LAPORAN PROJEK ANALISIS REGRESI DALAM MEMPREDIKSI HARGA SEWA RUMAH

Ditujukan Guna Memenuhi Tugas Mata Kuliah Analisis Regresi
Dosen Pengampu : Dr. Yusep Suparman, S.Si., M.Sc.

Disusun Oleh:

1. Tansya Putri Rizkya Zakaria (140610230021)
2. Naira Aqila (140610230022)
3. Cica Nuraeni (140610230029)

PROGRAM STUDI S-1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PADJADJARAN
JATINANGOR

2024

KATA PENGANTAR

Segala puji dan syukur bagi Allah SWT yang telah memberikan rahmat-Nya sehingga penulis dapat menyelesaikan tugas ini tepat pada waktunya. Laporan yang berjudul "LAPORAN PROJEK ANALISIS REGRESI DALAM MEMPREDIKSI HARGA SEWA RUMAH" ini dibuat untuk memenuhi tugas mata kuliah Analisis Regresi.

Penulis berharap adanya laporan ini dapat menambah wawasan bagi pembaca mengenai model prediksi terbaik perkiraan harga rumah. Walaupun pada saat menyelesaikan laporan ini penulis banyak mendapatkan hambatan dan kesulitan, akan tetapi berkat bantuan dari berbagai pihak akhirnya laporan ini dapat diselesaikan dengan baik.

Tidak lepas atas bimbingan dan dukungan dari pihak-pihak terkait dalam penulisan dan penyusunan laporan ini sehingga laporan ini dapat terealisasikan. Hanya ucapan terima kasih secara mendalam yang dapat kami berikan tanpa terkecuali.

Penulis menyadari bahwa penyusunan laporan ini masih jauh dari sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran membangun dari pembaca untuk perbaikan di masa mendatang. Penulis berharap laporan ini dapat bermanfaat bagi penulis khususnya dan bagi pembaca umumnya.

Jatinangor, 9 Juni 2024

Penulis

DAFTAR ISI

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam ekonomi modern, harga sewa rumah adalah salah satu komponen utama biaya hidup yang signifikan bagi banyak individu dan keluarga. Dengan meningkatnya urbanisasi dan permintaan akan perumahan, memahami faktor-faktor yang mempengaruhi harga sewa rumah menjadi penting untuk kebijakan perumahan dan keputusan investasi. Analisis regresi adalah salah satu metode statistik yang dapat digunakan untuk memprediksi harga sewa berdasarkan berbagai variabel yang mempengaruhi, seperti karakteristik rumah, lingkungan, dan lainnya.

Projek ini menggunakan data dari Indonesia Family Life Survey (IFLS) yang terbaru. Data ini mencakup berbagai karakteristik rumah, baik kuantitatif maupun kualitatif, serta atribut lingkungan rumah. Dengan menganalisis data ini menggunakan model regresi, diharapkan dapat memberikan prediksi yang akurat tentang harga sewa rumah di wilayah urban.

1.2 Rumusan Masalah

1. Bagaimana karakteristik kuantitatif dan kualitatif rumah mempengaruhi harga sewa rumah di Indonesia?
2. Model regresi apa yang paling efektif dalam memprediksi harga sewa rumah?

1.3 Tujuan Penelitian

Tujuan penulisan laporan ini adalah untuk menganalisis penerapan analisis regresi linier berganda dalam menentukan model prediksi terbaik untuk harga sewa rumah. Proses ini akan menggunakan metode seleksi variabel backward dan memeriksa pemenuhan asumsi klasik regresi. Laporan ini disusun guna memenuhi tugas mata kuliah Analisis Regresi.

BAB II

TINJAUAN PUSTAKA

2.1 Analisis Regresi

Analisis regresi merupakan metode statistika yang banyak digunakan dalam penelitian. Istilah regresi pertama kali diperkenalkan oleh Sir Francis Galton pada tahun 1986. Secara umum, analisis regresi adalah kajian terhadap hubungan satu variabel yang disebut sebagai variabel yang diterangkan dengan satu atau dua variabel yang menerangkan. Variabel yang diterangkan selanjutnya disebut sebagai variabel respon, sedangkan variabel yang menerangkan biasa disebut variabel bebas (Gujarati, 2003).

Model Regresi Linear Sederhana yaitu:

$$y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$$

Regresi Linear Berganda adalah model regresi linear dengan melibatkan lebih dari satu variabel bebas atau predictor. Dalam bahasa inggris, istilah ini disebut dengan multiple linear regression. Menurut Sugiyono (2012: 275), analisis regresi berganda digunakan oleh peneliti, bila peneliti bermaksud meramalkan keadaan (naik turunnya) variabel dependen (kriterium), bila dua atau lebih variabel independen sebagai faktor prediktor dimanipulasi (naik turunnya). Dengan demikian, Regresi Linier Berganda dinyatakan dalam persamaan matematika sebagai berikut :

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \varepsilon_i$$

Keterangan:

- $\beta 0, \beta 1, \beta 2, \beta k$ adalah estimasi parameter yang diperoleh dari model regresi
- $X1, X2, Xki$ adalah nilai-nilai variabel independen
- y_i adalah prediksi nilai variabel dependen berdasarkan nilai-nilai variabel independen

2.2 Dummy Variabel

Dummy variabel adalah variabel yang digunakan dalam analisis regresi untuk mewakili kategori dari variabel non-numerik dengan nilai biner 0 atau 1. Dalam konteks penelitian ini, dummy variabel dapat digunakan untuk mengkodekan informasi yang bersifat kategorikal. Penggunaan dummy variabel membantu dalam menangkap efek dari variabel-variabel tersebut dalam model regresi. (Asteriou, 2015)

2.3 Uji Asumsi Klasik

Uji asumsi klasik adalah analisis yang dilakukan untuk menilai apakah di dalam sebuah model regresi linear Ordinary Least Square (OLS) terdapat masalah-masalah asumsi klasik.

Asumsi klasik adalah syarat-syarat yang harus dipenuhi pada model regresi linear OLS agar model tersebut menjadi valid sebagai alat penduga. Jenis- Jenis uji asumsi klasik :

A. Outlier

Menurut Ghozali (2016: 41), *outlier* adalah kasus ketika terdapat data yang memiliki karakteristik unik yang sangat berbeda jauh dari observasi-observasi lainya dan muncul dalam bentuk nilai ekstrim. Uji *outlier* dapat dilakukan dengan mengkonversi data penelitian ke dalam *studentized residual*. Terdapat empat faktor yang dapat menyebabkan munculnya data *outlier*:

- Kesalahan ketika meng-entri data.
- Gagal menspesifikasi adanya *missing value* dalam program komputer.
- *Outlier* bukan merupakan anggota populasi yang kita ambil sebagai sampel
- *Outlier* berasal dari populasi yang kita ambil sebagai sampel, tetapi distribusi dari variabel dalam populasi tersebut memiliki nilai ekstrim dan tidak berdistribusi normal.

B. Uji Multikolinearitas

Model regresi dikatakan baik jika tidak terjadi multikolinearitas. Uji multikolinearitas dilakukan untuk menguji apakah terdapat korelasi antar variabel bebas dalam model regresi. Multikolinearitas berarti adanya hubungan linier yang sempurna antara beberapa atau semua variabel yang menjelaskan model regresi (Ajija, 2011). Besaran (quality) yang dapat digunakan untuk mendeteksi adanya multikolinearitas adalah faktor inflasi ragam (Variance Inflation Factor / VIF). VIF digunakan sebagai kriteria untuk mendeteksi multikolinearitas pada regresi linier yang melibatkan lebih dari dua variabel bebas. Nilai VIF lebih besar dari 10 mengidentifikasi adanya masalah multikolinearitas yang serius (Ryan, 1997). VIF untuk koefisien regresi-j diidentifikasikan sebagai berikut:

$$VIF_j = \frac{1}{1-Rj^2}$$

Dengan : *Rj*2 adalah koefisien determinasi antara Xj dengan variabel bebas lainnya pada persamaan / model dugaan ; dimana j = 1,2,..., p.

C. Uji Normalitas

Uji Normalitas adalah sebuah uji yang dilakukan dengan tujuan untuk menilai sebaran data pada sebuah kelompok data atau variabel, apakah sebaran data tersebut berdistribusi normal atau tidak. Uji statistik yang dapat digunakan diantaranya adalah: Uji Chi-Square, Kolmogorov Smirnov, Lilliefors, Shapiro Wilk, Jarque Bera.

D. Uji Linearitas

Uji linearitas merupakan salah satu syarat yang dilakukan dalam analisis korelasi atau regresi linear. Pengujian ini bertujuan untuk mengetahui apakah variabel yang sudah ditetapkan dalam hal ini yaitu variabel independen dan variabel dependen memiliki hubungan yang linear atau tidak secara signifikan. Simpulan data dapat dikatakan linear apabila memiliki taraf signifikansi linearitas lebih kecil dari 0,05 ($p<0,05$).

E. Uji Homoskedastisitas

Salah satu asumsi klasik adalah homoskedastisitas atau non heteroskedastisitas yaitu asumsi yang menyatakan bahwa varian setiap sisaan (ε_i) masih tetap sama baik untuk nilai-nilai pada variabel independen yang kecil maupun besar. Asumsi ini dapat ditulis sebagai berikut:

$$var\,(\varepsilon_1) = \sigma^2,\ i = 1, 2, \dots, n$$

untuk n menunjukkan jumlah observasi. Salah satu cara menguji kesamaan variansi yaitu dengan melihat pola tebaran sisaan terhadap nilai estimasi Y. Jika tebaran sisaan bersifat acak (tidak membentuk pola tertentu), maka dikatakan bahwa variansi sisaan homogen.

2.4 Model Regresi Linier dengan Metode Backward

Pada analisis regresi linier berganda terdapat beberapa metode yang digunakan dalam menentukan persamaan regresi linier berganda, salah satunya diantaranya menggunakan Metode *Backward*. Metode *Backward* adalah metode eliminasi langkah mundur dengan meregresikan seluruh variabel bebas dengan variabel terikat (Sembiring, 1995). Metode *Backward* adalah salah satu metode regresi yang baik digunakan karena menerangkan perilaku variabel terikat dengan sebaik-baiknya yaitu memilih variabel bebas dari banyaknya variabel bebas yang tersedia.

Metode Backward Elimination memiliki proses yaitu dengan memasukkan semua variabel bebas ke dalam model kemudian mengeliminasi satu per satu variabel bebas/independen dari model regresi yang terbentuk (Ekanayake, Rankothge, Weliwatta, & Jayasinghe, 2021). Langkah mundur adalah salah satu metode dengan memasukkan seluruh variabel bebas ke dalam model selanjutnya dikeluarkan secara satu persatu dengan menguji terhadap parameter- parameternya dan menggunakan *Fparsial*. Metode *Backward* lebih praktis digunakan daripada metode kemungkinan regresi lainnya, ini berarti bahwa metode ini hanya memeriksa regresi paling baik yang memuat sejumlah peubah peramal tertentu (N.R. Draper & H. Smith, 1992)

Langkah-langkah melakukan prosedur regresi menggunakan Metode *Backward* yaitu:

a. Membentuk persamaan regresi linier berganda variabel bebas, yaitu *X1, X2, ... , Xk*. Kemudian membentuk koefisien korelasi ganda, lalu uji keberartian regresi ganda, yaitu *b1, b2, ... ,bk*. Jika antara variabel (*X* dan *Y*) memiliki skala pengukuran paling sedikit interval dan hubungannya linier, maka keeratan hubungan antara variabel (*X* dan *Y*) dapat diperoleh menggunakan formulasi korelasi pearson sebagai berikut:

$$r = \frac{n\sum_{i=1}^{n} Y_i X_i - \sum_{i=1}^{n} Y_i \sum_{i=1}^{n} X_i}{\sqrt{n\sum_{i=1}^{n} Y_1^2 - (\sum_{i=1}^{n} Y_1^2)}\sqrt{n\sum_{i=1}^{n} X_1^2 - (\sum_{i=1}^{n} X_1^2)}}$$

b. Menentukan nilai dari parsial terkecil pertama keluar dari model regresi.

$$F_{parsial} = \frac{a_k^2}{s_k^2}$$

Uji hipotesis:
H0 : Tidak terdapat pengaruh signifikan antara *Yi* dengan *Xi*
H1 : Terdapat pengaruh signifikan antara *Yi* dengan *Xi*

Keputusan:
Jika *Fparsial* ≤ *Ftabel*, maka *H0* diterima Jika *Fparsial* > *Ftabel*, maka *H1* ditolak

c. Membentuk persamaan regresi linier berganda kedua

Jika pada langkah sebelumnya *H0* ditolak, maka proses berhenti dan penduganya yaitu persamaan regresi linier berganda lengkap. Sedangkan, apabila *H0* diterima, maka langkah berikutnya yaitu membentuk persamaan regresi linier berganda dengan memasukkan seluruh variabel *Xi* (untuk i ≠ 1). Sehingga prosedur yang digunakan sama dengan langkah pertama.

d. Pemilihan variabel kedua keluar dari model Dalam menentukan variabel kedua yang keluar dari model dilandaskan pada nilai *Fparsial* dari variabel bebas yang terdapat pada persamaan regresi linier berganda kedua (langkah sebelumnya).

Langkah di atas diulang terus-menerus sehingga diperoleh nilai *Fparsial* terkecil dari variabel bebas lebih besar dari *Ftabel*.

2.5 Faktor-faktor yang mempengaruhi harga sewa rumah

Harga sewa rumah dipengaruhi oleh berbagai faktor yang saling berkaitan. Pertama, sumber air minum yang berkualitas dan tersedia secara mudah sangat penting karena mempengaruhi kenyamanan dan kesehatan penghuni rumah. Rumah dengan akses ke air minum bersih biasanya memiliki nilai sewa yang lebih tinggi (Gu, Shen, & Feng, 2018).

Fasilitas sanitasi, seperti tempat buang air besar, juga mempengaruhi nilai sewa. Rumah dengan fasilitas sanitasi yang baik cenderung lebih diminati dan memiliki harga sewa yang lebih tinggi. Selain itu, pengelolaan limbah yang baik dan tempat pembuangan sampah yang memadai turut meningkatkan nilai sewa rumah karena menjaga kebersihan dan kesehatan lingkungan (MDPI, 2022).

Renovasi rumah meningkatkan kondisi fisik dan estetika rumah, yang pada gilirannya meningkatkan harga sewa. Rumah yang telah direnovasi cenderung lebih menarik bagi penyewa. Kesehatan lingkungan sekitar, seperti kebersihan dan ketiadaan genangan air, juga memainkan peran penting dalam menentukan harga sewa rumah (MDPI, 2022).

Kondisi rumah secara keseluruhan, termasuk pemeliharaan dan usia bangunan, mempengaruhi nilai sewa. Rumah yang terawat baik dan dalam kondisi fisik yang baik cenderung memiliki harga sewa yang lebih tinggi (Gu, Shen, & Feng, 2018). Luas lantai rumah adalah faktor lain yang signifikan, rumah dengan luas lantai yang lebih besar biasanya memiliki nilai sewa yang lebih tinggi (Li, Li, & Nuttapong, 2022).

Jenis material yang digunakan untuk lantai, dinding, dan atap rumah juga mempengaruhi harga sewa. Material yang berkualitas tinggi dan tahan lama meningkatkan nilai estetika dan kenyamanan rumah, sehingga meningkatkan harga sewanya (MDPI, 2022; Gu, Shen, & Feng, 2018).

Secara keseluruhan, kombinasi dari faktor-faktor ini menentukan nilai sewa rumah. Pemahaman tentang faktor-faktor ini sangat penting bagi pengambilan keputusan dalam investasi properti dan pengembangan kebijakan perumahan yang efektif.

2.6 Perbedaan wilayah urban dan rural dalam harga sewa rumah

Perbedaan harga sewa rumah antara wilayah urban dan rural dipengaruhi oleh berbagai faktor yang unik untuk masing-masing wilayah. Di wilayah urban, harga sewa cenderung lebih tinggi karena aksesibilitas yang lebih baik ke fasilitas umum seperti transportasi, sekolah, pusat perbelanjaan, dan layanan kesehatan. Selain itu, kepadatan penduduk dan permintaan yang tinggi di daerah perkotaan juga berkontribusi pada tingginya harga sewa rumah. (Huang & Wang, 2022)

Di sisi lain, harga sewa di wilayah rural biasanya lebih rendah karena keterbatasan akses ke fasilitas umum dan infrastruktur yang kurang memadai. Namun, faktor seperti keindahan alam dan luas tanah yang lebih besar dapat meningkatkan nilai sewa di beberapa daerah pedesaan. Faktor ekonomi, seperti tingkat pendapatan per kapita dan tingkat pembangunan ekonomi daerah, juga memainkan peran penting dalam menentukan harga sewa rumah di wilayah rural.

BAB III

PEMBAHASAN

3.1 Populasi dan sampel

Populasi dalam penelitian ini mencakup seluruh rumah tangga di Indonesia dengan sampel yang diambil dari populasi rumah tangga yang tercakup dalam Indonesia Family Life Survey (IFLS) terbaru. IFLS adalah survei panel yang dilakukan secara periodik untuk mengumpulkan data mengenai berbagai aspek kehidupan rumah tangga di Indonesia. Survei ini mencakup informasi demografis, sosial, ekonomi, kesehatan, dan karakteristik perumahan rumah tangga. Teknik pengambilan sampel yang digunakan adalah teknik stratified random sampling, di mana populasi dibagi menjadi strata-strata berdasarkan karakteristik tertentu seperti lokasi (urban dan rural) dan jenis perumahan. Kemudian, sampel diambil secara acak dari masing-masing strata untuk memastikan representasi yang seimbang. (Strauss & Sikoki, 2016)

3.2 Variabel

Variabel adalah karakteristik atau atribut yang dapat diukur atau diobservasi dan memiliki nilai yang berbeda-beda antara satu entitas dengan entitas lainnya. Dalam penelitian, variabel dibagi menjadi dua jenis utama: variabel independen (bebas) dan variabel dependen (terikat) sebagai berikut:

- Variabel Independen (X)
  Variabel independen merupakan variabel yang mempengaruhi variabel dependen atau variabel terikat. Menurut Sugiyono (2019:61) variabel independen adalah variabel-variabel yang mempengaruhi atau yang menjadi sebab perubahannya atau timbulnya variabel dependen (terikat).Variabel independen dalam penelitian ini adalah sumber air minum (X3.1), sumber air mandi (X4.2), tempat BAB (X5.1), tempat air limbah (X6.1, X6.2, X6.3), cara pengambilan sampah (X7), status renovasi (X8), kesehatan lingkungan (X10), kondisi rumah/lingkungan (X11.1, X11.2, X11.3), luas lantai (X12), jumlah ruangan (X13), jenis lantai (X14.1, X14.2, X14.3), jenis dinding (X15.1), dan jenis atap (X16.1, X16.3).
- Variabel Dependen (Y)
  Menurut Sugiyono (2019:39) variabel dependen sering disebut sebagai variabel output, kriteria dan konsekuen. Dalam bahasa Indonesia sering disebut sebagai variabel terikat.Variabel terikat merupakan variabel yang dipengaruhi atau yang menjadi akibat karena adanya variabel bebas.Variabel dependen pada penelitian ini adalah harga sewa rumah per bulan (Y).

Setelah melakukan seleksi variabel, penulis memutuskan untuk menggunakan data daerah urban untuk dianalisis. Daerah urban umumnya memiliki kepadatan penduduk yang lebih tinggi dibandingkan daerah rural. Hal ini menciptakan permintaan yang lebih besar untuk perumahan dan, oleh karena itu, harga sewa rumah cenderung lebih bervariasi dan sensitif terhadap perubahan pasar. (Brueckner, 2000).

Tabel 1. Meta Data

No	Nama Variabel	Kode Variabel	Sumber	Definisi
1	Sumber air minum	kr13;kr15	Book 2 (Household Economy)	X3.1= 1 jika menggunakan ledeng, x = 0 lainnya
2	Sumber air mandi	kr17	Book 2 (Household Economy)	X4.2 = 1 Jika menggunakan sumur pompa, X = 0 lainnya
3	Tempat bab	kr20	Book 2 (Household Economy)	X5.1 = 1 Jika memiliki tempat pembuangan air besar dengan septic tank, X = 0 lainnya
4	Tempat limbah	kr21	Book 2 (Household Economy)	X6.1 = 1 Jika selokan mengalir, 0 = lainnya;
				X6.2 = 1 Jika selokan tidak mengalir, 0 = lainnya;
				X6.3 = 1 jika terdapat lubang permanent, 0 = lainnya
5	Tempat pembuangan sampah	kr22	Book 2 (Household Economy)	X7 = 1 Jika sampah diangkut oleh petugas, 0 = Tidak diangkut oleh petugas
6	Renovasi	kr24b	Book 2 (Household Economy)	X8 = 1 Jika melakukan renovasi, 0 = tidak melakukan renovasi
7	Kesehatan Lingkungan	krk02a-krk02d	Buku K	X10 = Skor kualitas kesehatan lingkungan; 3 = baik 4 = sangat baik
8	Kondisi Rumah	krk02e-krk02f	Buku K	X11.1 = 1 Jika ventilasi cukup, 0 lainnya;
				X11.2 = 1 Jika halaman terawat dan cukup, 0 lainnya;
				X11.3 = 1 Jika rumah dapur pisah, 0 lainnya;
9	Luas Lantai	krk05a	Buku K	X12 = Luas lantai 50 m^2 - 200 m^2
10	Kondisi Ruangan	krk06	Buku K	X13 = Lebih dari 2 ruangan
11	Jenis lantai	krk08	Buku K	X14.1 = 1 jika lantai menggunakan ceramic/marmer/granit, 0 = lainnya;
				X14.2 = 1 jika lantai menggunakan ubin, 0 = lainnya;
				X14.3 = 1 jika lantai menggunakan cement , 0 = lainnya;

12	Jenis dinding	krk09	Buku K	X15.1 = 1 Jika dinding berjenis Masonry, 0 = lainnya;
13	Jenis atap	krk10	Buku K	X16.1 = 1 jika atap berjenis metal plat, 0 = lainnya;
				X16.3 = 1 jika atap berjenis roof tiles, 0 = lainnya
14	Harga sewa	kr05a	Book 2 (Household Economy)	Y = Harga sewa per bulan dan harga 0 dibuang

3.3 Model Regresi Awal

```
> b
              [,1]
       -489608.600
X3.1  -164084.127
X4.2  -183541.442
X5.1    99593.792
X6.1    75387.051
X6.2   -39919.028
X6.3    11095.747
X7     205162.586
X8      70270.662
X10    155687.719
X11.1 -115455.843
X11.2  -41932.790
X11.3  -52182.200
X12      3122.008
X13     78039.103
X14.1 -157056.326
X14.2    2509.859
X14.3 -174114.074
X15.1   64075.797
X16.1 -411882.803
X16.3 -409521.430
```

$$y = -489608,6 - 164084,127(X3.1) - 1835141,442(X4.2) + 99593,792(X5.1) + 75385,051(X6.1) - 39919,028(X6.2) + 11095,747(X6.3) + 205162,586(X7) + 70270,662(X8) + 155687,719(X10) - 115455,843(X11.1) - 41932,79(X11.2) - 52182,2(X11.3) + 3122,008(X12) + 78039,103(X13) - 157056,326(X14.1) + 2509,859(X14.2) - 174114,074(X14.3) + 64075,797(X15.1) - 411882,803(X16.1) - 409521,43(X16.3)$$

```
> R2       <- SSR/SST;R2
           [,1]
[1,] 0.03599254
```

Model tersebut memiliki R square sebesar 0.003599254 atau 3% menunjukkan bahwa sekitar 3% dari variasi dapat dijelaskan oleh model tersebut.

3.4 Uji Asumsi Klasik

A. Outlier

1. Hipotesis

   H_0: y_i bukan merupakan outlier

   H_1 : y_i adalah outlier

2. Taraf Signifikan

   α = 5%

3. Statistik Uji

```
> ## studentized deleted residual
> MSEi      <- (matrix(c(SSE),n,1)-e*e/diag(I-H))/(n-p-1)
> vd <- MSEi/diag(I-H)
> t  <- d/sqrt(vd)
> alpha     <- 0.05
> t_cri     <- qt(1-alpha/n/2,n-p-1)
> outlier <- abs(t)>=t_cri
> which(abs(t)>=t_cri)
[1]  483  603 1271 1973 2061 2219 2270 2635 2653 2738 2858
```

4. Kriteria Uji

   Tolak H_0 jika $|t_i| \geq$ t_cri

5. Keputusan

   H_0 ditolak untuk data ke-483, 603, 1271, 1973, 2061, 2219, 2270, 2653, 2738, dan 2858 karena ti > t_cri, selain itu H_0 diterima.

6. Kesimpulan

   Dengan derajat kepercayaan 95%, dapat disimpulkan bahwa terdapat 11 outlier pada data, yaitu pada data ke-483, 603, 1271, 1973, 2061, 2219, 2270, 2653, 2738, dan 2858. Hal tersebut menjelaskan bahwa ada pengamatan yang signifikan secara statistik berbeda dari pola umum data, yang dapat mempengaruhi hasil analisis dan interpretasi.

B. Uji Multikolinearitas

Multikolinearitas terjadi ketika ada korelasi tinggi antara variabel-variabel independen dalam model yang dapat menyebabkan ketidakstabilan dalam estimasi koefisien regresi.

```
> detX

[1] 8.767004e+60

> detX>0       # tidak ada multikol sempurna
```

```
[1] TRUE
```

VIF mengukur seberapa banyak variabilitas koefisien regresi meningkat karena adanya kolerasi antara variabel independen. Jika VIF > 5 menunjukkan bahwa variabel tersebut memiliki kolerasi yang kuat dengan variabel independen lainnya dan dapat menyebabkan multikolinearitas.

```
> VIF
    X3.1     X4.2     X5.1     X6.1     X6.2     X6.3 
1.127037 1.167441 1.142679 1.896612 1.111621 1.741991 
      X7       X8      X10    X11.1    X11.2    X11.3 
1.175055 1.046578 1.050707 1.529754 1.495832 1.068268 
     X12      X13    X14.1    X14.2    X14.3    X15.1 
1.133682 1.197202 1.363207 1.569638 1.443814 1.169039 
   X16.1    X16.3 
2.651037 2.689365 
> VIF>=5       # ada gejala multikol kuat
 X3.1  X4.2  X5.1  X6.1  X6.2  X6.3    X7    X8   X10 X11.1 
FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE 
X11.2 X11.3   X12   X13 X14.1 X14.2 X14.3 X15.1 X16.1 X16.3 
FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE 
```

Berdasarkan nilai VIF yang diperoleh, tidak ada nilai VIF yang lebih dari 5. Sehingga dapat disimpulkan tidak ada gejala multikolinearitas pada data, hal tersebut mengindikasikan bahwa variabel independen dapat dianggap relatif independen satu sama lain, sehingga memperkuat keandalan analisis regresi yang dilakukan.

C. Uji Homoskedastisitas

Asumsi yang menyatakan bahwa varian setiap sisaan (ε_i) masih tetap sama baik untuk nilai-nilai pada variabel independen yang kecil maupun besar.

1. Hipotesis

   H_0: Homoskedastisitas

   H_1: Heteroskedastisitas
2. Taraf signifikan

   $\alpha = 5\%$
3. Statistik Uji

```
> k
[1] 20
> R2
           [,1]
[1,] 0.01011011
> chi2
         [,1]
[1,] 32.15016
> chi_c
[1] 31.41043
```

4. Kriteria uji
   Tolak H_0 jika $p - value \leq 0,05$ atau chi_2 > chi_c
5. Keputusan
   Karena chi_2 (32,15016) > chi_c (31,41043), maka H_0 ditolak.
6. Kesimpulan
   Dengan menggunakan Uji Breush-Pagan test dan taraf signifikansi 5% dapat disimpulkan bahwa tolak H_0 , artinya data heteroskedastisitas

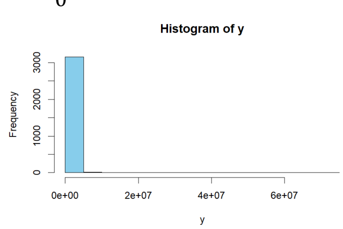
D. Uji Linearitas

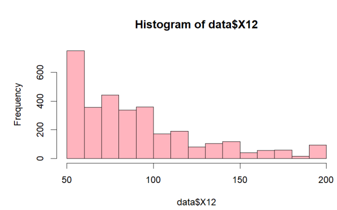
Untuk mengetahui apakah dua variabel atau lebih yang diuji mempunyai hubungan yang linear atau tidak secara signifikan.

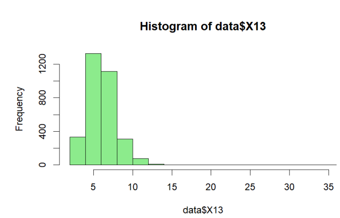
1. Hipotesis :
   H_0: Model linear
   H_1: Model tidak linear
2. Taraf signifikan
   $\alpha = 5\%$
3. Statistik Uji

```
> f1       <- (SSE0-SSE1)/(q-1)/(SSE1/(n-k-q));f1
          [,1]
[1,] 54.27107
> #Menentukan kriteria uji
> alpha    <- 0.05
> fcrit    <- qf(1-alpha,q-1,n-k-q);fcrit
[1] 2.998577
> p_val1<- 1-pf(f1,q-1,n-k-q);p_val1
     [,1]
[1,]    0
```

4. Kriteria Uji
   Tolak H_0 jika $fhitung > fcrit$, terima dalam hal lainnya
5. Keputusan
   Karena $fhitung$ $(54, 27) > fcrit$ $(2, 9985)$, maka terima H_0
6. Kesimpulan
   Berdasarkan uji linearitas dengan taraf signifikan 5% dapat disimpulkan bahwa tolak H_0 artinya hubungan antara variabel X dan Y bersifat tidak linear.







Grafik 1, 2, 3 Histogram Data Numerik

Berdasarkan histogram di atas, dapat dilihat bahwa nilai dari variabel numerik menceng kanan sehingga harus ditransformasi menggunakan Logaritma.

E. Uji Normalitas

Pengujian normalitas sering dianggap tidak diperlukan dalam regresi linier jika ukuran sampel cukup besar ($n \geq 30$). Berdasarkan teorema limit pusat, distribusi sampling rata-rata mendekati normal meskipun populasi asal tidak normal. Ini memungkinkan penggunaan uji statistik parametrik, seperti regresi linier, tanpa memastikan data asli berdistribusi normal terlebih dahulu.

3.5. Transformasi Log-log model: log y(log x)

Setelah dilakukan uji asumsi klasik, kita ketahui bahwa data tidak memenuhi asumsi homoskedastisitas dan asumsi linearitas. Oleh karena itu, kita gunakan log-log model untuk memenuhi asumsi asumsi tersebut.

a. Uji Linearitas

Untuk mengetahui apakah dua variabel atau lebih yang diuji mempunyai hubungan yang linear atau tidak secara signifikan. Terlebih dahulu lakukan transformasi pada variabel numerik yaitu Y, X12 (luas lantai), dan X13 (jumlah ruangan). Dengan menggunakan software excel diperoleh head data:

Tabel 2. Hasil Transformasi Data Numerik

log(y)	log(X12)	log(X13)
5.39794	1.90309	0.60206
5.653213	1.845098	0.69897
4.920819	1.748188	0.60206
5.69897	1.845098	0.845098
5.619789	2.133539	1.041393
5.318759	1.924279	0.69897
5.477121	1.924279	0.778151
5.221849	1.845098	0.845098
5.318759	1.732394	0.60206

1. Hipotesis :

   H_0: Model linear

   H_1: Model tidak linear

2. Taraf signifikan

   $\alpha = 5\%$

3. Statistik Uji

```
> #Menghitung Statistik uji
> f1  <- (SSE0-SSE1)/(q-1)/(SSE1/(n-k-q))
> f1
         [,1]
[1,] 1.629352
> #Menentukan kriteria uji
> alpha   <- 0.05
> fcrit   <- qf(1-alpha,q-1,n-k-q)
> fcrit
[1] 2.998588
> p_val1<- 1-pf(f1,q-1,n-k-q)
> p_val1
          [,1]
[1,] 0.1962221
```

4. Kriteria Uji

   Tolak H_0 jika $fhitung > fcrit$, terima dalam hal lainnya

5. Keputusan

   $fhitung < fcrit$, maka terima H_0

6. Kesimpulan

   Berdasarkan uji linearitas dengan taraf signifikan 5% dapat disimpulkan bahwa terima H_0 artinya hubungan antara variabel X dan Y bersifat linear setelah variabel numerik ditransformasi menggunakan logaritma.

b. Uji Homoskedastisitas

Asumsi yang menyatakan bahwa varian setiap sisaan (ε_i) masih tetap sama baik untuk nilai-nilai pada variabel independen yang kecil maupun besar.

1. Hipotesis

   H_0: Homoskedastisitas

   H_1: Heteroskedastisitas

2. Taraf signifikan

   $\alpha = 5\%$

3. Statistik Uji

```
> k
[1] 20
> R2
           [,1]
[1,] 0.00829995
```

```
> chi2
          [,1]
[1,] 26.29424
> chi_c
[1] 31.41043
```

4. Kriteria uji
   Tolak H_0 jika $p - value \leq 0,05$ atau chi_2 > chi_c
5. Keputusan
   Karena chi_2 (26,29424) > chi_c (31,41043), maka H_0 diterima.
6. Kesimpulan
   Dengan menggunakan Uji Breusch-Pagan test dan taraf signifikansi 5% dapat disimpulkan bahwa tolak H_0 , artinya data homoskedastisitas setelah variabel numerik ditransformasi menggunakan logaritma.

3.6 Model Regresi Linier dengan Metode Backward

a. Cycle 1(menghitung semua eksplanatori variabel)
   Step 1: mencari p-value dari hasil semua eksplanatori variabel

```
> p_v
                  [,1]
          0.000000e+00
X3.1      1.871709e-01
X4.2      4.043602e-06
X5.1      1.437478e-06
X6.1      3.448090e-03
X6.2      8.677440e-03
X6.3      6.609688e-01
X7        0.000000e+00
X8        7.462840e-02
X10       1.551128e-03
X11.1     2.818503e-01
X11.2     4.769012e-01
X11.3     1.808818e-01
log.X12.  2.492216e-04
log.X13.  1.607504e-08
X14.1     3.012677e-06
X14.2     8.908510e-01
X14.3     5.085710e-12
X15.1     3.078399e-02
X16.1     9.283859e-02
X16.3     1.254192e-06
```

Interpretasi : Dari perhitungan seleksi variabel, Variabel X4.2, X5.1, X6.1, X6.2, X7, X10, log.X12, log.X13, X14.1, X14.3, X15.1, dan X16.1 dikeluarkan dari model karena p-value lebih besar dari alpha=0.05 yang artinya variabel tersebut tidak signifikan.

b. Cycle 2 (Mencari model terbaik setelah variabel tidak signifikan dikeluarkan dari model)
   Step 1: mencari p-value tanpa variabel X4.2, X5.1, X6.1, X6.2, X7, X10, log.X12, log.X13, X14.1, X14.3, X15.1, dan X16.1

```
> p_v
                  [,1]
          0.000000e+00
X4.2      4.933299e-06
X5.1      5.811401e-07
X6.1      2.661989e-05
X6.2      4.071855e-03
X7        0.000000e+00
X10       1.921443e-03
log.X12.  4.244857e-04
log.X13.  3.858926e-09
X14.1     1.973326e-08
X14.3     4.440892e-16
X15.1     2.373847e-02
X16.3     1.429299e-08
```

   Interpretasi : Tidak ada variabel yang dikeluarkan, semua variabel telah signifikan dimana p-value dari alpha = 0.05. Jadi model terbaik prediksi ini adalah model dari cycle 2 step 1.

c. Koefisien Determinasi

```
> R2  <- 1-MSE/var(y)
> R2
          [,1]
[1,] 0.1905938
```

   Interpretasi: Berdasarkan hasil perhitungan, nilai R^2 sebesar 0,19 menunjukkan bahwa hanya 19% dari total variasi harga rumah yang dapat dijelaskan oleh model ini. Dengan kata lain, 81% variasi harga rumah disebabkan oleh faktor lain yang tidak termasuk dalam model ini. R^2 yang rendah menunjukkan bahwa model ini memiliki kemampuan prediksi yang lemah dalam menjelaskan harga rumah.

d. Uji Asumsi Klasik
   - **Uji Linearitas**
     Untuk mengetahui apakah dua variabel atau lebih yang terpilih setelah seleksi variabel mempunyai hubungan yang linear atau tidak secara signifikan.
     1. Hipotesis :
        H_0: Model linear
        H_1: Model tidak linear
     2. Taraf signifikan

$\alpha = 5\%$

3. Statistik Uji

```
> #Menghitung Statistik uji
> f1  <- (SSE0-SSE1)/(q-1)/(SSE1/(n-k-q))
> f1
         [,1]
[1,] 1.930982
> #Menentukan kriteria uji
> alpha   <- 0.05
> fcrit   <- qf(1-alpha,q-1,n-k-q)
> fcrit
[1] 2.99858
> p_val1<- 1-pf(f1,q-1,n-k-q)
> p_val1
          [,1]
[1,] 0.1451772
```

4. Kriteria Uji

   Tolak H_0 jika $fhitung > fcrit$, terima dalam hal lainnya

5. Keputusan

   $fhitung\ (1,93) < fcrit\ (2,9958)$, maka terima H_0

6. Kesimpulan

   Berdasarkan uji linearitas dengan taraf signifikan 5% dapat disimpulkan bahwa terimaH_0 artinya hubungan antara variabel X dan Y bersifat linear setelah dilakukan *backward elimination*.

- Homoskedastisitas

1. Hipotesis

   H_0: Homoskedastisitas

   H_1: Heteroskedastisitas

2. Taraf signifikan

   $\alpha = 5\%$

3. Statistik Uji

```
> k
[1] 12
> R2
            [,1]
[1,] 0.004373701
> chi2
         [,1]
[1,] 13.85588
> chi_c
[1] 21.02607
```

4. Kriteria uji

   Tolak H_0 jika $p - value \leq 0,05$ atau chi_2 > chi_c

5. Keputusan

   Karena chi_2 (13,85588) < chi_c (21,02607), maka H_0 diterima.

6. Kesimpulan

Dengan menggunakan Uji Breusch-Pagan test dan taraf signifikansi 5% dapat disimpulkan bahwa terima H_0, data menunjukkan adanya homoskedastisitas setelah dilakukan *backward elimination*.

e. Model Terbaik Setelah Seleksi dan Pemenuhan Asumsi Klasik

```
> bh
                [,1]
          4.62965759
X4.2     -0.06691313
X5.1      0.10212490
X6.1      0.06055218
X6.2      0.12932303
X7        0.19323747
X10       0.04921327
log.X12.  0.15805853
log.X13.  0.33760762
X14.1    -0.11134342
X14.3    -0.15360537
X15.1     0.06317210
X16.3    -0.08643872
```

Berdasarkan hasil perhitungan stepwise dengan metode *Backward* diperoleh model terbaik yaitu:

$$\hat{y} = 4.62965759 - 0.06691313(X_{4.2}) + 0.10212490(X_{5.1}) + 0.06055218(X_{6.1}) + 0.12932303(X_{6.2}) + 0.19323747(X_{7}) + 0.04921327(X_{10}) + 0.15805853(log(X_{12})) + 0.33760762(log(X_{13})) - 0.11134342(X_{14.1}) - 0.15360537(X_{14.3}) + 0.06317210(X_{15.1}) - 0.08643872(X_{16.3})$$

Model ini menunjukkan bahwa harga rumah dipengaruhi oleh beberapa variabel independen. Variabel seperti X4.2, X14.1, X14.3, dan X16.3 memiliki pengaruh negatif terhadap harga rumah, sementara variabel lainnya memiliki pengaruh positif. Penggunaan logaritma pada variabel X12 dan X13 menunjukkan bahwa hubungan antara variabel-variabel ini dengan harga rumah bersifat log-linear.

BAB IV

PENUTUP

4.1 Kesimpulan

Berdasarkan hasil analisis dapat disimpulkan bahwa data yang diambil dari Indonesia Family Life Survey (IFLS) khususnya di wilayah urban (perkotaan) memiliki model regresi sebagai berikut:

$$\hat{y} = 4.62965759 - 0.06691313(X_{4.2}) + 0.10212490(X_{5.1}) + 0.06055218(X_{6.1}) + 0.12932303(X_{6.2}) + 0.19323747(X_{7}) + 0.04921327(X_{10}) + 0.15805853(log(X_{12})) + 0.33760762(log(X_{13})) - 0.11134342(X_{14.1}) - 0.15360537(X_{14.3}) + 0.06317210(X_{15.1}) - 0.08643872(X_{16.3})$$

Model regresi ini mengindikasikan bahwa harga rumah dipengaruhi oleh berbagai variabel independen. Beberapa variabel, seperti X4.2, X14.1, X14.3, dan X16.3, menunjukkan pengaruh negatif terhadap harga rumah. Sebaliknya, variabel lainnya memberikan pengaruh positif. Penggunaan transformasi logaritma pada variabel X12 dan X13 mengungkapkan bahwa hubungan antara variabel-variabel tersebut dan harga rumah bersifat log-linear.

Model tersebut diperoleh dari proses seleksi variabel menggunakan metode Backward Elimination. Proses ini melibatkan:

1. Memulai dengan semua variabel potensial dalam model regresi.
2. Mengeliminasi satu per satu variabel yang tidak signifikan berdasarkan nilai p-value hingga hanya tersisa variabel-variabel yang signifikan.
3. Melakukan uji asumsi klasik untuk memastikan bahwa model memenuhi asumsi-asumsi dasar regresi, seperti tidak adanya multikolinearitas yang kuat, homoskedastisitas, dan linearitas.

Model yang dihasilkan memiliki nilai R^2 sebesar 19%, yang menunjukkan bahwa hanya 19% variasi harga sewa rumah dapat dijelaskan oleh model ini. Dengan demikian, model ini memiliki kemampuan prediksi yang terbatas. Pada analisis ini uji asumsi klasik menunjukkan adanya masalah heteroskedastisitas dan linieritas pada data asli ditangani dengan transformasi logaritma pada variabel numerik. Untuk meningkatkan kemampuan prediksi model, disarankan menambahkan variabel lain yang mungkin berpengaruh signifikan terhadap harga sewa rumah. Selain itu, analisis lebih lanjut diperlukan untuk mengatasi masalah heteroskedastisitas dan memastikan semua asumsi klasik regresi terpenuhi dengan baik. Dengan langkah-langkah ini, diharapkan model prediksi harga sewa rumah yang dihasilkan dapat menjadi lebih akurat dan andal.

DAFTAR PUSTAKA

Gu, H., Shen, T., & Feng, C. (2018). The impact of micro-level influencing factors on home value: A housing price-rent comparison. *Sustainability*, 10(12), 4343. https://doi.org/10.3390/su10124343

Li, N., Li, R.Y.M., & Nuttapong, J. (2022). Factors affect the housing prices in China: A systematic review of papers indexed in Chinese Science Citation Database. *Property Management*, 40(5), 780-796. https://doi.org/10.1108/PM-11-2020-0078

MDPI. (2022). Understanding the effects of influential factors on housing prices by combining extreme gradient boosting and a hedonic price model (XGBoost-HPM). *Land*. https://doi.org/10.3390/land11112241

Huang, L., Zheng, M., & Wang, R. (2022). Rural Housing Rental Rates in China: Regional Differences, Influencing Factors, and Policy Implications. *Land*, 11(7), 1053. https://doi.org/10.3390/land11071053

Strauss, J., Witoelar, F., & Sikoki, B. (2016). The Fifth Wave of the Indonesia Family Life Survey (IFLS5): Overview and Field Report. RAND Corporation.

Brueckner, J. K. (2000). Urban Sprawl: Diagnosis and Remedies. International Regional Science Review, 23(2), 160-171. https://doi.org/10.1177/016001700761012710

LAMPIRAN

File dan data dapat diakses di:
https://bit.ly/PROJEKANREG2024

Syntax dan output R:

```
> data<-read.csv("C:/Users/Asus/OneDrive/documents/awal.csv")
> head(data)
          Y X3.1 X4.2 X5.1 X6.1 X6.2 X6.3 X7 X8 X10 X11.1 X11.2 X11.3 X12 X13
1 250000.00    0    0    1    1    0    0  1  0   4     0     1     0  80   4
2 450000.00    0    0    1    1    0    0  1  1   4     0     0     0  70   5
3  83334.23    0    0    1    1    0    0  0  1   3     1     1     1  56   4
4 500000.00    0    1    1    1    0    0  0  1   4     1     1     0  70   7
5 416666.67    0    0    1    1    0    0  0  1   4     0     0     0 136  11
6 208334.23    0    0    1    1    0    0  0  0   4     1     1     0  84   5
  X14.1 X14.2 X14.3 X15.1 X16.1 X16.3
1     1     0     0     1     1     0
2     0     1     0     1     0     1
3     0     0     1     1     0     0
4     0     1     0     1     0     1
5     0     1     0     1     0     1
6     0     0     1     0     1     0
> #Membentuk vektor y
> y    <- as.matrix(data[,1])
> #Membentuk matriks X
> ##Kolom matriks X dari data
> X0   <- as.matrix(data[,-1])
> head(X0)
     X3.1 X4.2 X5.1 X6.1 X6.2 X6.3 X7 X8 X10 X11.1 X11.2 X11.3 X12 X13 X14.1
[1,]    0    0    1    1    0    0  1  0   4     0     1     0  80   4     1
[2,]    0    0    1    1    0    0  1  1   4     0     0     0  70   5     0
[3,]    0    0    1    1    0    0  0  1   3     1     1     1  56   4     0
[4,]    0    1    1    1    0    0  0  1   4     1     1     0  70   7     0
[5,]    0    0    1    1    0    0  0  1   4     0     0     0 136  11     0
[6,]    0    0    1    1    0    0  0  0   4     1     1     0  84   5     0
     X14.2 X14.3 X15.1 X16.1 X16.3
[1,]     0     0     1     1     0
[2,]     1     0     1     0     1
[3,]     0     1     1     0     0
[4,]     1     0     1     0     1
[5,]     1     0     1     0     1
[6,]     0     1     0     1     0
> ##Menentukan banyak unit pengamata (baris dari vektor y)
> n    <- nrow(y)
> ##Membentuk vektor unit konstan
> u    <- matrix(c(1),n,1)
> X    <- cbind(u,X0)
> #Taksiran kodefisien regresi
> b    <- solve(t(X)%*%X)%*%t(X)%*%y;b
             [,1]
      -489608.600
X3.1  -164084.127
X4.2  -183541.442
X5.1    99593.792
X6.1    75387.051
X6.2   -39919.028
X6.3    11095.747
X7     205162.586
X8      70270.662
X10    155687.719
```

```
X11.1 -115455.843
X11.2  -41932.790
X11.3  -52182.200
X12      3122.008
X13     78039.103
X14.1 -157056.326
X14.2    2509.859
X14.3 -174114.074
X15.1   64075.797
X16.1 -411882.803
X16.3 -409521.430
> H     <- X%*%solve(t(X)%*%X)%*%t(X) # hat matrix
> I     <- diag(c(1),n,n)                # identity matrix of n-order
> J     <- matrix(c(1),n,n)        # unit constant martix of n-order
> SSE   <- t(y)%*%(I-H)%*%y;SSE       #sum square error
             [,1]
[1,] 1.119824e+16
> SSR   <- t(y)%*%(H-(J/n))%*%y              #sum square regression
> SST   <- SSR+SSE                           #sum square total
> MSE   <- SSE/(n-p)                 #mean square error
> MSR   <- SSR/(p-1)                 #mean square regression
> MST   <- SST/(n-1)                 #mean square total
> R2    <- SSR/SST;R2                              #R-square
           [,1]
[1,] 0.03599254
> #Outlier
> #Taksiran kodefisien regresi
> b     <- solve(t(X)%*%X)%*%t(X)%*%y
>
> #Taksiran standar error
> #Matriks Hat (H)
> H     <- X%*%solve(t(X)%*%X)%*%t(X)
> ##Menghitung banyak regresor+1
> p     <- ncol(X)
>
> ##Menghitung taksiran varians kekeliruan (MSE)
> s2    <- SSE/(n-p)
> s     <- sqrt(s2)
> ##Membentuk matriks identitas (nxn)
> I     <- diag(c(1),n,n)
> SSE   <- t(y)%*%(I-H)%*%y
> p     <- ncol(X)
> s2    <- SSE/(n-p)
> s     <- sqrt(s2)
> e     <- (I-H)%*%y
> ## Semi Studentized residual
> es    <- e/s[1,1]
> ## Semistudentized residual
> es    <- e/s[1,1]
> ## studentized residual
> r     <- e/matrix(c(s),n,1)/sqrt(diag(I-H))        #s dijadikan matrik (nX1)
> r1    <- e/s[1,1]/sqrt(diag(I-H))                  #s dijadikan skalar
> ## deleted residual
> d     <- e/diag(I-H)
> ## studentized deleted residual
> MSEi  <- (matrix(c(SSE),n,1)-e*e/diag(I-H))/(n-p-1)
> vd    <- MSEi/diag(I-H)
> t     <- d/sqrt(vd)
> alpha <- 0.05
> t_cri <- qt(1-alpha/n/2,n-p-1)
> outlier <- abs(t)>=t_cri
> which(t>= t_cri)
```

```
 [1]  483  603 1271 1973 2061 2219 2270 2635 2653 2738 2858
> # Multicolinearitas sempurna
> X     <- cbind(matrix(1,n,1),X0)
> detX  <- det(t(X)%*%X)
> detX
[1] 8.767004e+60
> detX>0       # tidak ada multikol sempurna
[1] TRUE
> # Multicolineairtas kuat
> ##inverst matriks korelasi
> R     <- cor(X0)
> VIF   <- diag(solve(R))
> VIF
    X3.1     X4.2     X5.1     X6.1     X6.2     X6.3 
1.127037 1.167441 1.142679 1.896612 1.111621 1.741991 
      X7       X8      X10    X11.1    X11.2    X11.3 
1.175055 1.046578 1.050707 1.529754 1.495832 1.068268 
     X12      X13    X14.1    X14.2    X14.3    X15.1 
1.133682 1.197202 1.363207 1.569638 1.443814 1.169039 
   X16.1    X16.3 
2.651037 2.689365 
> VIF>=5       # ada gejala multikol kuat
 X3.1  X4.2  X5.1  X6.1  X6.2  X6.3    X7    X8   X10 X11.1 
FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE 
X11.2 X11.3   X12   X13 X14.1 X14.2 X14.3 X15.1 X16.1 X16.3 
FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE 
> e     <- (I-H)%*%y
> e2    <- e*e
> #HOMOSKEDASTISITAS AWAL
> # Uji Breucsh-Pagan
> ##Membentuk vektor y
> y     <- e2                         # regresan merupakan resiudal^2 dari reresi utama
> ##Membentuk matrix Z
> X     <- X                          # regresor dalam model tambahan sama dengan regresor 
dlm model utama
> p     <- ncol(X)
> k     <- p-1                # banyaknya variabel eksplanatori
> ## Menghitung R^2
> H     <- X%*%solve(t(X)%*%X)%*%t(X) #Matriks Hat (H)
> I     <- diag(c(1),n,n)                     #Membentuk matriks identitas (nxn)
> J     <- matrix(c(1),n,n)           # unit constant martix of n-order
> SSE   <- t(y)%*%(I-H)%*%y           #sum square residual
> SSR   <- t(y)%*%(H-(J/n))%*%y               #sum square regression
> SST   <- SSR+SSE                            #sum square total
> R2    <- SSR/SST                            #R-square
> ## Statistik uji
> chi2  <- n*R2
> ## Titik kritis
> chi_c <- qchisq(1-alpha,k)
> k
[1] 20
> R2
           [,1]
[1,] 0.01011011
> chi2
         [,1]
[1,] 32.15016
> chi_c
[1] 31.41043
> # LINEARITAS AWAL
```

```
> SSE0  <- t(y)%*%(I-H)%*%y
> SSE0
             [,1]
[1,] 1.119824e+16
>
> ##Menghitung banyak regresor+1
> p     <- ncol(X)
> k     <- p-1  #banyaknya variabel eksplanatori
>
> #Menghitung Sum Square Residual di bawah H1
> yh    <- H%*%y        #Nilai prediksi di bawah H0
> yh2   <- yh*yh        #yh^2
> yh3   <- yh2*yh       #yh^3
>
> ##Membentuk matriks X di bawah H1
> X1    <- cbind(X,yh2,yh^3)
> q     <- ncol(X1)-ncol(X)+1  #Menghitung ordo tertinggi
>
> #Menghitung Sum Square Residual di bawah H1
> H1    <- X1%*%solve(t(X1)%*%X1, tol=1e-39)%*%t(X1) #Matriks Hat (H)
>
> SSE1  <- t(y)%*%(I-H1)%*%y
> SSE1
             [,1]
[1,] 1.082603e+16
>
> #Menghitung Statistik uji
> f1    <- (SSE0-SSE1)/(q-1)/(SSE1/(n-k-q))
> f1
         [,1]
[1,] 54.27107
>
> #Menentukan kriteria uji
> alpha <- 0.05
> fcrit <- qf(1-alpha,q-1,n-k-q)
> fcrit
[1] 2.998577
>
> p_val1<- 1-pf(f1,q-1,n-k-q)
> p_val1
     [,1]
[1,]    0
>
> #---LOG-LOG MODEL----
> data<-read.csv("C:/Users/Asus/OneDrive/documents/awal.csv")
>
> #Melihat bagian atas data
> head(data)
          Y X3.1 X3.3 X5.1 X6.1 X6.2 X6.3 X7 X8 X10 X11.1 X11.2 X11.3 X12 X13
1 250000.00    0    0    1    1    0    0  1  0   4     0     1     0  80   4
2 450000.00    0    0    1    1    0    0  1  1   4     0     0     0  70   5
3  83333.33    0    0    1    1    0    0  0  1   3     1     1     1  56   4
4 500000.00    0    1    1    1    0    0  0  1   4     1     1     0  70   7
5 416666.67    0    0    1    1    0    0  0  1   4     0     0     0 136  11
6 208333.33    0    0    1    1    0    0  0  0   4     1     1     0  84   5
  X14.1 X14.2 X14.3 X15.1 X16.1 X16.3
1     1     0     0     1     1     0
2     0     1     0     1     0     1
3     0     0     1     1     0     0
```

```
4     0     1     0     1     0     1
5     0     1     0     1     0     1
6     0     0     1     0     1     0
>
> #Membentuk vektor y
> y     <- as.matrix(data[,1])
>
> #Membentuk matriks X
> X0    <- as.matrix(data[,-1])       #Kolom matriks X dari data
> n     <- nrow(y)                            #Menentukan banyak unit pengamata (baris dari vektor
y)
> u     <- matrix(c(1),n,1)           #Membentuk vektor unit konstan
>
> X     <- cbind(u,X0)
>
> #Menghitung Sum Square Residual di bawah H0
> H     <- X%*%solve(t(X)%*%X)%*%t(X)  #Matriks Hat (H)
> I     <- diag(c(1),n,n)                     #Membentuk matriks identitas (nxn)
> e     <- (I-H)%*%y
>
> SSE0  <- t(y)%*%(I-H)%*%y
> SSE0
             [,1]
[1,] 1.119824e+16
>
> ##Menghitung banyak regresor+1
> p     <- ncol(X)
> k     <- p-1  #banyaknya variabel eksplanatori
>
> #Menghitung Sum Square Residual di bawah H1
> yh    <- H%*%y       #Nilai prediksi di bawah H0
> yh2   <- yh*yh       #yh^2
> yh3   <- yh2*yh      #yh^3
>
> ##Membentuk matriks X di bawah H1
> X1    <- cbind(X,yh2,yh^3)
> q     <- ncol(X1)-ncol(X)+1  #Menghitung ordo tertinggi
>
> #Menghitung Sum Square Residual di bawah H1
> H1    <- X1%*%solve(t(X1)%*%X1, tol=1e-39)%*%t(X1)  #Matriks Hat (H)
>
> SSE1  <- t(y)%*%(I-H1)%*%y
> SSE1
             [,1]
[1,] 1.082603e+16
>
> #Menghitung Statistik uji
> f1    <- (SSE0-SSE1)/(q-1)/(SSE1/(n-k-q))
> f1
         [,1]
[1,] 54.27107
>
> #Menentukan kriteria uji
> alpha <- 0.05
> fcrit <- qf(1-alpha,q-1,n-k-q)
> fcrit
[1] 2.998577
>
> p_val1<- 1-pf(f1,q-1,n-k-q)
```

```
> p_val1
     [,1]
[1,]    0
> #####------LOG-LOG MODEL------#####
> data<-read.csv("C:/Users/Asus/OneDrive/documents/akhir.csv")
> head(data)
    log.y. X3.1 X3.3 X5.1 X6.1 X6.2 X6.3 X7 X8 X10 X11.1 X11.2 X11.3 log.X12.
1 5.397940    0    0    1    1    0    0  1  0   4     0     1     0 1.903090
2 5.653213    0    0    1    1    0    0  1  1   4     0     0     0 1.845098
3 4.920819    0    0    1    1    0    0  0  1   3     1     1     1 1.748188
4 5.698970    0    1    1    1    0    0  0  1   4     1     1     0 1.845098
5 5.619789    0    0    1    1    0    0  0  1   4     0     0     0 2.133539
6 5.318759    0    0    1    1    0    0  0  0   4     1     1     0 1.924279
  log.X13. X14.1 X14.2 X14.3 X15.1 X16.1 X16.3
1 0.602060     1     0     0     1     1     0
2 0.698970     0     1     0     1     0     1
3 0.602060     0     0     1     1     0     0
4 0.845098     0     1     0     1     0     1
5 1.041393     0     1     0     1     0     1
6 0.698970     0     0     1     0     1     0
> # taraf signifikansi
> alpha <- 0.05
> # Menghitung residu^2 dari regresi utama
> ##Membentuk vektor y
> y     <- as.matrix(data[,1])
> ##Membentuk matriks X
> X0    <- as.matrix(data[,-1])       #Kolom matriks X dari data
> n     <- nrow(y)                            #Menentukan banyak unit pengamata (baris dari vektor
y)
> u     <- matrix(c(1),n,1)           #Membentuk vektor unit konstan
> X     <- cbind(u,X0)
> ##Menghitung Residual
> H     <- X%*%solve(t(X)%*%X)%*%t(X) #Matriks Hat (H)
> I     <- diag(c(1),n,n)                     #Membentuk matriks identitas (nxn)
> e     <- (I-H)%*%y
> e2    <- e*e
>
> # HOMOSKEDASTISITAS
> # Uji Breucsh-Pagan
> ##Membentuk vektor y
> y     <- e2                                 # regresan merupakan resiudal^2 dari reresi utama
> ##Membentuk matrix Z
> X     <- X                                  # regresor dalam model tambahan sama dengan regresor
dlm model utama
> p     <- ncol(X)
> k     <- p-1                        # banyaknya variabel eksplanatori
> ## Menghitung R^2
> H     <- X%*%solve(t(X)%*%X)%*%t(X) #Matriks Hat (H)
> I     <- diag(c(1),n,n)                     #Membentuk matriks identitas (nxn)
> J     <- matrix(c(1),n,n)           # unit constant martix of n-order
> SSE   <- t(y)%*%(I-H)%*%y           #sum square residual
> SSR   <- t(y)%*%(H-(J/n))%*%y               #sum square regression
> SST   <- SSR+SSE                            #sum square total
> R2    <- SSR/SST                            #R-square
> ## Statistik uji
> chi2  <- n*R2
> ## Titik kritis
> chi_c <- qchisq(1-alpha,k)
> k
```

```
[1] 20
> R2
           [,1]
[1,] 0.00829995
> chi2
         [,1]
[1,] 26.29424
> chi_c
[1] 31.41043
>
> # LINEARITAS
> SSE0  <- t(y)%*%(I-H)%*%y
> SSE0
         [,1]
[1,] 433.8876
> ##Menghitung banyak regresor+1
> p     <- ncol(X)
> k     <- p-1  #banyaknya variabel eksplanatori
> #Menghitung Sum Square Residual di bawah H1
> yh    <- H%*%y       #Nilai prediksi di bawah H0
> yh2   <- yh*yh       #yh^2
> yh3   <- yh2*yh      #yh^3
> ##Membentuk matriks X di bawah H1
> X1    <- cbind(X,yh2,yh^3)
> q     <- ncol(X1)-ncol(X)+1  #Menghitung ordo tertinggi
> #Menghitung Sum Square Residual di bawah H1
> H1    <- X1%*%solve(t(X1)%*%X1, tol=1e-39)%*%t(X1) #Matriks Hat (H)
> SSE1  <- t(y)%*%(I-H1)%*%y
> SSE1
         [,1]
[1,] 433.4385
> #Menghitung Statistik uji
> f1    <- (SSE0-SSE1)/(q-1)/(SSE1/(n-k-q))
> f1
         [,1]
[1,] 1.629352
> #Menentukan kriteria uji
> alpha <- 0.05
> fcrit <- qf(1-alpha,q-1,n-k-q)
> fcrit
[1] 2.998588
> p_val1<- 1-pf(f1,q-1,n-k-q)
> p_val1
          [,1]
[1,] 0.1962221
> #STEPWISE BACKWARD
> # cycle 1
> ## step 1
> ### X1-X6
> X <- cbind(u,X0[,1:20])
> p <- ncol(X)
> bh <- solve(t(X)%*%X)%*%t(X)%*%y
> H <- X%*%(solve(t(X)%*%X))%*%t(X)
> e <- (I-H)%*%y
> MSE <- (t(e)%*%e)/(n-p)
> SE <- sqrt(diag((solve(t(X)%*%X)%x%MSE)))
> t <- bh/SE
> p_v <- 2*(1-pt(abs(t),n-p))
> p_v>=0.05
          [,1]
```

```
             FALSE
X3.1          TRUE
X3.3         FALSE
X5.1         FALSE
X6.1         FALSE
X6.2         FALSE
X6.3          TRUE
X7           FALSE
X8            TRUE
X10          FALSE
X11.1         TRUE
X11.2         TRUE
X11.3         TRUE
log.X12.     FALSE
log.X13.     FALSE
X14.1        FALSE
X14.2         TRUE
X14.3        FALSE
X15.1        FALSE
X16.1         TRUE
X16.3        FALSE
> p_v
                     [,1]
             0.000000e+00
X3.1         1.871709e-01
X3.3         4.043602e-06
X5.1         1.437478e-06
X6.1         3.448090e-03
X6.2         8.677440e-03
X6.3         6.609688e-01
X7           0.000000e+00
X8           7.462840e-02
X10          1.551128e-03
X11.1        2.818503e-01
X11.2        4.769012e-01
X11.3        1.808818e-01
log.X12.     2.492216e-04
log.X13.     1.607504e-08
X14.1        3.012677e-06
X14.2        8.908510e-01
X14.3        5.085710e-12
X15.1        3.078399e-02
X16.1        9.283859e-02
X16.3        1.254192e-06
> # cycle 2
> ## step 1
> ### model yang dikeluarkan adalah X4.2, X5.1, X6.1, X6.2, X7, X10, log.X12, log.X13, X14.1,
X14.3, X15.1, dan X16.1
> X <- cbind(u,X0[,c(2,3,4,5,7,9,13,14,15,17,18,20)])
> p <- ncol(X)
> bh <- solve(t(X)%*%X)%*%t(X)%*%y
> H <- X%*%(solve(t(X)%*%X))%*%t(X)
> e <- (I-H)%*%y
> MSE <- (t(e)%*%e)/(n-p)
> SE <- sqrt(diag((solve(t(X)%*%X)%x%MSE)))
> t <- bh/SE
> p_v <- 2*(1-pt(abs(t),n-p))
> p_v
                     [,1]
             0.000000e+00
X3.3         4.933299e-06
X5.1         5.811401e-07
```

```
X6.1      2.661989e-05
X6.2      4.071855e-03
X7        0.000000e+00
X10       1.921443e-03
log.X12.  3.344857e-04
log.X13.  3.858926e-09
X14.1     1.973326e-08
X14.3     4.440892e-16
X15.1     2.373847e-02
X16.3     1.429299e-08
> p_v>=0.05
           [,1]
          FALSE
X3.3      FALSE
X5.1      FALSE
X6.1      FALSE
X6.2      FALSE
X7        FALSE
X10       FALSE
log.X12.  FALSE
log.X13.  FALSE
X14.1     FALSE
X14.3     FALSE
X15.1     FALSE
X16.3     FALSE
> est <- cbind(bh,t,p_v)
> R2 <- 1-MSE/var(y)
> est
                 [,1]      [,2]         [,3]
           4.62965759 44.623371 0.000000e+00
X3.3      -0.06691313 -4.575529 4.933299e-06
X5.1       0.10212490  5.007685 5.811401e-07
X6.1       0.06055218  4.206822 2.661989e-05
X6.2       0.12932303  2.874648 4.071855e-03
X7         0.19323747 13.660214 0.000000e+00
X10        0.04921327  3.104727 1.921443e-03
log.X12.   0.15805853  3.590960 3.344857e-04
log.X13.   0.33760762  5.906830 3.858926e-09
X14.1     -0.11134342 -5.628823 1.973326e-08
X14.3     -0.15360537 -8.137792 4.440892e-16
X15.1      0.06317210  2.262432 2.373847e-02
X16.3     -0.08643872 -5.684767 1.429299e-08
> R2
           [,1]
[1,] 0.1905938
> bh
                 [,1]
           4.62965759
X3.3      -0.06691313
X5.1       0.10212490
X6.1       0.06055218
X6.2       0.12932303
X7         0.19323747
X10        0.04921327
log.X12.   0.15805853
log.X13.   0.33760762
X14.1     -0.11134342
X14.3     -0.15360537
X15.1      0.06317210
X16.3     -0.08643872
> #Heteroskedastisitas------------------------------------
> # taraf signifikansi
```

```
> alpha   <- 0.05
> # Menghitung residu^2 dari regresi utama
> ##Membentuk vektor y
> y   <- as.matrix(data[,1])
> ##Membentuk matriks X
> X0  <- as.matrix(data[,c(2,3,4,5,7,9,13,14,15,17,18,20)])   #Kolom matriks X dari data
> n   <- nrow(y)              #Menentukan banyak unit pengamata (baris dari vektor y)
> u   <- matrix(c(1),n,1)     #Membentuk vektor unit konstan
> X   <- cbind(u,X0)
> ##Menghitung Residual
> H   <- X%*%solve(t(X)%*%X)%*%t(X)   #Matriks Hat (H)
> I   <- diag(c(1),n,n)           #Membentuk matriks identitas (nxn)
> e   <- (I-H)%*%y
> e2  <- e*e
> # Uji Breucsh-Pagan
> ##Membentuk vektor y
> y   <- e2                   # regresan merupakan resiudal^2 dari reresi utama
> ##Membentuk matrix Z
> X   <- X                    # regresor dalam model tambahan sama dengan regresor dlm model
utama
> p   <- ncol(X)
> k   <- p-1              # banyaknya variabel eksplanatori
> ## Menghitung R^2
> H   <- X%*%solve(t(X)%*%X)%*%t(X)   #Matriks Hat (H)
> I   <- diag(c(1),n,n)           #Membentuk matriks identitas (nxn)
> J   <- matrix(c(1),n,n)     # unit constant martix of n-order
> SSE <- t(y)%*%(I-H)%*%y         #sum square residual
> SSR <- t(y)%*%(H-(J/n))%*%y     #sum square regression
> SST <- SSR+SSE              #sum square total
> R2  <- SSR/SST              #R-square
> ## Statistik uji
> chi2    <- n*R2
> ## Titik kritis
> chi_c   <- qchisq(1-alpha,k)
> k
[1] 12
> R2
            [,1]
[1,] 0.004373701
> chi2
         [,1]
[1,] 13.85588
> chi_c
[1] 21.02607
> #LINEARITAS---------------------
> #Membentuk vektor y
> y   <- as.matrix(data[,1])
> #Membentuk matriks X
> X0  <- as.matrix(data[,c(2,3,4,5,7,9,13,14,15,17,18,20)])  #Kolom matriks X dari data
> n   <- nrow(y)              #Menentukan banyak unit pengamata (baris dari vektor y)
> u   <- matrix(c(1),n,1)     #Membentuk vektor unit konstan
> X   <- cbind(u,X0)
> #Menghitung Sum Square Residual di bawah H0
> H   <- X%*%solve(t(X)%*%X)%*%t(X)   #Matriks Hat (H)
> I   <- diag(c(1),n,n)           #Membentuk matriks identitas (nxn)
> e   <- (I-H)%*%y
> SSE0    <- t(y)%*%(I-H)%*%y
> SSE0
         [,1]
[1,] 472.8041
> ##Menghitung banyak regresor+1
```

```
> p   <- ncol(X)
> k   <- p-1  #banyaknya variabel eksplanatori
> #Menghitung Sum Square Residual di bawah H1
> yh  <- H%*%y    #Nilai prediksi di bawah H0
> yh2 <- yh*yh    #yh^2
> yh3 <- yh2*yh   #yh^3
> ##Membentuk matriks X di bawah H1
> X1  <- cbind(X,yh2,yh^3)
> q   <- ncol(X1)-ncol(X)+1   #Menghitung ordo tertinggi
> #Menghitung Sum Square Residual di bawah H1
> H1  <- X1%*%solve(t(X1)%*%X1, tol=1e-39)%*%t(X1)    #Matriks Hat (H)
> SSE1    <- t(y)%*%(I-H1)%*%y
> SSE1
         [,1]
[1,] 472.2257
> #Menghitung Statistik uji
> f1  <- (SSE0-SSE1)/(q-1)/(SSE1/(n-k-q))
> f1
         [,1]
[1,] 1.930982
> #Menentukan kriteria uji
> alpha   <- 0.05
> fcrit   <- qf(1-alpha,q-1,n-k-q)
> fcrit
[1] 2.99858
> p_val1<- 1-pf(f1,q-1,n-k-q)
> p_val1
          [,1]
[1,] 0.1451772
> #---END-------------------------
```