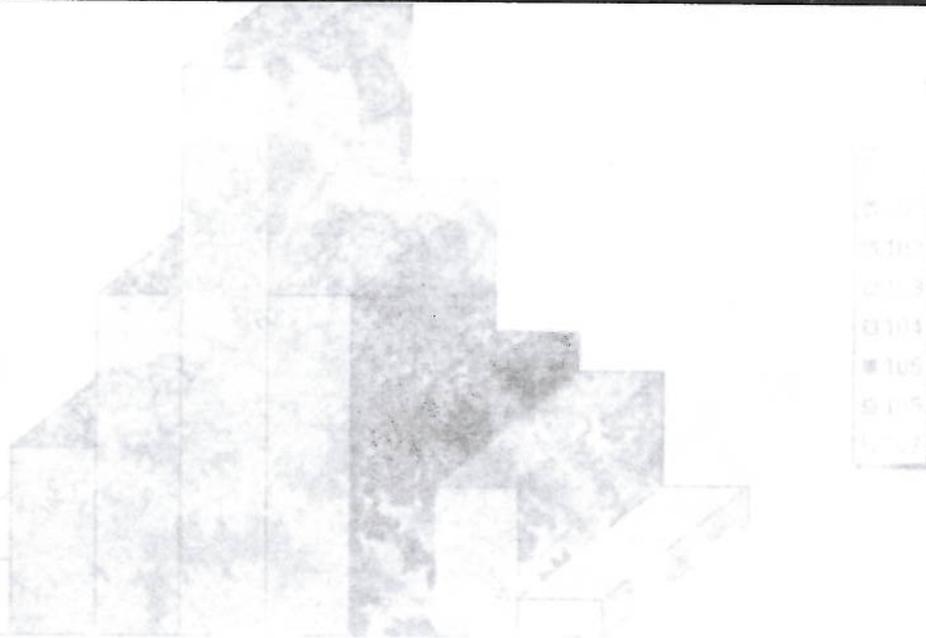


Mat Stat

Vol. 12 No. 2 Juli 2012

MATEMATIKA, STATISTIKA



| | | | | | |
|--------------------|---------|-------|--------------|----------------------|----------------------|
| Jurnal Mat Stat | Vol. 12 | No. 2 | Hlm. 103-192 | Jakarta Juli 2012 | ISSN: 1412 - 1220 |
|--------------------|---------|-------|--------------|----------------------|----------------------|

JURNAL ILMIAH Mat Stat

Vol. 12 No. 2 Juli 2012

Jurnal Ilmiah MATEMATIKA, STATISTIKA ($Mat Stat$) diterbitkan oleh Direktorat Riset dan HKI BINUS University memfokuskan pada artikel ilmiah berupa hasil penelitian, pengembangan ilmu dalam bidang matematika, statistika, dan aplikasinya. Tujuan publikasi jurnal ini adalah penyebarluasan hasil kajian dan penelitian untuk meningkatkan penguasaan, pengembangan, penerapan matematika dan statistika dalam bidang teknik, komputer dan industri, serta aplikasi lainnya.

DAFTAR ISI

| | |
|--|---------|
| <p>^{D0205} ^{D0434} Wikaria Gazali; Haryono Soeparno; Jenny Ohliati Penerapan Metode Konvolusi dalam Pengolahan Citra Digital <i>(Implementation of Convolution Model in Digital Image Processing)</i>.....</p> | 103-113 |
| <p>^{D3611} ^{D1347} Margaretha Ohlyver; Heruna Tanty* Pendeteksian Outlier pada Model Regresi Ganda: Studi Kasus Tingkat Penghunian Kamar Hotel di Kendari <i>(Outlier Detection on Double Regression Model: A Case Study on the Level of Room Occupacy in a Hotel in Kendari)</i>.....</p> | 114-122 |
| <p>^{D4446} Rokhāna Dwi Bakti * Prediksi dan Interpolasi melalui Ordinary Kriging: Studi Kasus Kemiskinan di Provinsi Jawa Timur <i>(Prediction and Interpolation through Ordinary Kriging: A Case Study on Poverty in East Java)</i>.....</p> | 123-132 |
| <p>^{D3539} Viska Noviantri Solusi Penyebaran Panas pada Batang Konduktor menggunakan Metode Crank-Nicholson <i>(Solution for Heat Spreading on Conductor Bar using Cranck-Nicholson Method)</i>.....</p> | 133-142 |
| <p>^{D1103} Ngarap Im. Manik; Manal Penggunaan Model Fraktal untuk Pengembangan Motif Ulos <i>(The Use of Fractal Model for Ulos Pattern Development)</i>.....</p> | 143-151 |
| <p>^{D3535} Siti Komsiyah * Model Petri Net Tak Berwaktu pada Sistem Produksi (Batch Plant) dan Simulasinya dengan Pipe2 <i>(Timeless Petri Net Model on Production System (Batch Plant) and Its Simulation using Pipe2)</i>.....</p> | 152-164 |
| <p>Muhammad Nur; Ahmad Fadhilah; Ahmad Suseno; Heri Sutanto Mobilitas ion-ion Ar⁺, OH⁻, H⁺, CO₂, O₂ dan Laju Aliran Angin Ion dalam Plasma Korona pada Tekanan Atmosfer <i>(Mobility of Ion Ar⁺, OH⁻, H⁺, CO₂, O₂ and Rate of Wind Current in Corona Plasma on Athmosphere Pressure)</i>.....</p> | 165-175 |
| <p>Yanita; Abdul Ghafur Ahmad Penggunaan Transformasi Tietze Dalam Perhitungan Generator Modul Homotopi Kedua <i>(The Use of Tietze Transformation in Generator Calculation of Second Homotopy Module)</i>....</p> | 176-187 |
| <p>Indeks.....</p> | 189-192 |

PENDETEKSIAN OUTLIER PADA MODEL REGRESI GANDA: STUDI KASUS TINGKAT PENGHUNIAN KAMAR HOTEL DI KENDARI

Margaretha Ohyver; Heruna Tanty

Mathematics & Statistics Department, School of Computer Science, Binus University
Jl. K.H. Syahdan No. 9, Palmerah, Jakarta Barat 11480
mohyver@binus.edu; herunatanty@yahoo.com

ABSTRACT

Regression is a statistic method to model the correlation between responding variable and free variable. Several things need more attention when making model using regression, one of them is the outlier. Found in the data, outlier will produce less-optimal model. Therefore, outlier detection is needed. This research is objected to detect any outlier as well as identify its influence towards the data used. The outlier detection uses residual value and residual studentized. The data used are about the level of room occupancy in a hotel in Kendari. Two observations are obtained in this research. In addition, the outlier found causes small R^2 value.

Keywords: outlier, regression, residual, residual studentized

ABSTRAK

Regresi merupakan salah satu metode statistika yang dapat digunakan untuk memodelkan hubungan antara variabel respon dan variabel bebas. Ada beberapa hal yang menjadi perhatian ketika akan membuat pemodelan dengan menggunakan regresi, salah satunya adalah outlier. Adanya outlier pada data akan mengakibatkan model regresi yang diperoleh kurang baik. Oleh sebab itu perlu dilakukan pendeteksian outlier pada data. Tujuan dari penelitian ini adalah mendeteksi outlier dan mengetahui pengaruh outlier terhadap data yang digunakan. Pendeteksian outlier dilakukan dengan menggunakan nilai residual dan residual studentized. Data yang digunakan adalah data tingkat penghunian kamar hotel di Kendari. Hasil yang diperoleh adalah ada dua pengamatan yang dideteksi sebagai outlier. Adanya outlier ini menyebabkan nilai R^2 kecil.

Kata kunci: outlier, regresi, residual, residual studentized

PENDAHULUAN

Hubungan antara dua variabel (variabel bebas X dan variabel tak bebas Y) dapat dinyatakan dalam suatu persamaan regresi. Persamaan tersebut dapat berupa persamaan regresi sederhana maupun regresi ganda. Apabila persamaan regresi hanya memuat satu variabel bebas maka model regresinya disebut model regresi sederhana. Apabila persamaan regresi memuat lebih dari satu variabel bebas maka model regresinya disebut model regresi ganda (Ohyver, 2012: 178).

Misalkan terdapat variabel respon Y dan variabel bebas X_1, X_2, \dots, X_p . Hubungan antara Y dan X_1, X_2, \dots, X_p dapat dilihat pada persamaan (1).

$$y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon, \quad (1)$$

dengan $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ adalah koefisien-koefisien regresi dan ε adalah error (Chatterjee dan Hadi, 2006).

Jika terdapat n pengamatan, persamaan (1) dapat dituliskan sebagai berikut.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (2)$$

dengan y_i adalah nilai variabel respon Y ke- i , $x_{i1}, x_{i2}, \dots, x_{ip}$ adalah nilai variabel-variabel bebas untuk unit ke- i , dan ε_i adalah error ke- i .

Persamaan (2) dapat dituliskan dalam bentuk matriks. Sebelum menuliskan dalam bentuk matriks perlu didefinisikan matriks-matriks berikut.

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Notasi matriks dari persamaan (2) dapat dilihat pada persamaan (3)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (3)$$

dengan \mathbf{y} adalah vektor respon yang berukuran $n \times 1$; $\boldsymbol{\varepsilon}$ adalah vektor galat yang berukuran $n \times 1$; $\boldsymbol{\beta}$ adalah vektor parameter regresi yang akan diduga dan berukuran $(p+1) \times 1$; \mathbf{X} adalah matriks skalar yang berukuran $n \times (p+1)$ dan berpangkat penuh (Neter, *et al.*, 1990).

Nilai koefisien regresi dugaan $\hat{\boldsymbol{\beta}}$ dapat diperoleh dengan menggunakan metode penaksiran parameter seperti metode kuadrat terkecil (*least squares method*). Teorema 1 akan memberikan cara untuk memperoleh $\hat{\boldsymbol{\beta}}$.

Teorema 1

Misalkan $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ dengan \mathbf{X} adalah matriks berpangkat penuh yang berukuran $n \times (p+1)$, $\boldsymbol{\beta}$ adalah vektor dari parameter regresi yang akan diduga dan berukuran $(p+1) \times 1$, dan $\boldsymbol{\varepsilon}$ adalah vektor acak yang berukuran $n \times 1$ dengan nilai harapan $\mathbf{0}$ dan ragam $\sigma^2 \mathbf{I}$. Penduga kuadrat terkecil (*least squares estimator*) bagi $\boldsymbol{\beta}$, dinyatakan dengan $\hat{\boldsymbol{\beta}}$, adalah:

$$\hat{\beta} = (X'X)^{-1} X'y .$$

Untuk memperoleh suatu model regresi yang baik maka model tersebut harus memenuhi asumsi-asumsi yang berlaku, di antaranya asumsi normalitas, homoskedastisitas, dan autokorelasi. Jika asumsi-asumsi yang ada dalam regresi telah terpenuhi maka ada satu hal lagi yang perlu diperhatikan yaitu adanya outlier dan pengamatan berpengaruh pada data. Sehingga sebelum menerapkan metode-metode statistika pada data maka sangat penting untuk dilakukan pendeteksian outlier (Derquenne, 1993: 323).

Identifikasi outlier dan pengamatan berpengaruh pada model regresi didasarkan pada asumsi bahwa model regresi yang diperoleh sudah tepat. Hal ini berarti model regresi yang telah dipilih telah cukup menggambarkan hubungan antara variabel respon dan variabel independen. Jika model regresi telah ditentukan, sebagian besar data seharusnya mendekati garis regresi atau *hyperplane*. Titik-titik data yang berada jauh dari garis regresi atau *hyperplane* mungkin bukan titik-titik data ideal bagi model yang dipilih dan dapat diidentifikasi sebagai outlier.

Berbagai komunitas *scientific* mempunyai definisi yang berbeda-beda mengenai outlier. Definisi tersebut ada yang berdasarkan distribusi, jarak, dan densitas (Zhu et al, 2011: 218). Konsekuensinya adalah pendekatan yang diberikan berbeda-beda pula. Apa yang bagi seseorang adalah outlier bisa menjadi bukan outlier bagi orang lain.

Outlier adalah titik data yang secara statistik berada jauh dari model yang dipilih jika peneliti percaya bahwa model yang dipilih adalah model yang benar. Pengamatan berpengaruh adalah outlier yang memiliki pengaruh yang relatif besar terhadap koefisien regresi dugaan. Pengaruh yang dimaksud adalah penambahan atau pengurangan pengamatan tersebut akan menyebabkan perubahan yang tidak biasa terhadap pendugaan parameter regresi.

Penelitian ini dilakukan untuk membahas permasalahan sebagai berikut. Pertama, apakah ada pengamatan yang terdeteksi sebagai outlier pada data tingkat penghunian kamar hotel di Kendari? Kedua, apakah pengeluaran pengamatan yang terdeteksi sebagai outlier berpengaruh terhadap nilai R^2 ? Berdasarkan permasalahan tersebut maka tujuan penelitian ini adalah mendeteksi outlier dan mengetahui pengaruh outlier terhadap data tingkat penghunian kamar hotel di Kendari.

Pendeteksian Outlier pada Regresi Linear Ganda

Jika terdapat selisih yang besar antara nilai \hat{y}_i dan y_i maka pengamatan ke- i tersebut merupakan pengamatan yang berpotensi outlier. Tujuan dari pendeteksian outlier dalam analisis regresi adalah mengeliminasi pengamatan-pengamatan yang memiliki residual-residual yang relatif besar.

Definisi 1

Misalkan s adalah *mean square error* dari model regresi. Residual *standardized* didefinisikan sebagai berikut.

$$z_i = \frac{y_i - \hat{y}_i}{s} \quad (4)$$

Persamaan (4) disebut juga residual *normalized* atau z-skor dari residual (Yan dan Su, 2009).

Definisi 2

Misalkan s adalah *mean square error* dari model regresi. Residual *studentized* didefinisikan sebagai berikut.

$$z_i = \frac{y_i - \hat{y}_i}{s\sqrt{1-h_{ii}}} \quad (5)$$

di mana h_{ii} adalah elemen diagonal ke- i dari matriks HAT (disebut juga leverage).

Definisi 3

Jika residual pengamatan tiga kali lebih besar dari standar deviasi (atau *residual standardized* lebih besar dari 3), pengamatan tersebut adalah outlier.

METODE

Data yang akan digunakan adalah data sekunder. Data ini diperoleh dari katalog Badan Pusat Statistik (BPS) Sultra tahun 2011. Ada 90 hotel/penginapan yang menjadi sampelnya. Hotel/penginapan tersebut berada di wilayah Kendari. Variabel-variabel yang digunakan terdapat pada Tabel 1. Terdapat tujuh variabel bebas (X) dan 1 variabel tak bebas atau variabel respon (Y). Pendeteksian outlier pada penelitian ini akan dilakukan dengan menggunakan nilai residual dan residual *studentized*.

Tabel 1 Variabel Penelitian

| No | Nama variabel | Simbol |
|----|--|-----------|
| 1. | Usia hotel/penginapan | (X_1) |
| 2. | Tarif minimal hotel/penginapan | (X_2) |
| 3. | Tarif maksimal hotel/penginapan | (X_3) |
| 4. | Fasilitas yang dimiliki hotel atau penginapan | (X_4) |
| 5. | Jumlah tenaga kerja yang dimiliki hotel/penginapan | (X_5) |
| 6. | Jumlah kamar yang dimiliki hotel/penginapan | (X_6) |
| 7. | Jumlah tempat tidur yang dimiliki hotel/penginapan | (X_7) |
| 8. | Jumlah tamu | (Y) |

HASIL DAN PEMBAHASAN

Jasa perhotelan merupakan bagian dari industri pariwisata yang sangat penting dalam pembangunan ekonomi nasional maupun regional (BPS Provinsi Sultra, 2011). Hal ini tidak terbantahkan mengingat peranannya yang besar terutama dalam hal ketenagakerjaan dan sumber pendapatan daerah. Oleh karena itu, sektor ini perlu mendapat perhatian khusus dari pemerintah. Berbagai upaya yang terencana, terpadu, dan efektif perlu dilakukan untuk menunjang kebijakan pemerintah dalam pembangunan di sektor ini. Akan tetapi selain pemerintah, pihak penyedia jasa perhotelan pun harus melakukan pengembangan. Sebab segala upaya yang dilakukan oleh pemerintah tidak akan ada artinya jika pihak perhotelan tidak melakukan pengembangan atau perbaikan dari dalam.

Jumlah hotel/akomodasi di Sultra menunjukkan peningkatan yang cukup baik. Hal ini terlihat dari jumlahnya yang terus meningkat sejak tahun 2006 hingga tahun 2010 dengan peningkatan sebesar 4,65% per tahun. Demikian pula untuk jumlah kamar dalam rentang waktu yang sama dengan rata-rata peningkatan sebesar 9,76%. Di samping itu untuk jumlah tempat tidur meningkat rata-rata sebesar 4,45%. Dari hasil inventarisasi perusahaan akomodasi di Sultra tahun 2010, terdapat 265 buah

perusahaan/usaha akomodasi dengan 3389 kamar dan 4918 tempat tidur. Dari 265 hotel hanya ada 1 (satu) hotel berbintang (bintang satu) dan 264 hotel non bintang. Meskipun jumlah tersebut termasuk kecil jika dibandingkan dengan provinsi-provinsi lain yang ada di Indonesia, tetapi jumlah tersebut akan cukup memberikan banyak pilihan kepada pengunjung. Perusahaan akomodasi tentu tidak ingin kehilangan pengunjung. Oleh karena itu perusahaan perlu mengetahui faktor-faktor yang mempengaruhi jumlah pengunjung hotel. Dengan mengetahui faktor-faktor tersebut maka perusahaan dapat memberikan perhatian khusus sehingga ke depannya jumlah pengunjung dapat bertambah.

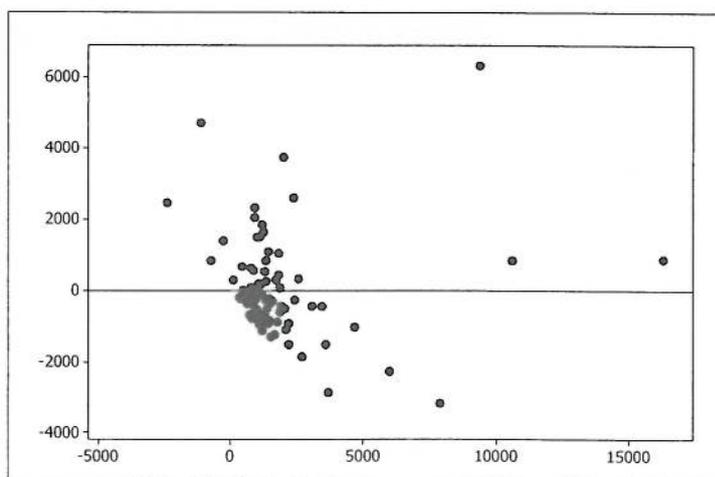
$$y = 667 - 7,8x_1 - 5,50x_2 + 5,37x_3 - 10,9x_4 + 125x_5 + 87,2x_6 - 94x_7. \quad (7)$$

Persamaan regresi untuk data tersebut dapat dilihat pada persamaan (7). Nilai R^2 dan s^2 yang diperoleh adalah 74,2% dan 2121247. Persamaan (7) menghasilkan 3 variabel yang tidak signifikan. Variabel-variabel tersebut adalah usia, tarif minimal, dan fasilitas. Hasil lengkapnya dapat dilihat pada Tabel 2. Nilai residual yang diperoleh digunakan untuk mendeteksi adanya outlier.

Tabel 2 Hasil Regresi dengan 90 Pengamatan

| | Koefisien | Standard error koefisien | T | P value |
|-----------|-----------|--------------------------|-------|---------|
| Konstanta | 667 | 449 | 1,49 | 0,141 |
| (X_1) | -7,8 | 18,08 | -0,43 | 0,665 |
| (X_2) | -5,50 | 3,770 | -1,46 | 0,149 |
| (X_3) | 5,37 | 1,245 | 4,31 | 0,000 |
| (X_4) | -10,9 | 89,70 | -0,12 | 0,903 |
| (X_5) | 125 | 34,01 | 3,67 | 0,000 |
| (X_6) | 87,2 | 35,75 | 2,44 | 0,017 |
| (X_7) | -94 | 20,47 | -4,59 | 0,000 |

Langkah awal dapat dilakukan dengan mendeteksi secara visual. Pendeteksian tersebut dilakukan dengan membuat *scatterplot* antara residual dan nilai dugaan. *Scatterplot* tersebut dapat dilihat pada Gambar 1. Pada gambar terlihat ada beberapa titik data yang terpisah dari kumpulan data. Jika dilakukan pemeriksaan secara visual bisa terjadi perbedaan pandangan. Sebab bisa terjadi apa yang merupakan outlier bagi seseorang bisa jadi bukan merupakan outlier bagi orang lain. Sehingga perlu dilakukan analisis secara statistik.



Gambar 1 *Scatterplot*

Pendeteksian dapat juga dilakukan dengan melihat *Boxplot*. Pada Gambar 2 disajikan *Boxplot* untuk masing-masing variabel. Dapat dilihat bahwa ada perbedaan di masing-masing *Boxplot*. Sehingga akan cukup membingungkan untuk menentukan mana yang akan dikeluarkan seandainya pengamatan tersebut adalah outlier yang berpengaruh terhadap pemodelan data.

Untuk mengetahui adanya outlier pada data yang digunakan maka perhatikan Tabel 3. Berdasarkan Definisi 3 maka akan dibandingkan nilai-nilai residual yang ada pada Tabel 3 dengan nilai $3 \times 1456,45 = 4369,35$. Hasil yang diperoleh adalah ada dua pengamatan yang $> 4369,35$, yaitu pengamatan ke-35 dan ke-86. Selain menggunakan nilai residual, dapat juga digunakan nilai residual *standardized*. Berdasarkan Definisi 3, nilai-nilai yang ada pada Tabel 4 akan dibandingkan dengan nilai 3. Hasil yang diperoleh adalah ada dua pengamatan yang > 3 , yaitu pengamatan ke-35 dan ke-86. Hasil yang diperoleh dalam mendeteksi outlier dengan menggunakan residual dan *residual standardized* menunjukkan hasil yang sama.

Tabel 3 Nilai Residual

| Pengamatan | Residual | Pengamatan | Residual | Pengamatan | Residual |
|------------|----------|------------|----------|------------|----------|
| 1 | -357,26 | 31 | -689,01 | 61 | -3149,32 |
| 2 | -37,90 | 32 | -201,65 | 62 | -2247,74 |
| 3 | 310,81 | 33 | 1080,07 | 63 | -657,53 |
| 4 | -164,72 | 34 | -1490,71 | 64 | -871,76 |
| 5 | 36,22 | 35 | 4726,78 | 65 | 113,42 |
| 6 | -469,15 | 36 | -120,03 | 66 | -100,96 |
| ⋮ | | ⋮ | | ⋮ | |
| 23 | -852,46 | 53 | 329,29 | 83 | -751,68 |
| 24 | -494,73 | 54 | 594,83 | 84 | -496,79 |
| 25 | -953,41 | 55 | 1520,54 | 85 | -675,87 |
| 26 | -887,55 | 56 | 1095,08 | 86 | 6346,25 |
| 27 | -692,70 | 57 | -999,28 | 87 | -1107,66 |
| 28 | -200,99 | 58 | -695,79 | 88 | -817,14 |
| 29 | 560,61 | 59 | -258,06 | 89 | 888,28 |
| 30 | -497,35 | 60 | 704,25 | 90 | -906,03 |

Tabel 4 Nilai Residual *Studentized*

| Pengamatan | Residual | Pengamatan | Residual | Pengamatan | Residual |
|------------|-----------|------------|----------|------------|----------|
| 1 | -0,262935 | 31 | -0,47725 | 61 | -2,51149 |
| 2 | -0,027545 | 32 | -0,14065 | 62 | -2,31127 |
| 3 | 0,227464 | 33 | 0,75678 | 63 | -0,46151 |
| 4 | -0,118708 | 34 | -1,04998 | 64 | -0,62467 |
| 5 | 0,026024 | 35 | 3,42348 | 65 | 0,08407 |
| 6 | -0,341813 | 36 | -0,08309 | 66 | -0,07048 |
| ⋮ | | ⋮ | | ⋮ | |
| 23 | -0,591047 | 53 | 0,22934 | 83 | -0,54102 |
| 24 | -0,347575 | 54 | 0,42872 | 84 | -0,37203 |
| 25 | -0,667658 | 55 | 1,08925 | 85 | -0,47371 |
| 26 | -0,631545 | 56 | 0,75968 | 86 | 5,14993 |
| 27 | -0,491375 | 57 | -0,73915 | 87 | -0,77560 |
| 28 | -0,140429 | 58 | -0,48708 | 88 | -0,56908 |
| 29 | 0,389842 | 59 | -0,17953 | 89 | 0,90141 |
| 30 | -0,349063 | 60 | 0,49215 | 90 | -0,64017 |

Lalu apakah kedua pengamatan yang dideteksi sebagai outlier akan dikeluarkan dari data? Jika pengamatan ke-35 dikeluarkan, akan diperoleh nilai $R^2 = 77,7\%$. Jika pengamatan ke-86 dikeluarkan maka akan diperoleh nilai $R^2 = 75,3\%$. Jika pengamatan ke-35 dan ke-86 dikeluarkan, akan diperoleh nilai $R^2 = 78,4\%$. Berdasarkan nilai R^2 tersebut terlihat bahwa jika pengamatan yang dideteksi sebagai outlier dikeluarkan akan meningkatkan R^2 . Hal tersebut juga terjadi pada nilai R^2 adjusted. Selain itu, pengeluaran kedua pengamatan tersebut berpengaruh terhadap jumlah variabel yang tidak signifikan. Jika kedua pengamatan tersebut dikeluarkan dari data maka jumlah variabel yang tidak signifikan menjadi 2 variabel, yaitu variabel usia dan fasilitas. Berdasarkan hal ini maka pengamatan ke-35 dan ke-86 dipertimbangkan untuk dikeluarkan dari pengamatan.

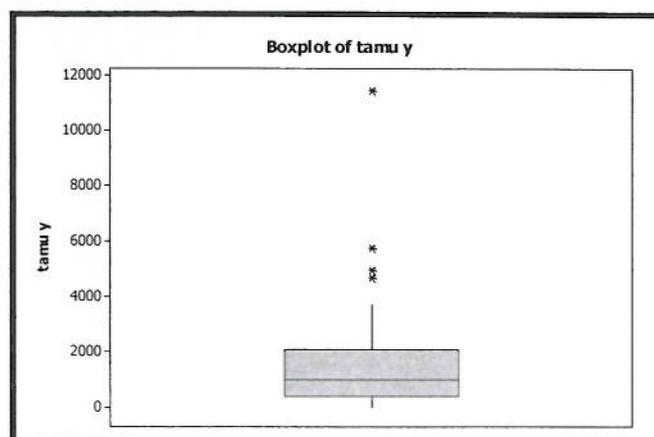
Jika kedua pengamatan tersebut dikeluarkan dari data maka akan diperoleh persamaan regresi yang ditunjukkan oleh persamaan (8). Hasil lengkapnya dapat dilihat pada Tabel 5.

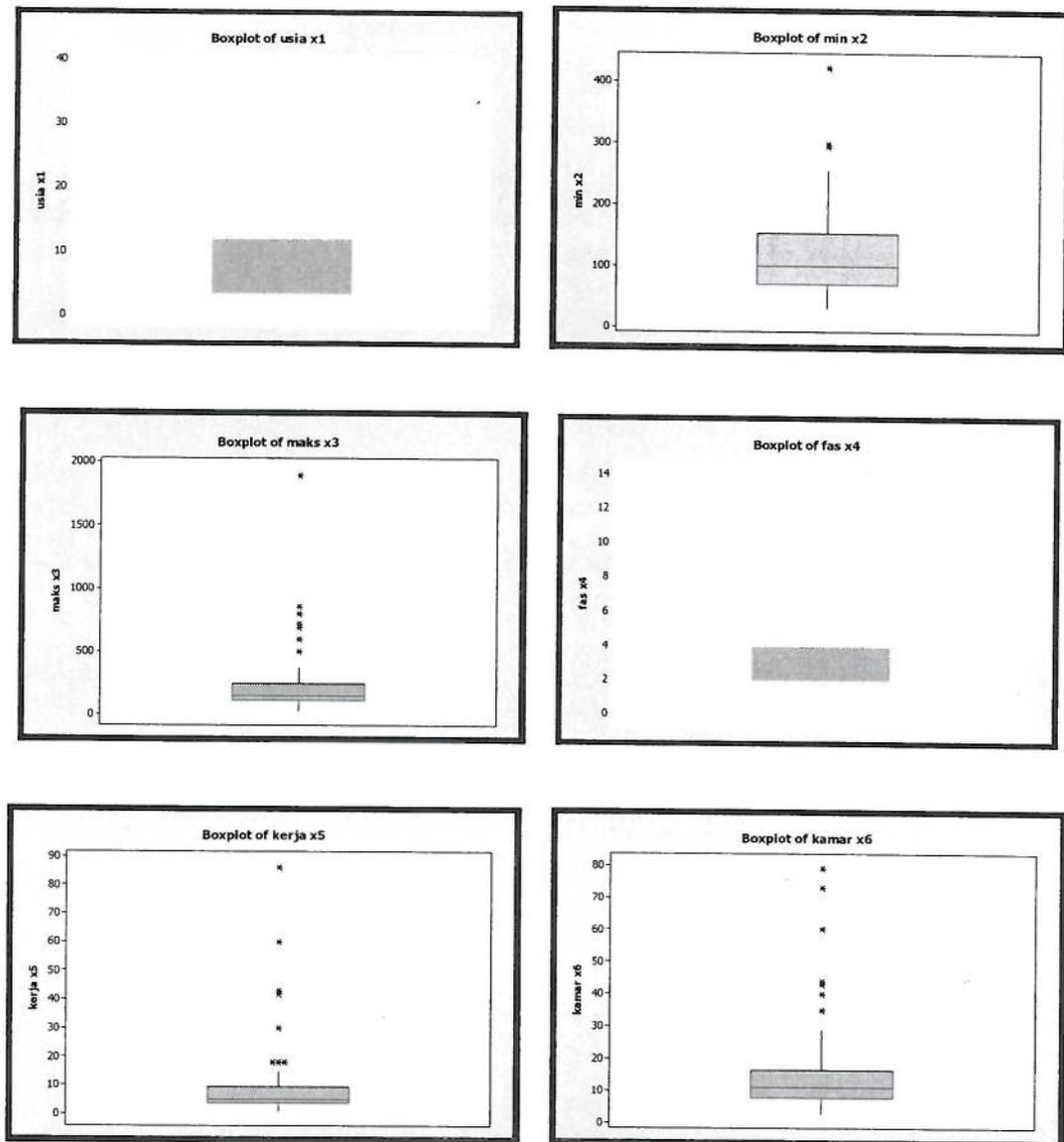
$$y = 920 - 9x_1 - 7,89x_2 + 6,20x_3 - 82,3x_4 + 107x_5 + 64,9x_6 - 71,8x_7 \quad (8)$$

Tabel 5 Hasil Regresi dengan 88 Pengamatan

| | Koefisien | Standard error koefisien | T | P value |
|-----------|-----------|--------------------------|-------|---------|
| Konstanta | 920 | 352,8 | 2,61 | 0,011 |
| (X_1) | -9 | 14,11 | -0,64 | 0,524 |
| (X_2) | -7,89 | 2,949 | -2,68 | 0,009 |
| (X_3) | 6,20 | 0,9722 | 6,38 | 0,000 |
| (X_4) | -82,3 | 70,65 | -1,16 | 0,248 |
| (X_5) | 107 | 27,29 | 3,91 | 0,000 |
| (X_6) | 64,9 | 28,43 | 2,28 | 0,025 |
| (X_7) | -71,8 | 17,94 | -4 | 0,000 |

Pengamatan ke-35 dan ke-86 masih dalam tahap dipertimbangkan untuk dikeluarkan. Hal ini terjadi sebab masih perlu dilakukan penelitian lebih lanjut. Ada beberapa hal yang perlu dipertimbangkan sebelum pengamatan tersebut dikeluarkan. Pertama, jika dilihat dari nilai variabel bebas dan variabel respon terlihat bahwa masih ada pengamatan lain yang nilainya lebih besar dari pengamatan ke-35 dan ke-86, misalnya pengamatan ke-43 dan ke-89. Kedua, Pengeluaran suatu pengamatan dari data sebaiknya tidak langsung dilakukan hanya berdasarkan pendeteksian outlier. Sebaiknya dilanjutkan apakah pengamatan yang dideteksi sebagai outlier tersebut berpengaruh terhadap model atau tidak.





Gambar 2 *Boxplot* variabel bebas dan variabel respon.

SIMPULAN

Berdasarkan penelitian yang telah dilakukan, disimpulkan sebagai berikut. Pertama, pendeteksian outlier dapat dilakukan dengan melihat nilai residual dan residual *studentized*. Kedua, pengamatan yang dideteksi sebagai outlier pada data tingkat penghunian kamar hotel di Kendari adalah pengamatan ke-35 dan ke-86. Ketiga, pengeluaran pengamatan yang dideteksi sebagai outlier dari data akan meningkatkan R^2 , R^2 *adjusted*, jumlah variabel yang signifikan.

DAFTAR PUSTAKA

- Chatterjee, S., and Hadi, A. S. (2006). *Regression Analysis by Example* (4th ed). New Jersey: John Wiley and Sons.
- Derquenne, C. (1993). Outlier Detection Before Running Statistical Methods. *Theory of Probability and its Applications*, 37(2): 323-4. doi:10.1137/1137066.
- Neter, J., Wasserman, W., and kutner, M. H. (1990). *Applied Linear Statistical Models* (3rd ed). Illinois: Irwin.
- Ohyver, M. (2012). Pemodelan Principal Component Regression dengan Software R. *Comtech*, 3(1): 177-185.
- Yan Xin., and Xiao Gang Su. (2009). *Linear Regression Analysis*. Singapore: World Scientific.
- Zhu, C., Kitagawa, H., Papadimitriou, S., & Faloutsos, C. (2011). Outlier Detection by Example. *Journal of Intelligent Information Systems*, 36(2): 217-247. doi:10.1007/s10844-010-0128-1