UJM 6(2) (2017)
Unnes Journal of Mathematics
http://journal.unnes.ac.id/sju/index.php/ujm

BATASAN PRASYARAT UJI NORMALITAS DAN UJI HOMOGENITAS PADA MODEL REGRESI LINEAR

Atmira Qurnia Sari✉, Y.L. Sukestiyarno, Arief Agoestanto

Jurusan Matematika, FMIPA, Universitas Negeri Semarang, Indonesia
Gedung D7 Lt.1, Kampus Sekaran Gunungpati, Semarang 50299

Info Artikel

Sejarah Artikel:
Diterima Agustus 2016
Disetujui September 2016
Dipublikasikan Nopember 2017

Keywords:
Model Linear;
Uji Normalitas;
Uji Homogenitas;
Uji ;
Uji F.

Abstrak

Setiap suku galat diasumsikan mempunyai ragam yang sama, sehingga respons Y_i mempunyai ragam yang sama karena sebaran peluang bagi Y mempunyai ragam yang sama, tidak tergantung pada nilai peubah bebas X. ε mempunyai distribusi sedangkan x tidak, Y juga mempunyai distribusi yang sesuai dengan ε yaitu Y_i. Asumsi galat berdistribusi normal dan homogen berdampak pada variabel dependen (Y) maka yang diuji normalitas dan homogenitas adalah variabel Y dan variabel independen (X) diasumsikan bukan variabel acak. Kekekaran uji t dan uji F terhadap pelanggaran normalitas dan homogenitas data ditunjukkan melalui simulasi data bangkitan dari program R dengan kriteria tidak normal dan heterogen. Data dilakukan uji t dan uji F sehingga diperoleh simpulan data tetap berditribusi t dan berdistribusi F serta nilai p signifikan. Diperoleh simpulan uji t dan uji F terbukti kekar terhadap ketidaknormalan dan heterogenitas data.

Abstract

Each error rate is assumed to have the same variance, so Y_i response have the same variety as the distribution of probability for Y has the same variety, it does not depend on the value of free variable X. ε have the distribution whereas x is not, then Y also has a distribution corresponding to ε that is Y_i. Assumption of normally distributed errors and homogeneous impact on the dependent variable (Y) then which is tested normality and homogeneity are variable Y and independent variable (X) is assumed not random variables. Robustness t test and f test to violations of normality and homogeneity of data is shown by using simulation data generation from R program which is not normal and heterogeneous. Data generation t test and F test is concluded that data remains t distribution and F distribution , as well as significant p-value.The conclusion was that the t test and F test proved the solidity to abnormalities and heterogeneity of data.

How to Cite

Sari A.Q, Sukestiyarno Y.L., & Agoestanto A. (2017). Batasan Prasyarat Uji Normalitas dan Uji Homogenitas pada Model Regresi Linear. *Unnes Journal of Mathematics*, 6(2): 168-177.

✉Alamat korespondensi:
E-mail: atmira.qurnia@yahoo.co.id

p-ISSN 2252-6943
e- ISSN 2460-5859

PENDAHULUAN

Matematika merupakan ilmu yang mempunyai peranan penting untuk menunjang ilmu-ilmu lainnya. Dalam perkembangannya kajian matematika terbagi menjadi dua arah yakni murni dan terapan. Matematika murni yang mengkaji tentang seluk beluk matematika serta memecahkan kasus-kasus dalam matematika. Salah satu cabang dari matematika terapan adalah statistika. Dunia penelitian atau riset dimanapun dilakukan, tidak akan terlepas dari masalah statistika.

Dalam statistik inferensial untuk menguji hipotesis, peneliti banyak menggunakan statistik parametris. Statistik parametris ini lebih banyak digunakan untuk menganalisis data yang berbentuk interval dan rasio dengan dilandasi beberapa persyaratan tertentu misalnya data variabel yang akan dianalisis nilai residual harus berdistribusi normal (Sugiyono, 2004: 8).

Statistika mempunyai peranan penting dalam memecahkan masalah yang terjadi pada bidang ilmu lainnya. Pada statistik parametrik terdapat teknik komputasi untuk pengambilan keputusan yang didasarkan pada model distribusi yang diketahui, sehingga penggunaanya dilandasi oleh berlakunya asumsi bahwa kesesuaian data sampel dengan model distribusi yang bersangkutan (*data-model fit*).

Dalam menggolah data, peneliti akan selalu berkepentingan menentukan hubungan atau pengaruh antara dua atau lebih peubah. Analisis regresi linear dipergunakan untuk menelaah hubungan antara dua variabel atau lebih, terutama untuk menelusuri pola hubungan yang modelnya belum diketahui dengan sempurna, atau untuk mengetahui bagaimana variasi dari beberapa variabel bebas (prediktor X atau independent variable) mempengaruhi variabel terikat (respon Y atau dependent variable) dalam suatu fenomena yang kompleks. Jika $X_1, X_2, \ldots, X_i$ adalah variabel-variabel independen dan Y adalah variabel dependen, maka terdapat hubungan fungsional antara X dan Y, di mana variasi dari X akan diiringi pula oleh variasi dari Y. Secara matematika hubungan di atas dapat dijabarkan sebagai berikut.

$$Y = f(X_1, X_2, \ldots, X_i, e)$$

di mana Y adalah variabel terikat, X adalah variabel bebas dan e adalah variabel residual (*disturbance term*) (Walpole, 1995).

Regresi linear mempunyai persamaan yang disebut sebagai persamaan regresi. Persamaan regresi mengekspresikan hubungan linear antara variabel tergantung/variabel kriteria yang diberi simbol Y dan salah satu atau lebih variabel bebas/prediktor yang diberi simbol X jika hanya ada satu prediktor dan X_1, X_2 sampai dengan X_k, jika terdapat lebih dari satu prediktor (Crammer & Howitt, 2006:139).

Istilah regresi diperkenalkan oleh sir Francis Galton dalam penemuannya yang ditulis dalam artikel yang berjudul *Family Likeness in Stature (Proceedings of Royal Society,* London, vol. 40,1886) (Supranto, 2005: 35). Kekhawatiran bahwa data sampel tidak terdistribusi mengikuti model data populasi yang diasumsikan atau tidak memenuhi kondisi yang disyaratkan bagi penggunaan teknik komputasi tertentu menyebabkan banyak peneliti sosial pemakai statistika melakukan lebih dahulu pengujian asumsi sebelum melakukan uji hipotesis.

Pada hampir semua skripsi S1, thesis S2, dan bahkan disertasi S3 dapat kita temui laporan hasil berbagai uji asumsi yang dilakukan sebelum pengujian hipotesisnya sehingga terdapat kesan kuat sekali bahwa uji asumsi merupakan prasyarat dan bagian yang tak terpisahkan yang mendahului analisis data penelitian. Pada tes parametrik klasik untuk memperoleh hasil yang akurat, asumsi yang mendasari mereka (misalnya, normalitas dan homoskedastisitas) harus dipenuhi. Asumsi ini jarang bertemu ketika menganalisis data real. Penggunaan metode parametrik klasik dengan asumsi melanggar dapat menjadikan hasil yang tidak akurat. Hal itu dapat menyebabkan kesalahan substantif untuk interpretasi data.

Kepanikan terjadi apabila hasil uji asumsi ternyata tidak sesuai dengan harapan. Berbagai reaksi timbul mulai dari reaksi wajar berupa usaha untuk menggunakan alternatif model uji yang lebih cocok dengan data, transformasi data agar sesuai dengan model yang diinginkan, sampai pada usaha-usaha memanipulasi data agar tampak memenuhi asumsi yang diinginkan. Sayangnya seringkali hal itu dilakukan tanpa pemahaman yang cukup mengenai permasalahan yang sedang dihadapi sehingga ada peneliti yang melakukan '*trimming*' atau pemangkasan terhadap subjek yang dianggapnya sebagai '*outliers*' agar datanya terdistribusi mengikuti model regresi linear, dan ada pula praktisi yang mecoba menggunakan model matematis yang terlalu kompleks bagi tujuan penelitiannya sehingga malah menjadikan kesimpulan analisisnya sulit dipahami oleh pembaca awam.

Asumsi bahwa sampel diambil secara random dan bahwa distribusi populasi adalah normal merupakan dua contoh asumsi yang merupakan formalitas dalam analisis. Beberapa orang percaya bahwa semua data yang dikumpulkan dan digunakan untuk analisis harus didistribusikan secara normal. Tapi distribusi normal tidak terjadi sesering orang pikirkan, dan itu bukan tujuan utama. Distribusi normal adalah sarana untuk mencapai tujuan , bukan tujuan itu sendiri. Data terdistribusi secara normal diperlukan untuk menggunakan sejumlah alat statistik, seperti analisis regresi, uji t, uji F atau analisis varians (ANAVA) dan masih banyak lagi. Jika seorang praktisi tidak menggunakan alat khusus seperti itu, maka tidaklah penting apakah data terdistribusi secara normal. Distribusi menjadi masalah hanya ketika praktisi mencapai suatu titik dalam sebuah proyek di mana mereka ingin menggunakan alat statistik yang memerlukan data terdistribusi normal dan mereka tidak memilikinya.

Dari makna kata, asumsi (*assumption*) berarti *a statement accepted true without proof (Encarta 97 Encyclopedia*) atau *something taken for granted (Random House Webster's Unabridged Dictionary)*. Kedua makna kata itu tentu berlaku juga bagi pengertian asumsi statistika. Oleh karena itu dalam inferensi statistika, data yang akan dianalisis dianggap memenuhi asumsi-asumsi yang disyaratkan bagi formula komputasinya. Analisis dapat dilakukan tanpa harus melakukan pemeriksaan terlebih dahulu terhadap terpenuhi tidaknya asumsi yang bersangkutan. Kalaupun ternyata kemudian bahwa data yang digunakan tidak sesuai dengan asumsi-asumsinya, maka kesimpulan hasil analisisnya tidak selalu *invalid*.

Dalam situasi aplikasi, asumsi-asumsi bagi distribusi sampling dibuat sebagai dasar keputusan pemilihan teknik komputasi tertentu guna pengujian suatu hipotesis. Asumsi ini jarang atau bahkan tidak pernah benar-benar diuji terhadap data sampel melainkan langsung dianggap benar (Hays & Winkler, 1971).

Pertanyaan yang mungkin timbul di kalangan pengguna statistika adalah sejauh manakah tuntutan uji asumsi dilakukan sebelum melakukan uji hipotesis. Berdasarkan uraian di atas maka penulis akan membahas mengenai sejauhmana persyaratan uji normalitas dan homogenitas pada model regresi linear harus dipenuhi.

Tujuan utama kegiatan penelitian antara lain ialah menemukan prinsip yang dapat diberlakukan secara umum atau bersifat universal. Secara ideal teoritik seorang peneliti harus meneliti keseluruhan populasi pada data penelitian sehingga generalisasi yang dikemukakan tidak terlalu jauh dengan kenyataan aslinya. Pada kenyataannya meneliti populasi keseluruhan sangat tidak praktis, dapat memakan waktu lama atau bahkan tidak mungkin, sehingga sebelum dilakukan pengukuran data peneliti harus mengubah populasi menjadi populasi yang lebih kecil (sampel) yang diambil secara acak dari populasi tersebut.

Disisi lain alat uji yang dipakai yaitu uji t dan uji F cukup kekar (robust) sehingga anggapan kenormalan dan homogenitas tidak dituntut secara ketat namun cukup agak kasar (Sembiring, 2003: 39). Uji t dan uji F mengasumsikan bahwa nilai residual mengikuti distribusi normal. Jika asumsi ini dilanggar maka uji statistik menjadi tidak valid untuk jumlah sampel kecil (Ghozali, 2011: 160). Berdasarkan uraian latar belakang di atas, maka penulis tertarik untuk menganalisis sejauhmana tuntutan uji normalitas dan uji homogenitas pada model regresi linear, serta menunjukkan uji t dan uji F bersifat robust.

METODE

Pada kegitan ini dilakukan kajian pustaka, yaitu mengkaji permasalahan secara teoritis berdasarkan sumber-sumber pustaka yang ada dengan kerangka teoritis. Kerangka teoritis adalah suatu model yang menerangkan bagaimana hubungan suatu teori dengan faktor-faktor penting yang telah diketahui dalam suatu masalah tertentu. Selain itu dilakukan pula studi pustaka, dalam studi pustaka ini digunakan sumber pustaka yang relevan yang digunakan untuk mengumpulkan informasi yang diperlukan dalam penelitian. Studi pustaka dilakukan dengan mengumpulkan sumber pustaka yang dapat berupa buku, jurnal, makalah dan sebagainya. Setelah sumber pustaka terkumpul dilanjutkan dengan pengkajian dari sumber pustaka tersebut. Pada akhirnya sumber pustaka itu dijadikan landasan untuk menganalisis permasalahan.

Selain teoritis akan dilakukan simulasi pada dua kelompok variabel yang diobservasi. Simulasi adalah tiruan dari proses dunia nyata atau sistem. Simulasi menyangkut pembangkitan proses serta pengamatan dari proses untuk menarik kesimpulan dari sistem yang diwakili (Banks, 1998).

Simulasi dilakukan untuk menunjukkan robust uji t dan uji F yang akan dilakukan pada dua kelompok data dengan berbagai macam kriteria data. Dasar uji t dan uji F mengasumsikan nilai residual berdistribusi normal. Mengecek robust uji t dan uji F terhadap ketidaknormalan data dan heterogenitas varian. Data yang disimulasikan adalah data tidak normal dan tidak homogen hasil bangkitan program R. Data dilakukan uji t dan uji F dengan bantuan program SPSS. Hasil output SPSS dicatat kemudian dilakukan pengambilan kesimpulan dari data yang telah tersedia.

Adapun langkah-langkah yang dilakukan dalam tahap simulasi adalah:

1. Menetukan kondisi yang digunakan untuk menguji robustness dari uji t dan uji F.
2. Menentukan besarnya n sampel yang hendak dilibatkan. Misalnya dipilih n = 30, 50 dst. Untuk melihat dampak dari uji t dan uji F untuk besar sampel yang berbeda.
3. Menentukan besarnya effect size yang hendak dilihat untuk mengamati *power* dari uji t dan uji F. Misalnya *effect size* = 1 standard deviasi.
4. Untuk setiap data yang diperoleh, lakukan uji t dan uji F kemudian catat apakah p<0,05 atau tidak.
5. Nilai p>0,05 menunjukkan data yang diperoleh konsisten dengan hipotesis nol.

Kondisi yang dioperasionalkan dalam simulasi yaitu data yang dibangkitkan dari program R dengan kriteria berikut ini :

1. data kedua kelompok tidak normal dan heterogen varian serta besar n sama untuk kedua kelompok, macam kelompoknya adalah:
   a. y_1: generating data untuk kelompok 1
   b. y_2: generating data untuk kelompok 2 (mean 2 = 2*sd, sd2 = 3 sd1)
2. Data kedua kelompok tidak normal dan heterogen varian serta besar n berbeda searah dengan varian untuk kedua kelompok, macam kelompoknya adalah:
   a. x_1: generating data untuk kelompok 1
   b. x_2: generating data untuk kelompok 2 (mean 2 = 2*sd, sd2 = 3 sd1)
3. data kedua kelompok tidak normal dan heterogen varian serta besar n berbeda searah dengan varian untuk kedua kelompok, macam kelompoknya adalah:
   a. x_1: generating data untuk kelompok 1
   b. x_2: generating data untuk kelompok 2 (mean 2 = 2*sd, sd2 = 3 sd1)

Data yang telah dibangkitkan kemudian dilakukan uji t dan uji F dengan menggunakan bantuan program SPSS. Diperoleh output nilai t, nilai F, dan nilai signifikan dari masing-masing uji tersebut.

HASIL DAN PEMBAHASAN

Normalitas

Regresi linear normal klasik mengasumsi dalam model $Y_i = \beta_0 + \beta_1 x_i + \varepsilon$, $i = 1,2,\dots,n$ bahwa tiap ε_i didistribusikan secara normal dengan rata-rata $E(\varepsilon_i) = 0$ dan varians $E(\varepsilon_i^2) = \sigma^2$ dan $Cov(\varepsilon_i, \varepsilon_j), E(\varepsilon_i, \varepsilon_j) = 0$ untuk $i \neq j$, ε_i dan ε_j tidak berkorelasi, $i \neq j$ dan bebas (*independent*) sehingga $cov(\varepsilon_i, \varepsilon_j) = 0$ jadi $E(Y_i) = \beta_0 + \beta_1 X_i, E(Y_i) = \sigma^2$ dan Y_i dan Y_j, $i \neq j$, tidak berkorelasi .

Asumsi ε_i merupakan suatu peubah acak, dengan nilai tengah nol dan ragam σ^2 secara ringkas bisa dinyatakan sebagai

$$\varepsilon_i \sim N(0, \sigma^2)$$

Dimana ~ berarti "didistribusikan sebagai" dan dimana N berarti "distribusi normal", unsur dalam tanda kurung menyatakan dua parameter distribusi normal, yaitu rata-rata dan varians (Drapper and Smith, 1992: 21).

Agar dapat menilai kebaikan suatu persamaan regresi maka data sampel diharap memenuhi beberapa asumsi, yaitu data sampel berasal dari suatu populasi berdistribusi normal dengan rataan μ_i dan variansi σ^2 (Sembiring, 2003: 37). Hal ini mengartikan bahwa peubah acak Y_i berdistribusi $N(\mu_i, \sigma^2)$ untuk setiap i. Lambang Y_i menyatakan peubah acak sedangkan y_i hanyalah salah satu dari sekian banyak nilainya yang diambil secara acak. Untuk setiap x_i distribusi Y_i normal, tepatnya:

$$p(y_i|x_i) = \frac{1}{\sqrt{2\pi}\sigma} eksp^{-\frac{1}{2}\left(\frac{y_i - \mu_i}{\sigma}\right)^2}, -\infty < y_i < +\infty.$$

Lambang $p(y|x)$ menyatakan fungsi padat y bila diketahui x. Jadi pada suatu x_i akan diambil satu atau lebih sampel dari populasi yang berdistribusi normal $N(\mu_i, \sigma^2)$, sehingga persamaannya menjadi

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i\ , \quad i = 1,2,\dots,n$$

Dalam persamaan di atas β_0 dan β_1 menyatakan parameter populasi yang merupakan tetapan, tetapi tidak diketahui besarannya dan ε_i merupakan galat acak yang berdistribusi $N(0, \sigma^2)$. Setiap suku galat ε_i diasumsikan mempunyai ragam yang sama σ^2 oleh karenanya, respons Y_i mempunyai ragam yang sama pula

$$\sigma^2 \{Y_i\} = \sigma^2$$

karena, berdasarkan sifat variansi, kita memperoleh

$$\sigma^2\{\beta_0 + \beta_1 X_i + \varepsilon_i\} = \sigma^2 \{\varepsilon_i\} = \sigma^2$$

Jadi, model regresi $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i\ ,\ i = 1,2,\dots,n$ mengasumsikan bahwa sebaran peluang bagi Y mempunyai ragam yang sama σ^2, tidak tergantung pada nilai peubah bebas X. Untuk memudahkan permasalahan, seterusnya dianggap X_i tidak mempunyai distribusi. Pada prakteknya nilai X_i sering berada di bawah pengendalian peneliti, artinya nilainya data ditentukan oleh peneliti. Karena itu penamaan peubah bebas untuk X merupakan kesalahan karena X tidaklah berubah secara bebas namun X bebas dari distribusi karena X diasumsikan tetap (*fixed*) artinya X tidak mempunyai disribusi.

Karena ε mempunyai distribusi, sedangkan x tidak, maka Y juga mempunyai distribusi yang sesuai dengan ε yaitu

Y_i berdistribusi $N(\beta_0 + \beta_1 x_i \sigma^2)$.

Pemisalan bahwa $var\ (\varepsilon_i)$, $var(Y_i)$, sama dengan σ^2 untuk setiap i, berarti variansinya tidak berubah pada setiap data sampel. Anggapan ini sangat penting untuk tujuan pengujian hipotesis. Tetapi anggapan $E(\varepsilon_i) = 0$, jadi $E(Y_i) = \mu_i = \beta_0 + \beta_1 x_i$, untuk setiap i, tidaklah begitu penting karena dapat selalu dijadikan seperti itu dengan translasi.

Bila Y berdistribusi normal dengan rataan μ dan simpangan baku σ, maka $aY + b$ juga berdistribusi normal dengan rataan $a\mu + b$ dan simpangan baku $a\sigma$, untuk a dan b tetapan sembarang. Bila Y_1 dan Y_2 dua peubah yang berdistribusi normal, masing – masing $N(\mu_1, \sigma_1^2)$ dan $N(\mu_2, \sigma_2^2)$, dan ρ koefisien korelasi antara Y_1 dan Y_2, maka funfsi pada gabungan Y_1 dan Y_2 adalah

$$f(y_1, y_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} eksp\left[-\frac{1}{2}Q(y_1, y_2)\right]$$

untuk

$$Q(y_1, y_2) = \frac{1}{1-\rho^2}\left[\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\frac{(y_1-\mu_1)(y_2-\mu_2)}{\sigma_1\sigma_2}\right]$$

Bila Y_1 dan Y_2 bebas satu sama lain, maka $\rho = 0$ dan bentuk fungsi padat gabungan di atas menjadi sangat sederhana

$$f(y_1, y_2) = \frac{1}{\sqrt{2\pi}\sigma_1} eksp\left[-\frac{1}{2}\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2\right] \frac{1}{\sqrt{2\pi}\sigma_2} eksp\left[-\frac{1}{2}\left(\frac{y_2-\mu_2}{\sigma_2}\right)^2\right]$$

$$f(y_1, y_2) = f(y_1).f(y_2)$$

Dalam hal normal, pengertian tidak berkorelasi dan bebas adalah identik. Umumnya korelasi yang nol tidak mengakibatkan kebebasan, tetapi dua peubah yang saling bebas mempunyai korelasi nol.

Bila $Y_1, Y_2, \dots, Y_n$ peubah acak yang saling bebas masing – masing berdistribusi normal dengan rataan $\mu_1, \mu_2, \dots, \mu_n$ dan simpangan baku $\sigma_1, \sigma_2, \dots, \sigma_n$ maka peubah acak

$$Y = a_1Y_1 + a_2Y_2 + \cdots + a_nY_n$$

juga berdistribusi normal dengan rataan

$$\mu_y = a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n$$

dan variansi

$$\sigma_y^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2$$

Khususnya bila $a_1 = a_2 = \cdots = a_n = \frac{1}{n}$ dan $Y_1, Y_2, \dots, Y_n$ berdistribusi normal yang sama yaitu $N(\mu, \sigma^2)$, maka

$$\bar{Y} = (Y_1 + Y_2 + \cdots + Y_n)/n$$

Juga berdistribusi normal dengan rataan $\mu_{\bar{y}} = \mu$ dan variansi $\sigma_{\bar{y}}^2 = \sigma^2/n$.

Bila $Y_1, Y_2, \dots, Y_n$ saling bebas dan masing-masing berdistribusi $N(0,1)$, maka $(Y_1^2 + Y_2^2 + \cdots + Y_n^2)$ mempunyai distribusi *khi-kuadrat* dengan derajat kebebasan $v = n$ lambang χ_v^2. Distribusi χ_v^2 ini mempunyai rataan v dan variansi $2v$. Misal untuk $v = 10$ dan $\alpha = 5\%$, maka $P\{\chi_{10}^2 \geq \chi^2(10; \alpha = 0{,}05)\}$ dipenuhi jika $\chi^2(10; \alpha = 0{,}05) = 18{,}31$. Bila $Y_1, Y_2, \dots, Y_n$ saling bebas dan masing – masing $N(\mu - \sigma^2)$ maka

$$\sum(Y_i - \mu)^2/\sigma^2 \text{ berdistribusi } \chi_n^2$$

Bila

$$S^2 = \frac{1}{n-1}\sum(Y_i - \bar{Y})^2$$

maka $(n-1)S^2/\sigma^2$ berdistribusi χ_{n-1}^2.

Begitu pula $(\bar{Y} - \mu)^2/(\sigma^2/n)$ berdistribusi χ_1^2. Suatu teorema yang banyak penggunaannya dalam statistika menyatakan bahwa kedua peubah acak $(n-1)\ S^2/\sigma^2$ dan $(\bar{Y} - \mu)^2/(\sigma^2/n)$ saling bebas.

Bila Z suatu peubah acak $N(0,1)$ dan U berdistribusi χ_v^2 dan keduanya saling bebas, maka

$$T = \frac{Z}{\sqrt{U/v}}$$

Dikatakan mempunyai *distribusi-t* dengan derajat kebebasan v lambang t_v. Distribusi-t sangat mirip dengan distribusi normal, setangkup terhadap rataannya. Rataannya selalu nol dan variansinya $v/(v-2), v > 2$, variansinya tergantung pada derajat kebebasan. Tabel t diberikan di Lampiran 2 Bila $Y_1, Y_2, \dots, Y_n$ saling bebas dan masing – masing $N(\mu - \sigma^2)$ maka

$$T = \frac{\bar{Y}-\mu}{S/\sqrt{n}}$$

berdistribusi-t dengan derajat kebebasan $n - 1$.

Bila $n \to \infty$ maka $t_n \to N(0,1)$. Misalkan $Y_{11}, Y_{12}, \dots, Y_{1n_1}$ bebas satu sama lain dan berdistribusi $N(\mu_1, \sigma^2)$. Juga misalkan $Y_{21}, Y_{22}, \dots, Y_{2n_2}$ saling bebas dan saling bebas pula dari peubahnya, masing-masing berdistribusi $N(\mu_2, \sigma^2)$.

Misalkan selanjutnya

$$S^2 = [(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2]/(n_1 + n_2 - 2)$$

dengan

$$S_1^2 = \sum_i (Y_{1i} - \bar{Y}_1)^2/(n_1 - 1)$$

dan

$$S_2^2 = \sum_i (Y_{2i} - \bar{Y}_2)^2/(n_2 - 1)$$

maka

$$(n_1 + n_2 - 2)S^2/\sigma\text{^}2$$

berdistribusi $\chi_{n_1+n_2-2}^2$ dan bebas dari $\bar{Y}_1$ dan $\bar{Y}_2$, sehingga

$$T = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

berdistribusi-t dengan derajat kebebasan yaitu $n_1 + n_2 - 2$.

Bila U_1 dan U_2 peubah yang saling bebas dan berdistribusi khi-kuadrat dengan derajat kebebasan masing-masing v_1 dan v_2 maka

$$F = \frac{U_1/v_1}{U_2/v_2}$$

Dikatakan mempunyai *distribusi-F* dengan derajat kebebasan v_1 dan v_2, lambang F_{v_1, v_2}.

Peubah acak ini mempunyai rataan $\frac{v_2}{v_2/2}, v_2 > 2$ dan variansi $\frac{2v_2^2(v_1+v_2-2)}{v_1(v_2-2)^2(v_2-4)}, v_2 > 4$.

$(U_1/v_1)/(U_2/v_2)$ juga mempunyai distribusi F, tapi derajat kebebasannya v_1 dan v_2, lambang F_{v_2,V_1}. Jadi $F_{v_1,V_2} = 1/F_{v_2,V_1}$. Bila $v_1 = 1$ maka $F_{1,V_2} = t_{v_2}^2$.

Jadi distribusi-t adalah hal khusus dari distribusi-F, yaitu bila derajat kebebasan F pada pembilang 1.

Misalkan $Y_{11}, Y_{12}, \dots, Y_{1n_1}$ dan $Y_{21}, Y_{22}, \dots, Y_{2n_2}$ dua sampel acak berukuran n_1 dan n_2 yang saling bebas, masing-masing berasal dari populasi $N(\mu_1, {\sigma_1}^2)$ dan $N(\mu_2, {\sigma_2}^2)$. Maka peubah acak

$$F = \frac{S_1^2/{\sigma_1}^2}{S_2^2/{\sigma_2}^2}$$

mempunyai distribusi F dengan derajat kebebasan $n_1 - 1$ dan $n_2 - 1$ bila

$$S_1 = \frac{1}{n_1-1}\sum(Y_{1i} - \bar{Y}_1)^2,\ \bar{Y}_1 = \sum Y_{1i}/n_1$$

dan

$$S_2 = \frac{1}{n_2-1}\sum(Y_{2i} - \bar{Y}_2)^2,\ \bar{Y}_2 = \sum Y_{2i}/n_2.$$

Setiap suku galat ε_i diasumsikan mempunyai ragam yang sama σ^2, oleh karenanya respons Y_i mempunyai ragam yang sama pula. Karena ε mempunyai distribusi, sedangkan x tidak, maka Y juga mempunyai distribusi yang sesuai dengan ε yaitu Y_i berdistribusi $N(\beta_0 + \beta_1\, x_i\ \sigma^2)$. Pemisalan bahwa $var\ (\varepsilon_i)$, $var(Y_i)$, sama dengan σ^2 untuk setiap i, berarti variansinya tidak berubah pada setiap data sampel. Karena asumsi galat berdistribusi normal berdampak pada variabel dependen (Y) maka yang diuji normalitas adalah variabel Y dan variabel independen (X) diasumsikan bukan variabel acak (Sukestiyarno, 2013: 69).

Model regresi $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ mengasumsikan bahwa sebaran peluang bagi Y mempunyai ragam yang sama σ^2, tidak tergantung pada nilai peubah bebas X. Karena ε mempunyai distribusi, sedangkan x tidak, maka Y juga mempunyai distribusi yang sesuai dengan ε yaitu Y_i. Karena asumsi galat berdistribusi normal dan homogen berdampak pada variabel dependen (Y) maka yang diuji normalitas dan homogenitas adalah variabel Y dan variabel independen (X) diasumsikan bukan variabel acak.

Ketika asumsi normalitas tidak dipenuhi oleh data, maka analisis regresi akan memberikan hasil yang meleset dari yang sebenarnya. Hal itu bisa saja terjadi karena banyak kejadian yang diluar kebiasaan, misalnya data ekstrim, data yang diurutkan, data mengikuti distribusi selain distribusi normal dan masih banyak penyebab-penyebab lainnya. Jika analisis tetap dilanjutkan, ada resiko hasil penelitian yang diperoleh tidak menggambarkan keadaan yang sebenarnya.

Analisis regresi perlu dilakukan uji normalitas karena persyaratan awal agar dapat menilai kebaikan suatu persamaan regresi adalah normalitas error. Uji normalitas diperlukan untuk menjawab pertanyaan apakah syarat sampel yang representatif terpenuhi atau tidak, sehingga hasil penelitian dapat digeneralisasi pada populasi atau dapat mewakili populasi (Hadi, 2001).

Jika sebuah data mempunyai sebaran yang tidak normal, perlakuan yang mungkin dilakukan agar data menjadi normal dapat dilakukan dengan berbagai macam cara. Beberapa cara tersebut diantaranya adalah menambahkan jumlah data pada variabel dependen (Y), menghilangkan data yang dianggap menjadi penyebab ketidaknormalan data (data *outlier*), dilakukan transformasi data.

Homogenitas

Salah satu asumsi penting dari model regresi linear klasik adalah varian error ε_i pada setiap nilai – nilai variabel bebas adalah sama (konstan), asumsi ini disebut juga sebagai asumsi homoskedastisitas atau homogenitas varian yang disimbolkan dengan: $(\varepsilon_i) = \sigma^2$, $i = 1,2, \dots, n$. Karena salah satu asumsi dasar regresi linear adalah homogenitas error maka pengujian homogenitas data sangat penting dilakukan pada setiap pengolahan data. Jika pada normalitas yang dilakukan uji adalah gabungan data antara variabel Y dan variabel X maka begitu pula dengan uji homogenitas yang diuji adalah gabungan data variabel Y dan variabel X.

Adanya heterokedastisitas sama sekali bukan berarti suatu model regresi adalah lemah. Pengira kuadrat terkecil tetap tak bias dan konsisten, tetapi tidak efisien (variansi membesar). Dampak dari membesarnya variansi adalah (Setiawan dan Kusrini, 2010)

1. Pengujian parameter regresi dengan statistik uji t menjadi tidak valid.

   $H_0: \beta_j = 0$

   $H_1: \beta_j \neq 0$

   $t_{hitung} = \frac{\hat{\beta}_j}{s(\hat{\beta}_j)}$ akan mengecil jika $s(\hat{\beta}_j)$ besar, sehingga cenderung untuk tidak menolak H_0.
2. Selang kepercayaan (perkiraan selang) untuk parameter regresi cenderung melebar. $P\left[\hat{\beta}_j - t_{\alpha/2}.s(\hat{\beta}_j) \leq \hat{\beta}_j \leq \hat{\beta}_j + t_{\alpha/2}.s(\hat{\beta}_j)\right] = 1 - \alpha$ akan melebar jika $s(\hat{\beta}_j)$ besar. Dengan melebarnya selang kepercayaan, hasil perkiraan yang diperoleh menjadi tidak dapat dipercaya.

Apabila asumsi homogenitas tidak dipenuhi dalam analisis regresi linear, maka didapatkan keadaan bahwa varian tidak bersifat konstan disebut heteroskedastisitas atau disimbolkan

$Var(\varepsilon_i) = \sigma_i^2, i = 1,2, \ldots, n.$

Dalam kenyataannya, asumsi homoskedastisitas dari kesalahan pengganggu ε mungkin tidak bisa dipenuhi, dengan kata lain varian dari kesalahan pengganggu bersifat heteroskedastisitas, yaitu $E(\varepsilon_i^2) = \sigma_i^2$. Hal ini dapat dipahami jika diperhitungkan faktor–faktor yang menjadi penyebab adanya kesalahan pengganggu ε dalam model regresi. Faktor kesalahan pengganggu ε dimasukkan ke dalam model untuk dapat memperhitungkan kesalahan–kesalahan yang mungkin terjadi dalam pengukuran dan kesalahan karena mengabaikan variabel–variabel tertentu. Dengan memperhatikan kedua perhitungan itu, maka terdapat alasan untuk memperkirakan bahwa varian ε bervariasi secara sistematis dengan variabel bebas X.

Konsekuensi dari pelanggaran asumsi homoskedastisitas adalah sebagai berikut:

1. Penduga (*estimator*) yang diperoleh tetap memenuhi persyaratan tidak bias.

Diberikan estimator $b_1 = \left[\frac{\sum_{i=1}^n (X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^n (X_i-\bar{X})^2}\right]$

Anggaplah bahwa : $k_i = \frac{(X_i-\bar{X})}{\sum_{i=1}^n (X_i-\bar{X})^2}$

Sehingga :

$b_1 = \sum_{i=1}^n k_i y_i$

$b_1 = \sum_{i=1}^n k_i(\beta_0 + \beta_1 X_i + \varepsilon_i)$

$b_1 = \beta_0 \sum_{i=1}^n k_i + \beta_1 \sum_{i=1}^n k_i X_i + \sum_{i=1}^n k_i \varepsilon_i$

Dengan diketahui bahwa :

$\sum_{i=1}^n k_i = 0$ dan

$$\sum_{i=1}^n k_i X_i = \sum_{i=1}^n k_i(x_i + \bar{X}) = \frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n x_i^2} = 1$$

Dengan demikian

$b_1 = \beta_1 + \sum_{i=1}^n k_i \varepsilon_i$

Sehingga diperoleh :

$E[b_1] = \beta_1 + \sum_{i=1}^n k_i E[\varepsilon_i]$

$E[b_1] = \beta_1$

Demikian juga untuk estimator parameter β_0 yaitu b_0.

Diberikan estimator $b_0 = \bar{Y} - b_1\bar{X}$

$b_0 = \frac{1}{n}(\sum_{i=1}^n Y_i - b_1 \sum_{i=1}^n X_i)$

substitusikan b_1 ke dalam persamaan di atas, maka :

$b_0 = \frac{1}{n}\left(\sum_{i=1}^n Y_i - \frac{\sum_{i=1}^n Y_i(X_i-\bar{X})}{\sum_{i=1}^n (X_i-\bar{X})^2}\sum_{i=1}^n X_i\right)$

$b_0 = \sum_{i=1}^n \left(\frac{1}{n} - \sum_{i=1}^n X_i k_i\right) Y_i$

$b_0 = \sum_{i=1}^n \left(\frac{1}{n} - \sum_{i=1}^n X_i k_i\right)(\beta_0 + \beta_1 X_i + \varepsilon_i)$

$b_0 = \beta_0 + \frac{1}{n}\sum_{i=1}^n \varepsilon_i - \bar{X}\sum_{i=1}^n k_i \varepsilon_i$

Sehingga akan diperoleh :

$E[b_0] = \beta_0 + \frac{1}{n}\sum_{i=1}^n E[\varepsilon_i] - \bar{X}\sum_{i=1}^n k_i E[\varepsilon_i]$

$E[b_0] = \beta_0$

Dapat disimpulkan bahwa sifat ketidakbiasan tidak tergantung pada varian galat. Jika dalam model regresi ada heteroskedastisitas, maka kita tetap memperoleh nilai parameter yang tidak bias karena sebagai penduga tidak bias tidak memerlukan asumsi bahwa varian galat harus konstan.

2. Varian penduga yang diperoleh akan menjadi tidak efisien, artinya penduga tersebut tidak memiliki varian terkecil diantara penduga – penduga tidak bias lainnya. Dengan asumsi adanya homoskedastisitas, maka :

$Var[b_1] = E[(b_1 - \beta_1)^2]$

$Var[b_1] = E[(\beta_1 + \sum_{I=1}^N k_i\varepsilon_i - \beta_1)^2]$

$Var[b_1] = E[(\sum_{I=1}^N k_i\varepsilon_i)^2]$

$Var[b_1] = E[k_1^2\varepsilon_1^2 + k_2^2\varepsilon_2^2 + \cdots + k_n^2\varepsilon_n^2 + 2k_1k_2\varepsilon_1\varepsilon_2 + \cdots + 2k_{n-1}k_n\varepsilon_{n-1}\varepsilon_n]$

$Var[b_1] = E[k_1^2\varepsilon_1^2 + k_2^2\varepsilon_2^2 + \cdots + k_n^2\varepsilon_n^2] + E[2k_1k_2\varepsilon_1\varepsilon_2 + \cdots + 2k_{n-1}k_n\varepsilon_{n-1}\varepsilon_n]$

$Var[b_1] = E[\sum_{i=1}^n k_i^2\varepsilon_i^2] + 2E[\sum k_ik_j\varepsilon_i\varepsilon_j]$

$Var[b_1] = k_i^2 E[\varepsilon_i^2] + 2\sum k_ik_jE(\varepsilon_i\varepsilon_j)$

Karena $E(\varepsilon_i\varepsilon_j) = 0$, dan $E[\varepsilon_i^2] = \sigma^2$

Sehingga diperoleh :

$var[b_1] = \sigma^2 \sum_{i=1}^n k_i^2$ (*)

Apabila dengan adanya asumsi heteroskedastisitas maka :

$var[b_1] = \sigma^2 \sum_{i=1}^n k_i^2$

$var[b_1] = \frac{\sigma^2}{\sum_{i=1}^n (X_i-\bar{X})^2}$ (**)

Walaupun β_1 dikatakan adalah unbiased, tetapi tidak efisien karena varian – variannya lebih besar daripada yang diperlukan. Untuk melihat penggunaan persamaan (*) dan (**), diuraikan sebagai berikut:

Misalnya, dinyatakan bahwa varian dengan asumsi heteroskedastisitas proporsional terhadap k_i dan σ^2 maka faktor proporsionalitasnya dinyatakan dengan persamaan: $\sigma_i^2 = k_i\sigma^2$

Dengan mensubstitusikan nilai σ_i^2 ke dalam persamaan (**), diperoleh:

$var[b_1] = k_i\sigma^2 \sum_{i=1}^n k_i^2$

$var[b_1] = k_i\sigma^2 \sum_{i=1}^n \left(\frac{(X_i-\bar{X})}{\sum_{i=1}^n (X_i-\bar{X})^2}\right)^2$

$var[b_1] = \frac{\sigma^2 \sum_{i=1}^n k_i(X_i-\bar{X})^2}{\sum_{i=1}^n (X_i-\bar{X})^2 \sum_{i=1}^n (X_i-\bar{X})^2}$

$var[b_1] = \left(\frac{\sigma^2}{\sum_{i=1}^n (X_i-\bar{X})^2}\right)\left(\frac{\sum_{i=1}^n k_i(X_i-\bar{X})^2}{\sum_{i=1}^n (X_i-\bar{X})^2}\right)$

Sehingga diperoleh :

$var[b_1]$

$= (var[b_1]\ dengan\ asumsi\ homoskedastisitas)$

$\left(\frac{\sum_{i=1}^n k_i(X_i-\bar{X})^2}{\sum_{i=1}^n (X_i-\bar{X})^2}\right)$

Dapat dikatakan bahwa, jika $(X_i - \bar{X})^2$ dan k_i berkorelasi positif atau mempunyai hubungan variabel yang positif dan jika komponen yang kedua dari persamaan diatas lebih besar daripada satu atau dapat dituliskan:

$$\left(\frac{\sum_{i=1}^{n} k_i (X_i - \bar{X})^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2}\right) > 1$$

Maka Var(b_1) dengan asumsi heteroskedastisitas akan lebih besar daripada $Var(b_1)$ dengan asumsi homoskedastisitas. Sebagai akibatnya, standar error dari b_1 terlalu rendah (underestimated). Sebagaimana diketahui bahwa standart error ini memiliki peran dalam pembentukan nilai t hitung yang berkaitan akan menjadi terlalu tinggi (*overestimated*) yang mungkin selanjutnya menghasilkan kesimpulan bahwa dalam kasus spesifik b_1 adalah kelihatannya signifikan, walaupun sebenarnya tidak signifikan. Oleh karena itu jika asumsi homoskedastisitas tidak dipenuhi maka hasil uji t dan uji F tidak menentu (Setiawan dan Kusrini, 2010: 107).

Selain uji signifikan tidak dapat diterapkan, batas-batas kepercayaan juga tidak dapat diterapkan. Artinya jika varian penaksir model tidak memenuhi asumsi homoskedastisitas, maka inferensi dan prediksi mengenai koefisien-koefisien sampel akan keliru.

Dalam analisis model regresi linear apabila semua asumsi model regresi linear klasik terpenuhi kecuali asumsi homoskedastisitas yang berarti adanya heteroskedastisitas, maka estimator dari paramater yang diperoleh masih tetap tak bias dan konsisten tetapi estimatornya tidak efisien, baik untuk sampel kecil maupun sampel besar (Ghozali, 2011: 160).

Berdasarkan hasil penelitian dan pengkajian beberapa literatur diperoleh hasil bahwa uji normalitas dan uji homogenitas sangat penting dilakukan dalam setiap proses penelitian. Hal itu disebabkan karena regresi linear normal klasik mengasumsi dalam model $Y_i = \beta_0 + \beta_1 x_i + \varepsilon, i = 1,2,\dots,n$ bahwa tiap ε_i didistribusikan secara normal dan varians $E(\varepsilon_i^2) = \sigma^2$.

Uji normalitas dan uji homogenitas menjadi sangat penting dipenuhi karena pada asumsi awal suatu persamaan regresi linear dikatakan baik jika error/galat regresi berdistribusi normal dan homogen. Setiap suku galat ε_i diasumsikan mempunyai ragam yang sama σ^2 oleh karenanya, respons Y_i mempunyai ragam yang sama pula. Model regresi $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ mengasumsikan bahwa sebaran peluang bagi Y mempunyai ragam yang sama σ^2, tidak tergantung pada nilai peubah bebas X. Karena ε mempunyai distribusi, sedangkan x tidak, maka Y juga mempunyai distribusi yang sesuai dengan ε yaitu Y_i. Karena asumsi galat berdistribusi normal berdampak pada variabel dependen (Y) maka yang diuji normalitas dan homogenitas adalah variabel Y dan variabel independen (X) diasumsikan bukan variabel acak.

Simulasi uji t dan uji F

Menunjukkan robust uji t dan uji F salah satunya dengan melakukan simulasi yang akan dilakukan pada dua kelompok data dengan berbagai macam kriteria data. Dasar bahwa uji t dan uji F mengasumsikan nilai residual berdistribusi normal. Mengecek robust uji t dan uji F terhadap ketidaknormalan data dan heterogenitas varian, kondisi yang dioperasionalkan adalah data yang dibangkitkan dari program R dengan kriteria tidak normal dan tidak homogen. Data dilakukan uji t dan uji F dengan bantuan program SPSS.

Kriteria yang akan digunakan untuk menentukan robust uji t dan uji F adalah nilai p. Nilai p menggambarkan taraf kesalahan yang tepat, paling kecil dimana hipotesis nol ditolak. Hipotesis nol yang sedang diuji kebenarannya menyatakan bahwa sebuah koefisien regresi adalah sama dengan nol. Nilai p mengukur probabilitas *Tipe Kesalahan I*, probabilitas kesalahan menolak hipotesis nol karena kenyataan hipotesis itu benar.

Berikut adalah tabel output hasil simulasi yang dilakukan pada data bangkitan program R dan dilakukan pada program SPSS.

Tabel 1. Nilai signifikan dari Uji t dan Uji F

No	Data	n	Uji F		Uji t	
			nilai F	sig. F	nilai t	sig. t
1	a_1a_2	30	0,252	0,620	0,312	0,757
2	b_1b_2	30	1,018	0,322	- 0,225	0,824
3	c_1c_2	30	1,285	0,267	- 0,252	0,803
4	x_1x_2	50	0,153	0,698	- 0,798	0,429
5	y_1y_2	50	1,796	0,187	0,296	0,768
6	z_1z_2	50	0,174	0,678	- 0,221	0,826

Hasil di atas merupakan perhitungan uji t dan uji F dua kelompok data dengan beberapa kriteria data. Kriteria yang digunakan untuk menilai robust uji t dan uji F adalah jika hasil nilai p lebih besar dari 0,05 maka uji t dan uji F dalam kondisi ini akan cenderung terlalu sering menghasilkan nilai yang signifikan dari yang seharusnya.

Hasil yang diperoleh tentu saja tidak bisa diharapkan exact, misalnya harus 0.05. Jika sedikit lebih besar atau sedikit lebih kecil, maka penyimpangan tersebut dapat diabaikan. Misalnya 0.0502 maka ini tidak perlu dianggap menyimpang dari 0.05. Dari data di atas banyak yang menghasilkan nilai p lebih dari 0,05. Nilai-p yang besar menunjukkan bahwa data yang diperoleh konsisten dengan hipotesis nol.

Untuk menunjukkan kekekaran uji t dan uji f terhadap pelanggaran normalitas dan homogenitas data dengan menggunakan simulasi data. Data yang digunakan adalah bangkitan dengan bantuan program R. Data yang dibangkitkan adalah dua kelompok data dengan kriteria kedua kelompok tidak normal dan heterogen varian. Tahap simulasinya yaitu data bangkitan dilakukan uji t dan uji F diperoleh nilai t, nilai F serta diperoleh nilai p dari masing-masing uji.

Berdasarkan nilai t, uji hipotesis yang dilakukan pada uji t adalah :

H_0 : data berdistribusi t

H_1 : data tidak berdistribusi t

Terima H_0 jika t hitung < t tabel, sebaliknya tolak H_0 jika t hitung > t tabel. Nilai t tabel untuk jumlah data 30 dan 50 adalah 2,048 dan 2,011. Dari output nilai t untuk data $a_1\ a_2$ nilai t 0,312 < 2,048, $b_1 b_2$ nilai t -0,225 < 2,048, $c_1 c_2$ nilai t -0,252 < 2,048, nilai t -0,798 < 2,011, $y_1 y_2$ nilai t 0,296 < 2,011, $z_1 z_2$ nilai t -0,221 < 2,011. Diperoleh simpulan nilai t hitung < t tabel maka terima H_0, yang artinya data berdistribusi t.

Berdasarkan nilai F, uji hipotesis yang dilakukan pada uji F adalah :

H_0 : data berdistribusi F

H_1 : data tidak berdistribusi F

Terima H_0 jika F hitung < F tabel, sebaliknya tolak H_0 jika F hitung > F tabel. Nilai F tabel untuk jumlah data 30 dan 50 adalah 4,195 dan 4,042. Dari output nilai F untuk data $a_1\ a_2$ nilai F 0,252 < 4,195, $b_1 b_2$ nilai F 1,018 < 4,195, $c_1 c_2$ nilai F 1,285 < 4,195, $x_1\ x_2$ nilai F 0,153 < 4,042, $y_1 y_2$ nilai F 1,796 < 4,042, $z_1 z_2$ nilai F 0,174 < 4,042. Diperoleh simpulan nilai F hitung < F tabel maka terima H_0, yang artinya data berdistribusi F.

Terlihat pada output SPSS signifikan yang di dapat adalah nilai p dari semua data simulasi tersebut memperoleh nilai p>0,05 menunjukkan data yang diperoleh konsisten dengan hipotesis nol. Dari tabel output simulasi diperoleh nilai t dan nilai F signifikan, sehingga data yang tidak normal dan tidak homogen setelah dilakukan uji t dan uji F tetap berditribusi t dan berdistribusi F. Dari pembahasan di atas, diperoleh kesimpulan bahwa uji t dan uji F terbukti kekar terhadap ketidaknormalan dan homogenitas data.

SIMPULAN DAN SARAN

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, maka dapat diambil kesimpulan setiap penelitian yang menggunakan regresi linear perlu diuji asumsi normalitas dan homoskedastisitas. Normalitas dan homogenitas diasumsikan pada error data. Jika semua asumsi klasik terpenuhi kecuali normalitas maka data model regresi dapat dikatakan tidak valid dengan jumlah sampel yang ada, namun penduga yang diperoleh masih bisa terdistribusi secara normal karena pelanggaran asumsi tersebut tidaklah terlalu serius akibatnya dalam analisis regresi. Jika semua asumsi klasik terpenuhi kecuali homogenitas data maka penduga yang diperoleh tetap tak bias dan konsisten, tetapi tidak efisien yang artinya penduga tersebut tidak memiliki varian terkecil diantara penduga – penduga tidak bias lainnya baik untuk sampel kecil maupun sampel besar. Pada dasarnya asumsi normalitas dan homogenitas dilakukan pada error data yang merupakan dasar ukuran kebaikan model linear maka uji normalitas dan uji homogenitas data sangat penting dilakukan pada setiap penelitian.

Dalam pengujian parameter regresi terdapat dua pengujian yang dilakukan untuk mengetahui signifikansi dari variabel bebas yaitu pengujian secara individu atau uji t dan pengujian secara serentak atau uji f. Pengujian secara individu atau uji t dan pengujian secara serentak atau uji f telah terbukti robust dengan menunjukkan simulasi uji t dan uji F yang dilakukan terhadap data bangkitan melalui program R. Telah dibuktikan dalam simulasi yakni data yang tidak normal dan tidak homogen setelah dilakukan uji t dan uji F tetap berditribusi t dan berdistribusi F serta nilai p uji t dan uji F dalam simulasi yang dilakukan ke dalam beberapa tipe data diperoleh nilai p lebih dari 0,05. Nilai-p yang tinggi (>0,05) menandakan bahwa sebuah koefisien tidak berbeda secara signifikan dari nol atau data yang diperoleh konsisten dengan hipotesis nol. Maka diperoleh kesimpulan bahwa uji t dan uji F terbukti kekar terhadap ketidaknormalan dan homogenitas data.

Berdasarkan hasil penelitian maka saran yang dapat disampaikan adalah pada pengkajian pada asumsi model regresi linear klasik (*Classical Linear Regression Model*) lain seperti autokorelasi dan multikolinearitas dengan menggunakan metode yang lain, dari pembahasan dan simulasi uji t menggunakan kriteria nilai-p yang telah dijabarkan diperoleh simpulan bahwa uji t terbukti robust dengan melakukan simulasi data bangkitan menggunakan program R, oleh sebab itu ada baiknya dilakukan penelitian menggunakan kriteria lain selain kriteria niali-p.

Dalam penelitian ini telah dilakukan penelitian robust uji t dengan simulasi data menggunakan program R, maka perlu dilakukan penelitian lebih lanjut terhadap uji f apakah terbukti robust atau tidak dengan program lainnya.

DAFTAR PUSTAKA

Banks, J. 1998. *Hand Book of Simulation: Aplication, Methodology, Advances, Aplications and Practices*. John and Willey Sons.

Cramer, D. & Howitt, D. 2006. *The Sage Dictionary of Statistics*. London: Sage Publication.

Drapper. N.R. & Smith. 1992. *Analisis Regresi Terapan*. Edisi Kedua. Jakarta: PT Gramedia Pustaka.

Ghozali, I. 2011. *Aplikasi Analisis Multivariate dengan program IBM SPSS 19.* Edisi kelima. Semarang: Universitas Diponeoro.

Hadi, S. 2001. *Statistik*. Cetakan ke-5. Yogyakarta: Andi Yogyakarta.

Hays. W.L. & Winkler, R.L. 1971. *Statistics-Probability, Inference, and Decision*. New York: Holt, Rinehart and Winston Inc.

Setiawan & Kusrini, D.E. 2010. *Ekonometrika*. Yogyakarta: ANDI.

Sembiring, R.K. 2003. *Analisis Regresi*. Edisi Kedua. Bandung: ITB Bandung.

Sugiyono. 2004. *Metode Penelitian Bisnis*. Alfabeta. Bandung.

Sukestiyarno. 2013. *Olah Data Penelitian Berbantuan SPSS*. Cetakan Keempat. Semarang. Universitas Negeri Semarang.

Supranto, J. 2005. *Ekonometri (1st ed)*. Bogor: Ghalia Indonesia.

Walpole, R.E. & Myers, R.H. 1995. *Ilmu peluang dan statistika untuk insinyur dan ilmuwan*. Edisi ke-4. ITB. Bandung.