

Peningkatan Kualitas Informasi pada Proses Pengolahan dan Analisis Data

Kasus : Kajian Residual dalam Mengatasi Data Pencilan (*outlier*) pada Penggunaan Regresi Linier Sederhana

Iwa Sungkawa

Fakultas Sain dan Teknologi UBINUS
iwasungkawa@yahoo.com

ABSTRAK

Perkembangan teknologi informasi yang cukup cepat perlu diimbangi oleh data atau informasi yang berkualitas. Peningkatan kualitas data dan informasi dapat dilakukan dalam setiap proses pengelolaannya (proses pengumpulan, proses analisis dan pada proses penyajiannya). Dalam tulisan ini akan dibahas proses analisis data/informasi jika terdapat data pencilan. Tujuan dari penulisan ini adalah untuk memberi gambaran bahwa mutu atau kualitas dari informasi diantaranya tergantung pada saat proses analisis data, serta mereview pengkajian residual dalam penggunaan regresi linier sederhana dan cara mengatasi data pencilan.

Residual dalam analisis regresi merupakan bagian yang dianggap menyebar normal sehingga dalam melakukan analisis data, sebelumnya perlu dilakukan uji normalitas terhadap residual atau langsung terhadap data hasil pengamatan. Hal ini perlu dilakukan karena jika asumsinya tidak dipenuhi maka hasil analisisnya dapat dianggap tidak syah dan dengan sendirinya kualitas data/informasinya diragukan. Dalam menanggulangi data pencilan (*outlier*) terlebih dulu dilakukan pendeteksian terhadap adanya gejala pencilan yang ditempuh dengan menentukan nilai mutlak residual dari setiap pengamatan atau $|e_i|$ dan tentukan mediannya (M). Selanjutnya lakukan kajian *outlier* untuk semua pengamatan.

Hasil kajian menunjukkan bahwa untuk uji normalitas, dapat dilakukan terhadap bagian residual atau bagian peubah respon (peubah tak bebas). Jika terdapat data pencilan akan berakibat koefisien korelasinya rendah sehingga kualitas model regresi lemah untuk digunakan sebagai prediktor. Untuk menanggulangi adanya gejala data pencilan perlu dilakukan transformasi terhadap data tersebut agar memperoleh hasil kajian yang baik, sehingga informasi yang dihasilkan akan lebih baik.

Kata kunci : residual, normalitas, pencilan, analisis regresi, korelasi, median.

I PENDAHULUAN

Latar Belakang

Dalam rangka penyusunan suatu perencanaan pada pelaksanaan pembangunan sektor pertanian, diperlukan data dan informasi yang berkualitas agar rencana yang disusun dapat dilaksanakan dengan baik dan dalam pelaksanaannya mencapai target yang sesuai dengan harapan. Untuk mendukung hal tersebut, teknologi informasi yang merupakan bagian dalam pengelolaan data dan informasi telah berkembang dengan pesat. Sejalan dengan perkembangan teknologi informasi yang cukup cepat dalam mendukung pembangunan sektor pertanian perlu diimbangi dengan peningkatan mutu atau kualitas dari data dan informasi yang merupakan bahan bakunya. Peningkatan kualitas data dan informasi dapat dilakukan dalam setiap proses pengelolaannya (proses pengumpulan, proses analisis dan pada proses penyajiannya).

Dalam proses pengumpulan data, yang biasa ditempuh dengan melakukan penarikan sampel (sampling) atau melakukan percobaan sering dijumpai berbagai kekeliruan atau kesalahan yang tentunya dapat menurunkan kualitas dari data atau informasi yang diperoleh. Untuk hal ini, perlu upaya untuk mencermati berbagai kekeliruan/kesalahan yang biasa dijumpai dalam setiap proses pengelolaan data dan informasi. Jenis kekeliruan yang biasa dijumpai dalam sampling, diantaranya adalah kekeliruan yang diakibatkan karena salah memilih individu ke dalam sampel (sampling error) dan kekeliruan yang bukan kesalahan pemilihan individu ke dalam sampel (non sampling error). Kekeliruan atau kesalahan yang biasa terjadi pada saat melakukan percobaan dan kekeliruan yang biasa dijumpai dalam proses pengolahan dan analisis data, juga mempunyai andil dalam penurunan kualitas informasi yang dihasilkan dari suatu kegiatan. Dalam tahapan penyajian atau penataan data dan informasi biasa dilakukan dalam mempublikasikan hasil dari suatu kegiatan, pada tahapan inipun tidak jarang dijumpai suatu kekeliruan atau

kesalahan yang berdampak sama terhadap kualitas dari informasi.

Dengan mencermati berbagai kekeliruan atau kesalahan pada proses pengolahan dan analisis data, kita dapat mensiasati untuk menghindari dan meminimalisir kekeliruan tersebut agar diperoleh hasil olahan yang baik sehingga dapat meningkatkan kualitas informasi yang dihasilkan dari suatu kegiatan penelitian atau pengamatan.

Tujuan dan Manfaat

Tujuan dari penulisan ini adalah untuk memberi gambaran bahwa mutu atau kualitas dari informasi diantaranya tergantung pada saat proses analisis data, serta mereview pengkajian residual dalam penggunaan regresi linier sederhana dan cara mengatasi data pencilan. Dalam proses pengolahan dan analisis data hasil penelitian yang mengamati bentuk hubungan fungsional antara peubah bebas dan peubah tak bebas dengan menggunakan analisis regresi linier sederhana terdapat residual e_i yang merupakan penyimpangan dari hasil prediksi dengan hasil pengamatan.

Dalam tulisan ini akan dibahas kasus dalam proses analisis data/informasi dengan menggunakan analisis regresi linier sederhana dan terdapat data pencilan. Akibat adanya data pencilan nilai residual dalam analisis regresi akan makin besar, berarti penyimpangan nilai prediksi akan semakin besar, sehingga koefisien regresi dan korelasinya akan bertambah kecil atau menurun. Secara otomatis model regresi yang dihasilkan kurang representatif dan lemah dalam menginterpretasikan kondisi ketergantungan antar peubah yang diamati. Residual dalam analisis regresi merupakan bagian yang dianggap menyebar normal sehingga dalam melakukan analisis data, sebelumnya perlu dilakukan uji normalitas terhadap residual atau langsung terhadap data hasil pengamatan. Hal ini perlu dilakukan karena jika asumsinya tidak dipenuhi maka hasil analisisnya dapat dianggap tidak syah dan dengan sendirinya kualitas data/informasinya diragukan dan dianggap menurun. Di samping itu untuk perlu penanggulangan data pencilan (outlier) dengan terlebih dulu dilakukan pendeteksian terhadap adanya gejala pencilan dari data hasil pengamatan. Untuk ini dapat dilakukan dengan menentukan nilai mutlak residual dari setiap pengamatan atau $|e_i|$ dan selanjutnya dilakukan kajian outlier untuk semua pengamatan.

Dengan memperhatikan berbagai persyaratan yang diperlukan pada saat melakukan proses pengolahan dan analisis data diharapkan dapat mengurangi kekeliruan atau kesalahan yang dapat timbul/terjadi sehingga hasilnya dapat memperbaiki kualitas dari informasi yang dihasilkan.

Kualitas Data Yang Baik

Menurut Budi Suharjo, data dianggap baik jika memiliki tiga sifat utama, yaitu : Valid; Reliable dan Representatif (<http://pr4bu.co.cc/?p=382>)

- Suatu data dikatakan valid (sah) bila mampu mencerminkan ciri atau karakteristik (fenomena) yang sebenarnya dari objek amatan.
- Suatu data dikatakan reliable, bila antar ciri yang diamati memiliki **saling ketergantungan** yang **konsisten** atau **stabil** sehingga **akurat** dalam menjelaskan perilaku objek amatan dan memungkinkan dilakukan **prediksi** terhadapnya
- Data yang dikumpulkan dalam suatu penelitian umumnya berasal dari sampel (sebagian dari populasi). Oleh karenanya data harus mampu mencerminkan kondisi (sebagian besar) populasi yang diwakilinya.

Selain tiga sifat di atas, data yang tergolong baik adalah harus lengkap, akurat (terandalkan), terkini (up to date), mudah dan cepat diakses, serta benar-benar diperlukan. Secanggih dan sebaik apapun data, tidak ada artinya jika data tersebut tidak diperlukan. Untuk menunjang perkembangan teknologi informasi yang cukup pesat diperlukan informasi tentang data atau informasi yang diperlukan oleh pengguna (informasi tentang kebutuhan informasi)

II PENURUNAN KUALITAS INFORMASI PADA PENGOLAHAN DAN ANALISIS DATA

Kualitas informasi yang dihasilkan melalui suatu kegiatan penelitian sangat tergantung pada berbagai faktor dan kondisi yang sedang dihadapi. Faktor dan kondisi dimaksud, diantaranya adalah pada saat berjalannya proses pengelolaan data. Proses pengelolaan data yang biasa dilakukan terdiri dari tiga tahapan, yaitu : tahap pengumpulan, tahap pengolahan dan analisis serta pada tahap penyajian data dan informasi. Dalam pelaksanaannya, tidak jarang kita melakukan kekeliruan atau kesalahan pada setiap tahapan, baik pada tahap pengumpulan data yang dapat ditempuh melalui survai (sampling) dan melalui percobaan; pada tahap pengolahan dan analisis data dengan metode sebagai pisau analisis yang sesuai; maupun pada tahap penyajian data/informasi dalam berbagai bentuk penyajian. Kekeliruan atau kesalahan tersebut dapat berakibat pada penurunan hasil kajian dari suatu kegiatan dan dengan sendirinya berpengaruh pada menurunnya kualitas informasi yang dihasilkan. Penurunan kualitas informasi pada saat melakukan pengumpulan data melalui survai (sampling) diantaranya disebabkan karena adanya kekeliruan atau kesalahan pada saat pemilihan individu kedalam sampel (sampling error) dan kesalahan bukan karena salah memilih individu ke dalam sampel (non sampling error). Pada saat melakukan percobaan sebagai upaya pengumpulan data dan informasi dan pada saat menyajikannya, kita tidak terlepas melakukan kekeliruan

atau kesalahan yang juga berakibat menurunkan kualitas dari data atau informasi yang dihasilkan.

Dalam kesempatan ini akan dibahas gejala menurunnya kualitas data atau informasi yang diakibatkan oleh kekeliruan atau kesalahan pada saat melakukan pengolahan dan analisis data/informasi. Beberapa jenis kekeliruan atau kesalahan pada proses pengolahan dan analisis data diantaranya adalah sebagai berikut :

- Tidak terpenuhinya asumsi sebaran yang diperlukan
- Kekeliruan/kesalahan dalam pengklasifikasian data
- Kurang tepatnya metode untuk pengolahan dan analisis data
- Kekeliruan/kesalahan pada entri dan edit data serta pada pelaksanaan pengolahan dan analisis data
- Kurang kehati-hatian dalam interpretasi hasil olahan

Di samping itu sumber kesalahan dalam pengumpulan data melalui penarikan sampel (survei) adalah :

- Variasi Acak (Random Variation)
Variasi acak merupakan kesalahan sampling yang paling umum dijumpai
- Kesalahan spesifikasi (*mis-specification of sample subject*)
- Kesalahan penentuan responden
- Kesalahan karena ketidaklengkapan cakupan daftar populasi (*coverage error*).
- Kesalahan karena ketidaklengkapan respon (*Non response error*)
- Kesalahan penarikan sampel (*sampling error*)
- Kesalahan pengukuran (*Measurement error*)

III UPAYA UNTUK MENINGKATKAN KUALITAS INFORMASI

Hindari kekeliruan atau kesalahan yang biasa dilakukan pada proses pengelolaan data agar diperoleh informasi yang berkualitas. Usahakan agar semua prosedur dilaksanakan secara benar dan semua persyaratan yang diperlukan dapat terpenuhi. Berikut adalah syarat-syarat yang diperlukan agar memperoleh data yang valid, reliabel dan representatif.

- Syarat mendapatkan data yang valid :
 1. Berkaitan langsung dengan permasalahan yang dihadapi
 2. Diukur atau diamati dengan menggunakan suatu kaidah (*concept*) atau landasan teori tertentu yang mendasarinya.
 3. Menggunakan alat ukur yang sesuai dengan perilaku ciri yang akan diamati.
 4. Menciptakan kondisi yang optimal bagi objek amatan agar ciri dapat diukur secara alami.
- Syarat mendapatkan data yang reliabel :
 1. Memastikan objek amatan memiliki kapabilitas (*experience* dan *knowledge*) yg memadai sesuai dg tujuan riset.

2. Menciptakan kondisi yang optimal (*convenience*) bagi objek amatan agar ciri dapat diukur secara alami

- Syarat mendapatkan data yang representatif :
 1. Definisikan Populasi dengan jelas
 2. Mengetahui sebaran populasi
 3. Mengetahui struktur populasi yang akan diteliti
 4. Menentukan satuan analisis dan satuan amatan

Di samping itu yang perlu diperhatikan agar mendapatkan informasi yang berkualitas adalah :

- Cek data hasil pengamatan apakah sudah memenuhi persyaratan yang diperlukan (asumsi sebaran dan asumsi lainnya)
- Sesuaikan bentuk klasifikasi yang cocok dengan kondisi data
- Pilih metode analisis yang sesuai dengan jenis data dan tujuan penelitian
- Lakukan secara hati-hati dan teliti pada pelaksanaan entri/edit data dan pada pelaksanaan analisisnya, termasuk interpretasi hasil.

Pada pelaksanaan pengolahan dan analisis data, sebaran yang sering digunakan diantaranya adalah sebaran normal baku, sebaran t, sebaran khi-kuadrat dan sebaran F. Jadi pada saat sebaran-sebaran tersebut digunakan, perlu diperhatikan aturan-aturan yang harus dipenuhi agar terjamin keabsyahannya. Beberapa hal yang dapat dikemukakan berkenaan dengan sifat-sifat dari sebaran diatas adalah sebagai berikut.

Menurut Hogg, R.V. dan A.T. Craig: jika X merupakan peubah acak kontinu yang menyebar normal dengan rata-rata μ dan ragam (variansi) σ^2 maka peubah acak $V = (X-\mu)^2/\sigma^2$ akan menyebar khi-kuadrat dengan derajat bebas satu. Selanjutnya bila X_1, X_2, \dots, X_n sampel acak berukuran n yang diambil dari populasi yang menyebar normal dengan rata-rata μ dan ragam (variansi) σ^2 maka:

- 1) rata-rata untuk \bar{x} adalah μ dengan ragamnya adalah $\frac{\sigma^2}{n}$
- 2) jika s^2 ragam (variance) dari sampel di atas, maka peubah acak $(n-1)s^2/\sigma^2$ berdistribusi khi-kuadrat dengan derajat bebas (n-1).

Selanjutnya untuk sebaran t diperoleh melalui transpormasi dari rasio dua peubah acak yang menyebar normal baku dan menyebar khi-kuadrat. Sedangkan sebaran F diperoleh melalui transpormasi dari rasio dua peubah acak yang keduanya menyebar khi-kuadrat.

Berdasarkan uraian di atas, keempat sebaran tersebut berasal dari sebaran normal, sehingga dalam penggunaannya sebelum dilakukan analisis data perlu dicek apakah data hasil pengamatan menyebar normal atau tidak.

IV KASUS : KAJIAN RESIDUAL DALAM MENGATASI DATA PENCILAN (OUTLIER) PADA PENGGUNAAN REGRESI LINIER SEDERHANA

Pengaruh Data Pencilan (Outlier) dan Residual Terhadap Kualitas Informasi

Di atas telah diuraikan bahwa untuk mendapatkan informasi yang berkualitas, diantaranya kita harus mengikuti prosedur atau tatacara yang berlaku pada setiap tahapan pengelolaan data. Pada tahap pengolahan dan analisis data hasil penelitian terdapat berbagai aturan yang harus dipenuhi, termasuk ketentuan tentang sebaran dari data tersebut. Jadi sebelum kita melakukan tahapan ini perlu ditelaah bentuk sebarannya, apakah sudah memenuhi asumsi yang diperlukan atau tidak. Asumsi sebaran yang diperlukan dalam analisis regresi adalah asumsi normalitas, sehingga data hasil penelitian harus menyebar normal. Yang diasumsikan menyebar normal adalah bagian residual (e_i), jadi residual dalam analisis regresi mempunyai peran yang cukup besar dan dianggap penentu pada saat melakukan prediksi dengan model regresi. Hal ini jelas karena pada saat melakukan pendugaan koefisien regresi digunakan metode kuadrat terkecil yang ditempuh dengan meminimalkan jumlah kuadrat dari residual atau meminimalkan $\sum e_i^2$. Ini dilakukan agar untuk mendapatkan bentuk model regresi yang baik dan handal haruslah berasal dari pengamatan-pengamatan dengan residu terkecil. Jadi jelas bahwa residual merupakan bagian yang menentukan dalam memilih model regresi yang akan digunakan.

Untuk mendapatkan informasi yang berkualitas, deteksi data pencilan melalui penelaahan residual dan uji normalitas merupakan hal yang mutlak perlu dilakukan jika kita akan melakukan analisis data dengan menggunakan regresi linier sederhana. Dalam bagian ini akan dibahas dengan menelaah nilai residual untuk mendeteksi adanya gejala data pencilan (outlier) dan uji normalitas terhadap nilai residualnya.

Residual atau sisaan dalam regresi linier sederhana merupakan selisih dari nilai prediksi dan nilai sebenarnya (actual) atau $e_i = Y_i - (a + b X_i)$. Jika nilai pengamatan terletak dalam garis regresi maka nilai residunya sama dengan nol, sehingga nilai residu merupakan penyimpangan hasil prediksi dari nilai actual. Jadi jika total jarak atau nilai mutlak dari residu atau $\sum |e_i| = 0$ berarti semua nilai pengamatan terletak pada garis regresi. Makin besar total jarak maka makin jauh regresi itu dari nilai actual, atau nilai residunya makin besar dan garis regresi kurang tepat digunakan untuk memprediksi baik secara interpolasi ataupun ekstrapolasi. Yang diharapkan adalah sebaliknya yaitu total residu semakin kecil sehingga garis regresi cukup handal untuk digunakan.

Nilai residu akan makin besar jika terdapat data pencilan dan dapat menurunkan nilai koefisien regresi atau

koefisien korelasi. Untuk mengamati nilai residu ini dapat dilakukan dengan memplot (menggambarkan) residu tersebut dalam suatu diagram (grafik). Di samping itu dapat juga dilihat nilai ragamnya, jika nilai ragamnya kecil maka variasi residu tidak besar, tetapi jika sebaliknya maka variasi residu membesar.

Untuk menunjukkan model regresi itu layak atau tidak digunakan maka perlu dicek persyaratan yang diperlukan apakah sudah semuanya dipenuhi atau belum. Diantaranya adalah anggapan tentang residu yang menyebar normal. Jika ini dipenuhi maka jelas total residunya sama dengan nol atau $\sum e_i = 0$. Jadi apabila nilainya jauh dari nol kita harus curiga dan perlu dicek (uji normalitas dan deteksi data pencilan serta upaya lainnya).

Apabila terdapat data pencilan dalam hasil pengamatan berakibat data cenderung tidak menyebar normal, di samping itu dengan adanya data pencilan kisaran data hasil pengamatan akan melebar dan tentunya variasi data akan bertambah besar sehingga upaya meminimalkan $\sum e_i^2$ tidak berhasil (tidak tercapai). Dampak lain yang terjadi adalah koefisien regresi dan korelasi nilainya akan menurun (mengecil) sehingga dapat menurunkan keterandalan dari pada model regresi yang dihasilkan dan lemah untuk digunakan sebagai prediktor. Hal ini dapat juga ditunjukkan pada saat proses penentuan koefisien regresi dan korelasi sebagai berikut :

- Koefisien regresi ditentukan dengan mencari rasio dari $Cov(X,Y)$ dan $(n-1)*S_x^2$ atau

$$b = \frac{n \sum_{i=1}^n X_i Y_i - (\sum_{i=1}^n X_i Y_i)^2}{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2}$$

- Koefisien korelasi ditentukan dengan mencari rasio dari $Cov(X,Y)$ dan $((n-1)*S_x*S_y)$ atau

$$r_{xy} = \frac{n \sum_{i=1}^n X_i Y_i - (\sum_{i=1}^n X_i)(\sum_{i=1}^n Y_i)}{\sqrt{[n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2] \{n \sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2\}}}$$

- Jika terdapat data pencilan, jelas nilai-nilai dari S_x , S_y dan juga residual e_i akan meningkat sehingga berakibat untuk menurunkan koefisien regresi dan koefisien korelasi.
- Dengan rendahnya nilai koefisien regresi dan korelasi dengan sendirinya dapat menurunkan kualitas dari model/garis regresi yang dihasilkan, sehingga perlu dicari model lain yang lebih cocok dengan kondisi yang diamati atau melakukan transformasi terhadap data tersebut.
- Penurunan nilai koefisien regresi dan korelasi akibat adanya data pencilan dapat berakibat pada penurunan kualitas informasi yang dihasilkan karena model regresi yang dihasilkan kurang baik untuk digunakan sebagai prediktor. Jadi upaya untuk menanggulangi data pencilan perlu dilakukan jika data tersebut tetap digunakan (tidak dibuang) dalam melakukan analisis.

Upaya Untuk Menanggulangi Data Pencilan

Deteksi Keberadaan Data Pencilan

Kehadiran data pencilan dapat membuat kualitas garis regresi menjadi rendah. Oleh karena itu kehadirannya perlu dideteksi diantaranya dengan cara sebagai berikut.

- Hitunglah residu untuk setiap $i=1,2,\dots,n$, maka akan kita peroleh harga-harga residu e_1, e_2, \dots, e_n .
- Ambil harga mutlak $|e_i|$; $i = 1,2,\dots,n$, kemudian urutkanlah dari yang terbesar hingga terkecil, e_{maks} menyatakan harga mutlak residu yang terbesar.
- Tentukan median M dari e_1, e_2, \dots, e_n .
- Hitung $Q = \sum_{i=1}^n |e_i - M|$
- Hitung $Q_{sisal} = Q - |e_{maks} - M|$
- Hitung $T = Q_{sisal}/Q$
- Bandingkan harga statistik penguji T dengan titik kritis untuk $k=1$ dan tingkat keberartian 0,01 atau 0,05 atau 0,10.
- Jika harga T melebihi titik kritis, maka data yang memberikan e_{maks} adalah bukan data pencilan.

Untuk lebih jelasnya dapat dilihat pada : http://www.math.itb.ac.id/~ma291/sas_rls.htm

Cara lain untuk mendeteksi adanya gejala pencilan dapat dilakukan dengan satu metode yang lebih sederhana, yaitu dengan menggunakan sebaran tengah d_Q (deviasi kuartil) sebagai berikut :

- Tentukan nilai kuartil atas (Q_A) kuartil bawah (Q_B) dan hitung besarnya $d_Q = Q_A - Q_B$
- Tentukan batas bawah pencilan $BBP = Q_B - (1,5)d_Q$.
- Tentukan batas atas pencilan $BAP = Q_A + (1,5)d_Q$.
- Untuk mendeteksi pencilan dilakukan dengan membandingkan nilai data : (jika data pengamatan lebih kecil dari BBP atau lebih besar dari BAP maka pengamatan tersebut adalah pencilan)
- Jadi $BAP - BBP = 4d_Q$. Mengapa diambil $4 d_Q$? Hal ini dapat dijelaskan melalui bentuk sebaran ideal, yakni normal. Dalam keadaan ideal ini, pengambilan $4 d_Q$ berarti bahwa tingkat keyakinan (probability) terjadinya outliers adalah sebesar 0,007 atau 0,7% atau kira-kira 1%.
- Untuk lebih jelasnya dapat dilihat pada <http://www.math.itb.ac.id/~ma291/memusatkandata.htm>

Dengan cara kedua data pengamatan diganti dengan residual maka setelah mendapatkan residual dari semua pengamatan selanjutnya tentukan nilai kuartil atas Q_A dan kuartil bawah Q_B dari nilai mutlak residual atau $|e_i|$ serta tentukan penyimpangannya $d_Q = Q_A - Q_B$. Seperti di atas tentukan BBP dan BAP dan untuk mendeteksi pencilan gunakan residu (bukan data pengamatan). Ketentuannya adalah : (jika nilai residu lebih kecil dari BBP atau lebih besar dari BAP maka data pengamatan yang bersangkutan adalah pencilan)

Prosedur Uji Normalitas Dalam Analisis Regresi

Dalam analisis regresi dan korelasi yang diasumsikan menyebar normal adalah residual e_i sehingga ada suatu pemikiran yang perlu di uji kenormalannya adalah residual, tetapi banyak juga yang melakukannya langsung terhadap data pengamatan, tepatnya terhadap peubah respon (peubah tak bebas Y). Keduanya sama saja karena berdasarkan sifat dari peubah acak yang menyebar normal, jika peubah tersebut menyebar normal maka kombinasi liniernya juga akan menyebar normal. Jadi jika residual menyebar normal maka Y juga menyebar normal karena Y adalah kombinasi linier dari residual e_i atau $Y_i = a + b X_i + e_i$.

Di samping itu, dalam melakukan uji koefisien regresi atau koefisien korelasi biasa digunakan sebaran t atau untuk pengujian secara simultan digunakan sebaran f . Kedua sebaran tersebut diturunkan/berasal dari sebaran normal. Atau untuk lebih jelasnya sebaran t dibangkitkan dari rasio dua peubah acak yang menyebar normal baku dan sebaran χ^2 -kuadrat, sedangkan sebaran f dibangkitkan dari rasio dua peubah acak yang masing-masing menyebar χ^2 -kuadrat. Sebaran χ^2 -kuadrat sendiri berasal dari sebaran normal baku (sebaran normal baku jelas berasal dari sebaran normal).

Berdasarkan informasi di atas, jika kita menghendaki hasil kajian yang syahih dan terandalkan maka uji normalitas terhadap residual atau langsung pada peubah respon jelas perlu dilakukan sebelum analisis data.

Untuk mengecek apakah hasil pengamatan menyebar normal atau tidak dapat dilakukan dengan berbagai cara seperti : dengan uji histogram, uji normal P Plot, uji χ^2 Square, Skewness dan Kurtosis atau uji Kolmogorov Smirnov. Tidak ada metode yang paling baik atau paling tepat. Tipsnya adalah bahwa pengujian dengan metode grafik sering menimbulkan perbedaan persepsi di antara beberapa pengamat, sehingga penggunaan uji normalitas dengan uji statistik bebas dari keragu-raguan, meskipun tidak ada jaminan bahwa pengujian dengan uji statistik lebih baik dari pada pengujian dengan metode grafik.

Transformasi Data Untuk Mengatasi Pencilan

Jika pencilan (outlier) ternyata ada dalam hasil pengamatan dan pencilan itu akan digunakan dalam analisis data maka perlu dicari cara untuk mengatasinya agar pencilan itu tidak mengganggu dan didapat hasil yang lebih baik. Salah satu cara untuk mengatasi pencilan atau pemenuhan asumsi normalitas dapat ditempuh dengan transformasi terhadap data hasil pengamatan.

Dalam hal demikian, peubah tak bebas Y dan atau peubah bebas X perlu ditransformasikan, dengan caranya sebagai berikut :

1. Mengamati *stem-leaf* (dahan -daun) sari numerik, dan *box plot* dari data X dan data Y

2. Pilihlah transformasi yang sesuai untuk Y dan atau untuk X. Misalkan hasil transformasi dari Y dan dari X berturut-turut adalah Z dan W
3. Tentukan regresi linear dari Z terhadap W
4. Bila regresi dari Z terhadap W membarikan harga R^2 yang sudah memuaskan, maka proses pemodelan selesai. Bila belum, maka ulangi langkah 2 dan 3.

Cara transformasi yang dapat dilakukan diantaranya adalah dengan : logaritma natural, akar kuadrat, inverse, atau bentuk lain yang sesuai dengan kondisi data hasil pengamatan.

V KESIMPULAN

Berdasarkan uraian dari penulisan ini, dapat disampaikan beberapa kesimpulan dan saran untuk mendapatkan informasi yang berkualitas adalah sebagai berikut :

- Dalam analisis regresi residual merupakan bagian yang menentukan layak tidaknya model tersebut digunakan, karena jika jumlah nilai mutlak residunya jauh dari nol dan juga jumlah kuadrat residunya besar sekali ini menunjukkan bahwa model regresi itu lemah dan kurang layak bila digunakan untuk memprediksi.
- Jika terdapat data pencilan (outlier) nilai residu akan makin besar dapat memperkecil/menurunkan nilai koefisien regresi dan juga nilai korelasi sehingga jika data itu mau digunakan maka perlu upaya untuk mengatasinya yang diantaranya dilakukan dengan menggunakan peubah yang sudah ditransformasi. Cara transformasi ini digunakan selain untuk menanggulangi data pencilan juga untuk mengupayakan agar terpenuhinya asumsi normalitas, karena jika terdapat pencilan maka data hasil pengamatannya tidak menyebar normal.
- Untuk mengamati nilai residu ini dapat dilakukan dengan memplot (menggambar) residu tersebut dalam suatu diagram (grafik). Di samping itu dapat juga

dilihat nilai ragamnya, jika nilai ragamnya kecil maka variasi residu tidak besar, tetapi jika sebaliknya maka variasi residu membesar.

- Dalam melakukan uji koefisien regresi atau koefisien korelasi biasa digunakan sebaran t atau untuk pengujian secara simultan digunakan sebaran f. Kedua sebaran tersebut diturunkan/berasal dari sebaran normal. Jika menghendaki hasil kajian yang syahih dan terandalkan maka uji normalitas jelas perlu dilakukan sebelum analisis data dilakukan dan dapat dilakukan terhadap residual atau langsung pada peubah respon.
- Keberartian suatu informasi ditentukan oleh tingkat diperlukannya informasi tersebut. Secanggih dan sebaik apapun data, tidak ada artinya jika data tersebut tidak diperlukan.
- Untuk menunjang perkembangan teknologi informasi yang cukup pesat diperlukan informasi tentang data atau informasi yang diperlukan oleh pengguna/user (database tentang kebutuhan informasi).

DAFTAR PUSTAKA

- [1]. Gaspersz Vincent (1991). Teknik Penarikan Contoh Untuk Penelitian Survei, Tarsito, Bandung.
- [2]. Hogg, R.V. and A.T. Craig. (1995). Introduction to Mathematical Statistics. Prentice Hall. Singapore
- [3]. Sudjana, 2002, Metode Statistika; Tarsito; Bandung
- [4]. http://www.math.itb.ac.id/~ma291/sas_rls.htm
- [5]. <http://mrpendi.wordpress.com/2008/02/13/23/>

