yang kita harapkan adalah prosedur yang menghasilkan keluaran yang cukup baik, meskipun beberapa asumsinya tidak terpenuhi secara sempurna. Metode diagnosis sisaan ditujukan untuk menilai keanehan data dan pengaruhnya terhadap analisis. (Aunuddin, 1989).

Titik data ekstrim yang merupakan outlier adalah suatu penyimpangan atau keganjilan dan menandakan suatu titik data yang sama sekali tidak tipikal dibandingkan dengan data yang lainnya. Penolakan begitu saja terhadap titik data ekstrim yang merupakan outlier bukanlah prosedur yang bijaksana. Adakalanya suatu outlier memberikan informasi yang tidak dapat diberikan oleh titik data yang lain. (N. Draper dan H. Smith, 1992).

Maka untuk membahas permasalahan diatas, dalam tahap tindakan pengamatan untuk mengetahui suatu titik data ekstrim itu merupakan outlier atau bukan, akan dilakukan dengan analisis pada sisaannya, yaitu dengan sisaan yang distudentkan secara eksternal (*externally studentized residual*). (Weisberg, 1985).

Dilain pihak prosedur statistik yang bersifat robust ditujukan untuk mengakomodasi adanya keanehan data dan sekaligus meniadakan pengaruhnya terhadap hasil analisis tanpa terlebih dahulu mengadakan identifikasi terhadap titik data yang tidak cocok dengan model. Prosedur ini lebih bersifat otomatis dalam menanggulangi kelainan data. Oleh karena itu diperlukan suatu estimasi yang bersifat tegar, dimana nilai dugaanya tidak banyak dipengaruhi oleh perubahan kecil dalam data. (N.Draper dan H.Smith, 1992).

Sehingga dalam tahap selanjutnya diperlukan suatu prosedur estimasi parameter yang bersifat tegar terhadap adanya outlier, maka estimasi parameter dengan penduga / estimator M merupakan alternatif pendugaan / estimasi yang bersifat tegar (Myers, 1990).

Untuk itu dalam penulisan tugas akhir ini akan dibahas mengenai sisaan yang distudentkan secara eksternal (*externally studentized residual*), untuk mengidentifikasi suatu titik data ekstrim itu merupakan outlier atau bukan dan akan dibahas pula mengenai penduga / estimator M, yang merupakan estimasi parameter yang bersifat tegar, sebagai alternatif pemecahan adanya titik data yang merupakan outlier.