ASUMSI SISAAN MENYEBAR NORMAL

Studi Kasus: Pengaruh Kualitas Udara di Kota Hyderabad Berdasarkan Indeks Kualitas Udara (AQI), Amonia, dan O_3

Makalah ini disusun untuk memenuhi tugas akhir mata kuliah Model Linear

Dosen Pengampu:

Ary Santoso, M.Si.

Disusun Oleh: Kelompok 1

Zahra Khotimah	**11210940000003**
M. Ikhwan Farhat	**11210940000007**
Reysia Amanda Yura	**11210940000008**
Dina Sekar Juliati	**11210940000020**
Intan Syafitri	**11210940000025**
Rifaldi Achmad Faisal	**11210940000027**
Chelsea Fatihah Rahma	**11210940000035**
Deffin Purnama Noer	**11210940000036**
Fanny Wahyu Aprilia	**11210940000037**

PROGRAM STUDI MATEMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SYARIF HIDAYATULLAH JAKARTA

2023

KATA PENGANTAR

Assalammu'alaikum Warahmatullahi Wabarakatuh

Alhamdulillahirabbil'alamin puji dan syukur penulis panjatkan kehadirat Allah SWT. atas segala rahmat dan hidayah-Nya penulis dapat menyelesaikan makalah dengan judul **"Asumsi Sisaan Menyebar Normal. Studi Kasus: Pengaruh Kualitas Udara di Kota Hyderabad Berdasarkan Indeks Kualitas Udara (AQI), Amonia, dan O_3"**.

Makalah ini disusun untuk memenuhi tugas akhir semester yang harus diselesaikan dalam mata kuliah Model Linear dengan dosen pengampunya adalah Bapak Ary Santoso, M.Si. Harapan kami sebagai penulis tentunya berharap agar makalah ini dapat berguna bagi para pembaca dalam menambah wawasan.

Jelas bahwa selesainya makalah ini tidak lepas dari adanya dukungan dan bantuan dari berbagai pihak. Untuk itu kami sangat berterima kasih kepada semua yang berperan penting dalam penyelesaian makalah ini, antara lain:

1. Bapak Ary Santoso, M.Si. selaku dosen pengampu mata kuliah Model Linear Program studi Matematika Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta.
2. Anggota Kelompok 1 kelas A mata kuliah Model Linear Program studi Matematika Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta.
3. Seluruh pihak yang telah membantu kami dalam menyusun makalah ini.

Penulis menyadari bahwa dalam penyusunan makalah ini masih banyak kekurangan. Oleh karena itu, penulis mengharapkan kritik dan saran yang bersifat membangun untuk memperbaiki di masa yang akan datang. Terakhir, penulis juga berharap semoga makalah ini mampu memberikan pengetahuan tentang asumsi sisaan menyebar normal dan permasalahan yang ditemukan dalam kehidupan sehari-hari.

Wassalammu'alaikum Warahmatullahi Wabarakatuh

Jakarta, 8 Juni 2023

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

BAB I
PENDAHULUAN

1.1 Latar Belakang

Analisis data merupakan salah satu proses penelitian yang dilakukan setelah semua data yang diperlukan untuk memecahkan permasalahan yang diteliti sudah diperoleh secara lengkap. Teknik analisis data terbagi menjadi 2, yaitu analisis kuantitatif dan kualitatif. Yang membedakan kedua teknik tersebut hanya terletak pada jenis datanya saja. Analisis kuantitatif yang biasa digunakan adalah analisis statistik. Salah satu metode analisis statistik adalah analisis statistik inferensial. Analisis statistik inferensial yaitu statistik yang berguna untuk mengadakan penarikan kesimpulan dan membuat keputusan berdasarkan analisis yang telah dilakukan. Analisis ini terbagi dalam beberapa jenis, salah satunya analisis regresi.

Analisis regresi digunakan untuk mengetahui sejauh mana ketergantungan atau hubungan sebuah variabel tak bebas dengan sebuah atau lebih variabel bebas. Jika dalam analisisnya hanya melibatkan sebuah variabel bebas, maka analisis yang digunakan adalah Analisis Regresi Linier Sederhana. Sedangkan bila dalam analisisnya melibatkan dua atau lebih variabel bebas, maka analisis yang digunakan adalah Analisis Linier Berganda. Analisis regresi linier berganda adalah hubungan secara linear antara dua atau lebih variabel independen (X_1, X_2, …, X_n) dengan variabel dependen (Y). Analisis ini untuk mengetahui arah hubungan antara variabel independen dengan variabel dependen apakah masing-masing variabel independen berhubungan positif atau negatif dan untuk memprediksi nilai dari variabel dependen apabila nilai variabel independen mengalami kenaikan atau penurunan.

Diperlukan uji normalitas pada sisaan, hal ini dilakukan untuk dapat menguji apakah data yang akan digunakan untuk uji hipotesis yaitu data dari variabel dependen dan independen yang digunakan telah berdistribusi secara normal atau tidak. Sisaan atau residual adalah selisih antara nilai amatan sebenarnya dan nilai perkiraan karena data sample yang kita miliki harus benar-benar mewakili populasi, sehingga hasil penelitian yang dibangun bisa digeneralisasikan pada populasi. Namun, seringkali terjadi suatu kesalahan yaitu ketika tes normalitas dilakukan untuk setiap variabel, meskipun tidak dilarang akan tetapi model regresi memerlukan suatu normalitas pada nilai residual dan bukan dalam variabel penelitian. Jika asumsi ini dilanggar maka uji statistik menjadi tidak valid, statistik parametrik tidak dapat digunakan, varians inkosisten dan model kurang mampu menaksir parameter dari populasi sebenarnya. Pelanggaran yang terjadi pada asumsi normalitas yaitu jika error tidak menyebar normal maka penduga metode kuadrat terkecil tidak berdistribusi normal dan statistik uji-t tidak menyebar secara t-student sehingga perhitungan peluangnya akan salah uji dari keputusan ujinya yang tidak tepat. Asumsi Normalitas adalah untuk melihat apakah nilai residual terdistribusi normal atau tidak. Pelanggaran dan kemungkinan penyebabnya juga bisa terjadi jika sebaran respon memang bukan sebarang normal, pelanggaran asumsi lineritas, adanya pencicilan dan ukuran sample yang terlalu kecil.

Untuk mengatasi masalah tersebut dapat dilakukan dengan Pengujian Asumsi Sisaan Menyebar Normal (Normalitas) dapat diperiksa dengan 2 cara, yaitu Pengeksplorasian dengan Plot Normal dan Histogram Sisaan (Histogram sisaan dan Plot sisaan) dan Pengujian Asumsi Klasik Normalitas (Uji Kolmogorov-Smirnov, Uji Shapiro-Wilk, Uji Lilliefors). Pada penelitian ini, peneliti ingin melihat pada tiap metode uji normalitas yaitu uji Kolmogorov-Smirnov, uji Lilliefors, dan uji Shapiro-Wilk apakah dalam menghasilkan keputusan memberikan hasil yang konsisten jika diterapkan pada berbagai besar sampel dan metode uji normalitas manakah yang menghasilkan tingkat konsistensi terbaik.

1.2 Rumusan Masalah

1. Apa itu uji asumsi sisaan menyebar normal?
2. Apa dampak atau akibat bagi model yang dihasilkan apabila terdapat pelanggaran pada uji asumsi sisaan menyebar normal?
3. Bagaimana cara memeriksa pengujian pada uji asumsi sisaan menyebar normal?
4. Bagaimana script R yang digunakan untuk melakukan pengujian pada uji asumsi sisaan menyebar normal?
5. Apa yang perlu dilakukan apabila terjadi pelanggaran pada uji asumsi sisaan menyebar normal?

1.3 Batasan Penelitian

Ruang lingkup dalam penelitian ini dibatasi pada data yang diambil dari buku atau internet, dan antar variabel bebas pada data yang memiliki asumsi menyebar normal.

1.4 Tujuan Penelitian

Tujuan dilakukannya penulisan makalah "Asumsi Sisaan Menyebar Normal. Studi Kasus: Pengaruh Kualitas Udara di Kota Hyderabad Berdasarkan Indeks Kualitas Udara (AQI), Amonia, dan O_3" adalah:

1. Mengetahui dan memahami definisi tentang asumsi sisaan menyebar normal.

2. Mengetahui dampak bagi model regresi yang dihasilkan apabila terdapat pelanggaran pada uji asumsi sisaan menyebar normal.
3. Mengetahui cara memeriksa dan melakukan pengujian terhadap uji asumsi sisaan menyebar normal.
4. Mengetahui dan memahami pengujian atau pemeriksaan dengan menggunakan R terhadap uji asumsi sisaan menyebar normal.
5. Dapat mengatasi apabila terjadi pelanggaran pada uji asumsi sisaan menyebar normal.
6. Dapat menyelesaikan studi kasus mulai dari pemeriksaan hingga cara mengatasi pelanggaran tersebut.

BAB II
METODOLOGI

2.1 Susunan Tim

Posisi	Nama	NIM
Ketua	Rifaldi Achmad Faisal	11210940000027
Anggota	Zahra Khotimah	11210940000003
	M. Ikhwan Farhat	11210940000007
	Reysia Amanda Yura	11210940000008
	Dina Sekar Juliati	11210940000020
	Intan Syafitri	11210940000025
	Chelsea Fatihah Rahma	11210940000035
	Deffin Purnama Noer	11210940000036
	Fanny Wahyu Aprilia	11210940000037

Tabel 1 Susunan Tim

2.2 Lembar Kerja Tim

No	Nama	Deskripsi Tugas
1	Zahra Khotimah	- Pengertian asumsi sisaan menyebar normal - Mengatasi jika terjadi pelanggaran pada asumsi - Bab 2 Metodologi - Studi kasus - Kesimpulan dan daftar pustaka - Membuat PPT dan Poster
2	M. Ikhwan Farhat	- Dampak (akibat) bagi model regresi yang dihasilkan jika asumsi dilanggar - Bab 2 Metodologi - Studi kasus - Kesimpulan dan daftar pustaka - Membuat PPT dan Poster

3	Reysia Amanda Yura	- Memeriksa atau melakukan pengujian pada asumsi - Membuat script R untuk melakukan pengujian atau pemeriksaan asumsi - Bab 2 Metodologi - Studi kasus - Kesimpulan dan daftar pustaka - Membuat PPT dan Poster
4	Dina Sekar Juliati	- Mengatasi jika terjadi pelanggaran pada asumsi - Penjelasan mengenai cara memeriksa atau melakukan pengujian pada asumsi - Bab 2 Metodologi - Studi kasus - Kesimpulan dan daftar pustaka - Membuat PPT dan Poster
5	Intan Syafitri	- Membuat Bab 1 Pendahuluan - Pengertian asumsi sisaan menyebar normal - Bab 2 Metodologi - Studi kasus - Kesimpulan dan daftar pustaka - Membuat PPT dan Poster
6	Rifaldi Achmad Faisal	- Memeriksa atau melakukan pengujian pada asumsi - Membuat script R untuk melakukan pengujian atau pemeriksaan asumsi - Bab 2 Metodologi - Studi kasus - Kesimpulan dan daftar pustaka - Membuat PPT dan Poster
7	Chelsea Fatihah Rahma	- Memeriksa atau melakukan pengujian pada asumsi - Bab 2 Metodologi - Studi kasus - Kesimpulan dan daftar pustaka - Membuat PPT dan Poster

8	Deffin Purnama Noer	- Dampak (akibat) bagi model regresi yang dihasilkan jika asumsi dilanggar - Membuat script R untuk melakukan pengujian atau pemeriksaan asumsi - Bab 2 Metodologi - Studi kasus - Kesimpulan dan daftar pustaka - Membuat PPT dan Poster
9	Fanny Wahyu Aprilia	- Membuat Bab 1 Pendahuluan - Membuat script R untuk melakukan pengujian atau pemeriksaan asumsi - Bab 2 Metodologi - Studi kasus - Kesimpulan dan daftar pustaka - Membuat PPT dan Poster

Tabel 2 Lembar Kerja Tim

2.3 Laporan Telemeeting

No	Tanggal	Agenda	Daftar Hadir	Foto Telemeeting
1	08/06/2023	Pembagian tugas dan deadline	Hadir semua	
2	16/06/2023	Pembahasan kerangka studi kasus	Hadir semua	

3	18/06/2023	Fiksasi studi kasus	Hadir semua	
4	19/06/2023	Pembahasan PPT	Hadir semua	
5	25/06/2023	Revisi studi kasus	Hadir Semua	

Tabel 3 Laporan Telemeeting

2.4 Tahapan Kegiatan Penyelesaian Tugas dan Rencana Waktu

No	Kegiatan	Bulan Juni											
		6	7	8	9	10	11	12	13	14	15	16	17
1	Pembuatan Grup WA												
2	Telemeeting 1 dan pembagian tugas												
3	Pencarian referensi jurnal dan data studi kasus												

4	Penyelesaian Bab I												
5	Penyelesaian Bab III												
6	Telemeeting 2 dan pembahasan studi kasus												

No	Kegiatan	Bulan Juni											
		18	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**		
7	Fiksasi data studi kasus dan pemaparan studi kasus												
8	Pembuatan PPT												
9	Poster												
10	Penyelesaian Bab IV dan makalah												

Tabel 4 Kegiatan Penyelesaian Tugas dan Rencana Waktu

BAB III
HASIL DAN PEMBAHASAN

3.1 Pengertian Asumsi Sisaan Menyebar Normal

Asumsi normalitas pada uji asumsi klasik merupakan salah satu uji pendukung dalam metode regresi berganda. Asumsi normalitas adalah sebuah uji yang dilakukan dengan tujuan untuk menilai sebaran data pada sebuah kelompok data atau variabel dependen dan independen, apakah sebaran data tersebut berdistribusi normal atau tidak. Uji normalitas berguna untuk menentukan data yang telah dikumpulkan berdistribusi normal atau diambil dari populasi normal. Model regresi yang baik adalah yang memiliki residual yang terdistribusi secara normal. Residual atau sisaan adalah selisih antara nilai amatan sebenarnya dan nilai perkiraan karena data sampel yang kita miliki harus benar-benar mewakili populasi, sehingga hasil penelitian yang dibangun bisa digeneralisasikan pada populasi.

Untuk mengetahui apakah asumsi kenormalan ini terpenuhi dapat dilakukan pengujian, ada dua cara untuk mendeteksi apakah residual berdistribusi normal atau tidak yaitu dengan analisis grafik dan uji statistik. Pengujian asumsi sisaan menyebar normal (normalitas) dapat diperiksa dengan tiga metode, antara lain: Histogram yaitu jika distribusi residual berbentuk mirip dengan kurva normal, maka asumsi normalitas terpenuhi; Plot-QQ yaitu membandingkan distribusi residual dengan distribusi teoritis. Jika titik-titik plot-QQ sejajar dengan garis referensi maka asumsi terpenuhi; dan Uji Statistik, ada beberapa contoh uji statistik, yaitu uji Kolmogorov-Smirnov (Shapiro-Wilk) atau dengan uji Lillie. Uji Kolmogorov-Smirnov bekerja dengan cara membandingkan dua buah distribusi/sebaran data, yaitu distribusi yang dihipotesiskan dan distribusi yang teramati. Sedangkan uji Lillie merupakan adaptasi dari uji Kolmogorov-Smirnov, uji ini relatif lemah dan data yang diperlukan juga cukup besar agar kita dapat menolak hipotesis dalam uji normalitas.

Penulisan matematis dari asumsi normalitas ini adalah:

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

Asumsi ε merupakan suatu peubah acak, dengan nilai tengah nol dan ragam σ^2. Dimana ~ berarti "didistribusikan sebagai" dan N berarti "distribusi normal", unsur dalam kurung menyatakan dua parameter distribusi normal, yaitu rata-rata dan variansi.

Hipotesis dalam uji normalitas adalah:

H_0: Residual data menyebar normal

H_1: Residual data tidak menyebar normal

Distribusi yang dihipotesiskan dalam kasus ini adalah distribusi normal. Sedangkan distribusi yang teramati adalah distribusi yang dimiliki oleh data yang sedang kita uji. Apabila distribusi yang teramati mirip dengan distribusi yang dihipotesiskan (distribusi normal), maka kita bisa menyimpulkan bahwa data yang kita amati memiliki distribusi/sebaran normal.

Pengujian asumsi sisaan menyebar normal (normalitas) dapat diperiksa dengan tiga metode, antara lain: Histogram yaitu jika distribusi residual berbentuk mirip dengan kurva normal, maka asumsi normalitas terpenuhi; Plot-QQ yaitu membandingkan distribusi residual dengan distribusi teoritis. Jika titik-titik plot-QQ sejajar dengan garis referensi maka asumsi terpenuhi; dan Uji Statistik, ada beberapa contoh uji statistik, yaitu uji Kolmogorov-Smirnov (Shapiro-Wilk) atau dengan uji Lillie. Uji Kolmogorov-Smirnov bekerja dengan cara membandingkan dua buah distribusi/sebaran data, yaitu distribusi yang dihipotesiskan dan distribusi yang teramati. Sedangkan uji Lillie merupakan adaptasi dari uji Kolmogorov-Smirnov, uji ini relatif lemah dan data yang diperlukan juga cukup besar agar kita dapat menolak hipotesis dalam uji normalitas.

3.2 Dampak (Akibat) bagi Model Regresi yang Dihasilkan jika Asumsi Dilanggar

Asumsi normalitas adalah sebuah uji yang dilakukan dengan tujuan untuk menilai sebaran data pada sebuah kelompok data atau variabel dependen dan independen, apakah sebaran data tersebut berdistribusi normal atau tidak. Jika asumsi normalitas dilanggar, uji statistik menjadi tidak valid, dan metode statistik parametrik tidak dapat digunakan dengan tepat.

Uji normalitas dapat diterapkan pada data dengan skala ordinal, interval, maupun rasio. Jika analisis yang dilakukan menggunakan metode parametrik, maka asumsi normalitas harus terpenuhi, yaitu data harus berasal dari distribusi normal. Jika data tidak mengikuti distribusi normal, jumlah sampel yang sedikit, atau jenis data yang digunakan bersifat nominal atau ordinal, maka metode yang tepat untuk digunakan adalah statistik non-parametrik.

Pelanggaran asumsi normalitas terjadi ketika error (galat) tidak menyebar secara normal. Hal ini menyebabkan estimasi menggunakan metode kuadrat terkecil tidak terdistribusi normal, dan statistik uji-t tidak mengikuti distribusi t-student yang diasumsikan. Pelanggaran asumsi normalitas biasanya terjadi jika terdapat outliers (pencilan) dan ukuran sampel yang terlalu kecil.

Jika asumsi normalitas dilanggar dalam model regresi, dampaknya dapat mempengaruhi kepastian dan ketepatan model. Asumsi normalitas dalam model regresi adalah bahwa residu (selisih antara nilai yang diamati dan nilai yang diprediksi oleh model) harus terdistribusi secara normal. Sehingga jika asumsi nomalitas dalam model regresi dilanggar maka yang akan terjadi adalah sebagai berikut:

1. Estimasi parameter yang tidak akurat: Jika asumsi normalitas dilanggar, estimasi parameter model regresi seperti koefisien regresi dan kesalahan standar dapat menjadi tidak akurat. Hal ini dapat menyebabkan kesalahan dalam penilaian mengenai pengaruh variabel independen terhadap variabel dependen. Estimasi parameter yang tidak akurat dapat menjadi pelanggaran asumsi normalitas karena asumsi normalitas membutuhkan bahwa data yang diamati di dalam sampel berasal dari populasi yang terdistribusi secara normal. Jika asumsi normalitas dilanggar, hasil statistik yang diperoleh dari analisis kita dapat menjadi tidak valid atau tidak dapat diandalkan. Asumsi normalitas biasanya diperlukan untuk menghasilkan estimasi parameter yang efisien dan konsisten.
2. Ketidakmampuan menguji signifikansi: Asumsi normalitas diperlukan untuk menggunakan berbagai metode statistik yang bergantung pada distribusi normal, seperti uji t-statistik. Jika asumsi normalitas dilanggar, pengujian signifikansi dapat menghasilkan hasil yang tidak valid atau tidak akurat. Ketidakmampuan untuk menguji signifikansi sebagai akibat dari pelanggaran asumsi normalitas dapat terjadi karena beberapa alasan. Asumsi normalitas adalah salah satu asumsi penting dalam banyak metode statistik parametrik, seperti uji-t, analisis varians (ANOVA), regresi linier, dan sebagainya. Asumsi ini menyatakan bahwa distribusi data yang diamati harus mengikuti distribusi normal.
3. Interval kepercayaan yang tidak akurat: Asumsi normalitas juga penting dalam membangun interval kepercayaan yang akurat. Jika asumsi normalitas dilanggar, interval kepercayaan yang dihasilkan dapat menjadi tidak akurat atau tidak valid. Penting untuk mencatat bahwa pelanggaran asumsi normalitas tidak selalu berarti bahwa interval kepercayaan akan menjadi tidak akurat secara signifikan. Beberapa metode statistik parametrik memiliki ketahanan terhadap pelanggaran asumsi normalitas yang kecil atau menggunakan metode penyesuaian lainnya. Selain itu, ada juga metode nonparametrik dan metode simulasi seperti bootstrapping yang dapat digunakan untuk mengestimasi interval kepercayaan secara lebih akurat ketika asumsi normalitas tidak terpenuhi.
4. Prediksi yang tidak akurat: Jika asumsi normalitas dilanggar, model regresi mungkin memberikan prediksi yang tidak akurat. Kesalahan prediksi dapat meningkat, dan

model mungkin tidak mampu memperhitungkan pola atau karakteristik data yang tidak sesuai dengan asumsi normalitas.

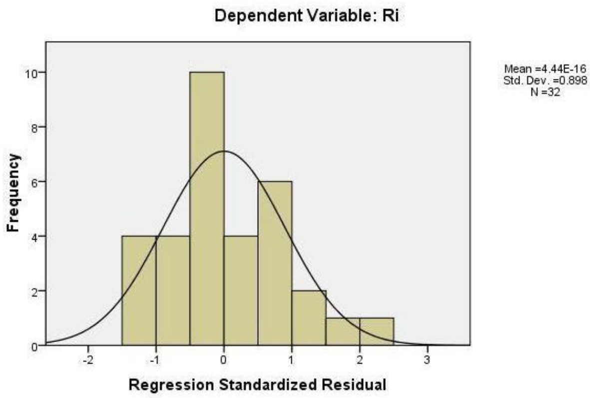
3.3 Memeriksa dan Melakukan Pengujian pada Asumsi Normalitas

Pengujian asumsi siasaan menyebar normal (Normalitas) dapat diperiksa dengan 2 cara yaitu:

3.3.1 Pengeksplorasian Dengan Plot Normal dan Histogram Sisaan

a. Histogram Sisaan

Histogram sisaan merupakan metode statistik yang digunakan untuk menguji apakah data tersebut berdistribusi normal atau tidak.



Gambar 1 Histogram

Grafik histogram dapat dikatakan normal jika distribusi data tersebut kurva-nya berbentuk lonceng yaitu tidak condong ke kiri atau tidak condong ke kanan. Grafik histogram pada gambar diatas kruva-nya kurang lebih membentuk lonceng sehingga dapat dikatakan berdistribusi normal.

Langkah-langkah untuk membuat histogram sisaan:

(a) Mengumpulkan Data

Tentukan batas interval untuk setiap interval. Misalnya, jika rentang data yang digunakan adalah 0-100 dan memiliki 10 interval, maka batas interval pertama adalah 0-10, batas interval kedua adalah 10-20, dan seterusnya.

(b) Mengurutkan Data

Urutkan data secara ascending atau descending untuk menghitung frekuensi setiap nilai dengan mudah.

(c) Menghitung Jumlah Interval

Tentukan jumlah interval yang akan digunanakan pada histogram. Jumlah interval yang disarankan adalah sekitar akar kuadrat dari jumlah data yang digunakan. Misalnya, jika Anda memiliki 100 data, gunakan sekitar 10 interval.

(d) Menghitung Lebar Interval

Hitung lebar setiap interval dengan cara membagi rentang data nilai maksimum dikurangi nilai minimum dengan jumlah interval yang telah ditentukan.

(e) Membuat Interval

Tentukan batas interval untuk setiap interval. Misalnya, jika rentang data yang digunakan adalah 0-100 dan memiliki 10 interval, maka batas interval pertama adalah 0-10, batas interval kedua adalah 10-20, dan seterusnya.

(f) Menghitung Frekuensi

Dapat menggunakan fungsi hitung atau filter pada program komputer atau spreadsheet untuk mempermudah perhitungan.

(g) Membuat Histogram

Gambarkan histogram dengan menggunakan sumbu-x untuk menampilkan interval dan sumbu-y untuk menampilkan frekuensi yang telah dibuat. Untuk setiap interval, gambar persegi panjang dengan lebar sesuai dengan lebar interval dan tinggi sesuai dengan frekuensi.

(h) Intrepetasi Hasil

Analisis histogram untuk melihat apakah data terdistribusi secara normal. Dalam distribusi normal, histogram akan menunjukkan pola simetris dengan puncak di tengah seperti bentuk lonceng. Jika histogram memiliki pola yang miring atau tidak simetris, maka data mungkin tidak memenuhi asumsi normalitas.

b. Plot Sisaan

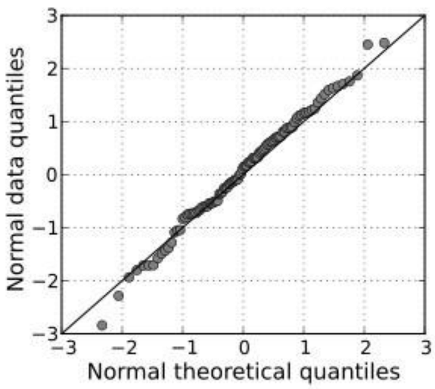
QQ plot (Quantile-Quantile plot) adalah salah satu alat yang berguna untuk mengevaluasi sejauh mana asumsi sisaan menyebar normal terpenuhi. QQ plot membandingkan distribusi sisaan dengan distribusi yang diharapkan jika sisaan terdistribusi secara normal.

Dalam QQ plot sisaan, nilai-nilai sisaan diurutkan dalam urutan naik (*sorted ascending*) dan kemudian dibandingkan dengan kuantil yang sesuai dari distribusi normal yang diharapkan. Jika asumsi sisaan menyebar normal terpenuhi, titik-titik dalam plot akan berada sepanjang garis referensi diagonal.

Interpretasi QQ plot sisaan untuk asumsi sisaan menyebar normal adalah sebagai berikut:

1. Jika titik-titik dalam QQ plot mengikuti garis referensi diagonal, ini menunjukkan bahwa sisaan secara keseluruhan terdistribusi secara normal. Semakin dekat titik-titik dengan garis referensi diagonal, semakin baik asumsi sisaan menyebar normal terpenuhi.
2. Jika titik-titik dalam QQ plot berada di atas garis referensi diagonal, ini menunjukkan bahwa ekor atas sisaan lebih tebal atau lebih ekstrem daripada yang diharapkan dalam distribusi normal. Hal ini dapat mengindikasikan kemencengan positif dalam sisaan.
3. Jika titik-titik dalam QQ plot berada di bawah garis referensi diagonal, ini menunjukkan bahwa ekor bawah sisaan lebih tebal atau lebih ekstrem daripada yang diharapkan dalam distribusi normal. Hal ini dapat mengindikasikan kemencengan negatif dalam sisaan.

Pada umumnya, semakin dekat titik-titik dalam QQ plot dengan garis referensi diagonal, semakin baik asumsi sisaan menyebar normal terpenuhi.



Gambar 2 Plot Q-Q

Langkah-langkah membuat QQ plot:

1. Mengumpulkan data, pastikan data sudah terkumpul dengan lengkap dan bersih dari gangguan atau kesalahan pengukuran.
2. Mengurutkan data, urutkan data dalam urutan dari terkecil hingga terbesar.
3. Menghitung kuantil teoritis, hitung kuantil teoritis yang sesuai untuk distribusi normal dengan peluang $\frac{nomor\ urut\ data - 0{,}5}{n}$ (Jika terdapat data yang sama maka ubah nomor urut tersebut dengan menghitung nilai rata-rata dari nomor urut tersebut).
4. Menghitung kuantil empiris, hitung kuantil empiris dari data yang dimiliki.

5. Membuat plot QQ, dengan menempatkan kuantil teoritis pada sumbu x dan kuantil empiris pada sumbu y.
6. Menginterpretasi hasil.

3.3.2 Pengujian Asumsi Klasik Normalitas

Dalam penelitian ini, akan dijelaskan lebih lanjut mengenai asumsi normalitas menggunakan uji Kolmogorov-Smirnov, uji Shapiro-Wilk, dan uji Lilliefors sebagai berikut.

a. **Asumsi Normalitas Menggunakan Uji Kolmogorov-Smirnov**

Dalam pendekatan uji Kolmogorov-Smirnov, pengujian normalitas dilakukan dengan menggunakan data residual ε. Uji Kolmogorov-Smirnov bekerja dengan cara membandingkan dua buah distribusi/sebaran data, yaitu distribusi yang dihipotesiskan dan distribusi yang teramati. Uji satu sampel Kolmogorov-Smirnov adalah suatu uji *goodness-of-fit*. Dasar pengambilan keputusan dengan menggunakan uji normalitas Kolmogorov-Smirnov adalah:

- Jika nilai signifikansi $\geq 0{,}05$ maka data berdistribusi normal
- Jika nilai signifikansi ≤ 0.05 maka data tidak berdistribusi normal

Misalkan suatu adalah fungsi distribusi frekuensi kumulatif sampel, sedangkan $F_t(x)$ adalah distribusi frekuensi kumulatif teoritis. Statistik uji Kolmogorov-Smirnov dapat dirumuskan sebagai berikut:

$$D = max|F_s(x) - F_t(x)|$$

Prosedur pengujian Kolmogorov-Smirnov ini dilakukan dengan langkah-langkah berikut antara lain:

1. Tetapkanlah fungsi kumulatif teoritisnya, yakni distribusi kumulatif yang diharapkan di bawah H_0.
2. Aturlah skor-skor yang diobservasi dalam suatu distribusi kumulatif dengan memasangkan setiap interval $F_t(x)$ dengan interval $F_s(x)$ yang sebanding.
3. Untuk tiap-tiap jenjang pada distribusi kumulatif, kurangilah $F_s(x)$ dengan $F_t(x)$.
4. Dengan memakai rumus carilah D.

$$D = \max_{1<i<N} \left(F(Y_i) - \frac{i-1}{N}, \frac{i-1}{N} - F(Y_i) \right)$$

Dimana:

D: nilai maksimum

$F(Y_i)$: peluang distribusi kumulatif

i: titik yang kurang dari Y_i

N: jumlah data

5. Lihat tabel distribusi normal untuk menemukan kemungkinan berapa probabilitas (dua sisi) kejadian untuk menentukan nilai-nilai teramati sebesar *D*, bila H_0 benar. Jika nilai-p sama atau kurang dari α, maka tolaklah H_0.

b. **Asumsi Normalitas Menggunakan Uji Shapiro Wilk**

Uji Shapiro Wilk adalah sebuah metode atau rumus perhitungan sebaran data yang dibuat oleh shapiro dan wilk. Metode ini merupakan metode uji normalitas yang efektif dan valid. Uji Shapiro Wilk digunakan untuk mengidentifikasi apakah suatu peubah acak mengikuti distribusi normal. Menurut Razali, N.M & Wah, Y.B. (2011) menyatakan bahwa uji Shapiro dan Wilk awalnya dibatasi untuk ukuran sampel yang kurang dari 50. Uji ini merupakan uji pertama yang mampu mendeteksi kenormalan data berdasarkan skewness dan kurtosis atau keduanya.

a) Syarat Uji Shapiro Wilk
   1. Data berskala interval atau ratio (kuantitatif)
   2. Data tunggal / belum dikelompokkan pada tabel distribusi frekuensi
   3. Data dari sampel random

b) Hipotesis

H_0= Data berdistribusi normal

H_1= Data tidak berdistribusi normal

c) Statistik Uji

Uji Shapiro Wilk dirumuskan sebagai berikut:

$$W = \frac{(\Sigma_{i=1}^{n} a_i\, y_i)^2}{\Sigma_{i=1}^{n} (y_i - \overline{y})^2}$$

Keterangan,

y_i = nilai residual yang diurutkan

$\overline{y}$ = rata-rata dari residual

a_i = koefisien yang bergantung pada ukuran sampel. Nilai koefisien ini telah dihitung dan tersedia dalam tabel statistik.

n =	2	3	4	5	6	7	8	9	10	11	12	13	14
a1	0.7071	0.7071	0.6872	0.6646	0.6431	0.6233	0.6052	0.5888	0.5739	0.5601	0.5475	0.5359	0.5251
a2			0.1677	0.2413	0.2806	0.3031	0.3164	0.3244	0.3291	0.3315	0.3325	0.3325	0.3318
a3					0.0875	0.1401	0.1743	0.1976	0.2141	0.2260	0.2347	0.2412	0.2460
a4							0.0561	0.0947	0.1224	0.1429	0.1586	0.1707	0.1802
a5									0.0399	0.0695	0.0922	0.1099	0.1240
a6											0.0803	0.0539	0.0727
a7													0.0240

Gambar 3 Tabel a_i jika $n \leq 14$

d) Tahapan melakukan Uji Shapiro Wilk

1. Mengurutkan data dari yang terkecil hingga terbesar
2. Hitung nilai Sum of Squares (jumlah kuadrat)

$$JK = \Sigma_{i=1}^{n}(y_i - \overline{y})^2$$

Keterangan,

y_i = nilai residual yang diurutkan

$\overline{y}$ = rata-rata dari residual

3. Menghitung nilai b

$$b = \Sigma_{i=1}^{n} a_i \, y_i$$

Keterangan,

a_i = koefisien yang bergantung pada ukuran sampel

4. Menghitung Uji Statistik

$$W = \frac{b^2}{JK}$$

5. Pengambilan keputusan

Jika $W_{statistik} < W_{\alpha(n)}$ atau $P_{value} < \alpha$, maka tolak.

Keterangan:

$W_{\alpha(n)}$ nilai dari tabel Shapiro-Wilk

α = tingkat signifikansi

c. Asumsi Normalitas Menggunakan Uji Lilliefors

Uji Lilliefors merupakan adaptasi dari uji Kolmogorov-Smirnov, uji ini relatif lemah dan data yang diperlukan juga cukup besar agar kita dapat menolak hipotesis dalam uji normalitas. Uji Lilliefors dilakukan dengan mencari nilai L_{hitung}, yakni nilai $|F(z_i) - S(z_i)|$ yang terbesar. Menurut Sudjana (1996: 466), uji normalitas data dilakukan dengan menggunakan uji Liliefors (L_0) dilakukan dengan langkah-langkah berikut:

1) Diawali dengan menentukan taraf nyata, dengan hipotesis pengujian sebagai berikut:

   H_0 : Residual berdistribusi normal

   H_0 : Residual tidak berdistribusi normal

2) Tahapan pengujian

   a) Pengamatan y_1, y_2, ……., y_n dijadikan bilangan baku z_1, z_2, ……, z_n dengan menggunakan rumus $\frac{y_i - \overline{y}}{s}$(dengan $\overline{y}$ dan s masing-masing merupakan rata-rata dari data residual dan simpangan baku sampel)

b) Untuk setiap bilangan baku ini dengan menggunakan daftar distribusi normal baku, kemudian dihitung peluang $F(z_i) = P(z < z_i)$

c) Selanjutnya dihitung proporsi z_1, z_2, ……, z_n yang lebih kecil atau sama dengan z_i. Jika proporsi ini dinyatakan oleh $S(z_i)$ maka: $\frac{banyaknya\ z_1, z_2, \ldots, z_n\ yang\ \leq z_i}{n}$

d) Hitung selisih $F(z_i) - S(z_i)$, kemudian tentukan harga mutlaknya.

e) Ambil harga yang paling besar di antara harga-harga mutlak selisih tersebut, misal harga tersebut L_0.

3) Untuk menerima atau menolak hipotesis dilakkan dengan cara membandingkan L_0 dengan tabel uji Lilliefors, dengan kriteria pengujian:

a) Jika $L_0 < L_{tabel}$ terima H_0, dan

b) Jika $L_0 > L_{tabel}$ tolak H_0.

3.4 Script R untuk Melakukan Pengujian/Pemeriksaan Asumsi

```
#Import library yang akan digunakan
library(readxl) #untuk import data excel
library(nortest) #untuk test normalitas

#Input data ke R
DL <-
read_xlsx("C:/Users/Rifaldi/Downloads/Dataset_habib
exl.xlt")

#Mengubah properti dalam variabel DL menjadi data integer
Y= as.integer(DL$AQI)
X1= as.integer(DL$`Ammonia - NH3 in æg/m3`)
X2= as.integer(DL$`O3 in æg/m3`)

#Membuat model regresi linear berganda
model = lm(Y ~ X1 + X2) #memasukkan rumus model
summary(model) #menampilkan hasil statistik model
```

Pertama, dengan melakukan import data excel pada R dengan menggunakan fungsi “readlxl()”. Setelah itu, buat suatu model regresi linier berganda menggunakan fungsi “lm()” dimana $Y = X_1 + X_2$. Untuk melihat statistika hasil model regresi linier berganda dapat dilihat melalui fungsi “summary()”.

```
# Membuat histogram
model_residual=resid(model) #membuat model regresi dari
sisaan
h=hist(model_residual) #untuk membuat histogram
xfit=seq(min(model_residual), max(model_residual), length=40)
#untuk membuat urutan variabel x
yfit=dnorm(xfit, mean=mean(model_residual),
sd=sd(model_residual)) #untuk menghitung nilai densitas
yfit=yfit*diff(h$mids[1:2]*length(model_residual)) #untuk
menyesuaikan nilai densitas dengan lebar interval histogram
lines(xfit, yfit, col='red', lwd=2) #untuk membuat garis
kurva
```

Kedua, kita lakukan eksplorasi data dengan membuat histogram untuk melihat sejauh mana distribusi residu model mendekati distribusi normal dan mengidentifikasi anomali atau pola yang mungkin terdapat dalam residu tersebut. Maka perlu membuat histogram dari residu model dan menambahkan garis kurva distribusi normal kedalamnya. Pertama, gunakan fungsi "resid()" untuk memilah sisaan/residual pada model tersebut. Selanjutnya fungsi "hist()" untuk menghasilkan histogram dari residu model, lalu gunakan fungsi "seq(min(), min(), length=n)" untuk menghasilkan urutan nilai residu dari terkecil hingga terbesar. Kemudian gunakan fungsi "dnorm(mean(), sd()" untuk menghitung nilai probabiilitas dari distribusi normal dengan menggunakan nilai residu sebagai input. Dan terakhir, fungsi "lines()" untuk menambahkan garis kurva ke dalam plot histogram.

```
# Membuat boxplot
boxplot(model_residual, main = "Boxplot",xlab = "Kategori X",
ylab = "Nilai Y")
```

Ketiga, eksplorasi data berupa boxplot untuk melihat distribusi dari residu model dan termasuk pula outlier di dalamnya dengan tujuan mengidentifikasi anomali atau pola dalam residu, maka menggunakan fungsi "boxplot()".

```
# Uji asumsi klasik normalitas untuk residual model regresi
linear berganda
ks.test(model_residual, ecdf(model_residual)) #Uji
kolmogorov-smirnov
shapiro.test(model_residual) # Uji saphiro-wilk
lillie.test(model_residual) # Uji lilliefors
```

Keempat, pengujian Asumsi Klasik Normalitas dilakukan dengan menggunakan sisaan/residual pada model regresi linier berganda yang telah kita buat. Untuk itu kita gunakan fungsi “resid()” untuk memilah sisaan/residual pada model regresi linier berganda tersebut. Pengujian dibagi menjadi 3, pertama Uji Kolmogorov-smirnov dimana menngunakan fungsi “ks.test()” dilanjutkan dengan Uji Shapiro-wilk dengan menggunakan fungsi “shapiro.test()” selanjutnya Uji Lilliefors dengan menggunakan fungsi “lillie.test()” (1).

```
# Transformasi data karena terdapat pelanggaran asumsi pada
Uji Saphiro-Wilk dan Uji Lilliefors

trans_Y = log10(Y) #untuk transformasi variabel Y
trans_X1 = log10(X1) #untuk transformasi variabel X1
trans_X2 = log10(X2) #untuk transformasi variabel X2
DL_new = cbind(trans Y, trans X1, trans X2) #Untuk
menggabungkan data yang sudah ditransformasikan (Y, X1,  X2)
model_new = lm(trans Y ~ trans X1 + trans X2) #Model baru
summary(model_new) #melihat hasil statistik model baru
model_residual_new = resid(model_new) #Model residual baru
```

Kelima, apabila dalam pengujian Asumsi Klasik Normalitas terdapat pelanggaran. Maka dapat diatasi dengan beberapa cara seperti melakukan transformasi data, menghapus data pengamatan yang memiliki nilai outliers pada data residual, menambahkan jumlah data baru, dan mengubah ke analisis regresi non-parametrik. Pada projek ini, kami akan menggunakan dua cara yakni melakukan transformasi data dan menghapus nilai outliers, yang bertujuan untuk melihat perbandingan hasil dari masing-masing cara tersebut. Pada transformasi data, langkah pertama yakni dengan mentransformasi masing-masing properti dalam variabel awal (DL) dengan menggunakan fungsi “log10()”. Selanjutnya kita membuat suatu variabel baru yang berisi gabungan dari properti yang sudah ditransformasi dengan fungsi “cbind()”. Kemudian, membuat model regresi linier berganda dengan menggunakan variabel baru dengan fungsi “lm()” dan selanjutnya mencari residual dari model baru tersebut dan melakukan pengujian Asumsi Klasik Normalitas seperti (1) Uji Asumsi Klasik di bawah ini.

```
# Uji kedua asumsi klasik normalitas dengan basis
transformasi
ks.test(model_residual_new, ecdf(model_residual_new)) # Uji
kolmogorov-smirnov
shapiro.test(model_residual_new) # Uji saphiro-wilk
lillie.test(model_residual_new) # Uji lilliefors
```

```
# Menghapus outliers
outliers <- boxplot(model_residual)$out #Memeriksa adanya
outliers
outliers

boxplot_stats <- boxplot.stats(model_residual) #menghitung
statistik bloxpot dari model residual
erase_outliers <- boxplot_stats$out #mengumpulkan nilai-nilai
outliers pada model residual
data_clean <- model_residual[!model_residual %in%
erase_outliers] #Menghapus outliers pada model residual
data_clean

boxplot(data_clean, main = "Boxplot",xlab = "Kategori X",
ylab = "Nilai Y") #Memeriksa kembali outliers pada data yang
sudah dibersihkan
new_outliers <- boxplot(data_clean)$out
new_outliers
model_clean <- lm(Y ~ X1 + X2, data = DL, subset =
!(model_residual %in% erase_outliers)) #Membuat model dengan
residual data yang sudah bersih dari outliers
summary(model_clean) #melihat hasil statistik model baru
```

Selanjutnya, kita gunakan cara kedua yakni dengan menghapus nilai outliers dari model resiual dengan menggunakan fungsi “outliers <- boxplot()$out” untuk memeriksa adanya outliers dalam model residual. Selanjutnya fungsi “boxplot_stats <- bocplot.stats()” untuk menghitung statistik boxplot dari model residual. Kemudian fungsi “erase_outliers<-boxplot_stats$out” untuk mengambil nilai-nilai outliers dari model residual pada statistik boxplot dan menyimpan dalam variabel “erase_outliers”. Selanjutnya untuk menghapus outliers kita gunakan fungsi “data_clean<-model_residual[!model_residual%in% erase_outliers]”. Setelah itu, kita buat boxplot dari data_clean dengan fungsi “boxplot()” dan terakhir, kita periksa kembali adanya outliers dalam data_clean menggunakan boxplot dengan membuat variabel baru yang berisi nilai-nilai outliers yang mungkin masih ada setelah data outliers dihapus. Kemudian membuat model regresi linier berganda dengan menggunakan variabel baru dengan fungsi “lm()”. Selanjutnya melakukan pengujian Asumsi Klasik Normalitas menggunakan data_clean seperti (1) Uji Asumsi Klasik di bawah ini.

```
# Uji kedua asumsi klasik normalitas dengan basis menghapus
outliers
ks.test(data_clean, ecdf(data_clean)) # Uji kolmogorov-
smirnov
shapiro.test(data_clean) # Uji saphiro-wilk
lillie.test(data_clean) # Uji lilliefors
```

```
# Membuat histogram 2
h=hist(model_residual_new) #membuat histogram dari model
residual baru
xfit=seq(min(model_residual_new), max(model_residual_new),
length=40) #untuk membuat variabel x
yfit=dnorm(xfit, mean=mean(model_residual_new),
sd=sd(model_residual_new)) #untuk menghitung nilai densitas
yfit=yfit*diff(h$mids[1:2]*length(model_residual_new))
lines(xfit, yfit, col='red', lwd=2) #untuk membuat garis
kurva
```

Terakhir, untuk melihat sejauh mana distribusi residu model mendekati distribusi normal dan melihat perbandingan model sebelum dan sesudah diatasi pelanggaran asumsi. Maka perlu membuat histogram dari residu model dan menambahkan garis kurva distribusi normal kedalamnya. Pertama, gunakan fungsi “resid()” untuk memilah sisaan/residual pada model tersebut. Selanjutnya fungsi “hist()” untuk menghasilkan histogram dari residu model, lalu gunakan fungsi “seq(min(), min(), length=n)” untuk menghasilkan urutan nilai residu dari terkecil hingga terbesar. Kemudian gunakan fungsi “dnorm(mean(), sd()” untuk menghitung nilai probabiilitas dari distribusi normal dengan menggunakan nilai residu sebagai input. Dan terakhir, fungsi “lines()” untuk menambahkan garis kurva ke dalam plot histogram.

3.5 Mengatasi jika Terjadi Pelanggaran pada Asumsi

Asumsi Normalitas adalah untuk melihat apakah nilai residual terdistribusi normal atau tidak. Model regresi yang baik adalah memiliki nilai residual yang terdistribusi normal. Jadi uji normalitas bukan dilakukan pada masing-masing variabel tetapi pada nilai residualnya dan model regresi memerlukan normalitas pada nilai residualnya. Jika residual tidak normal tetapi dekat dengan nilai kritis (misalnya signifikansi Kolmogorov Smirnov sebesar 0.049) maka dapat dicoba dengan metode lain yang mungkin memberikan justifikasi normal. Jika terdapat pelanggaran maka akan ada efeknya, diantaranya adalah varians inkosisten dan model kurang mampu menaksir parameter dari populasi sebenarnya. Ada beberapa cara untuk mengatasi pelanggaran pada asumsi normalitas ini, antara lain.

1. Melakukan transformasi data**.** Transformasi dapat dilakukan ke dalam bentuk logaritma natural, akar kuadrat, inverse, atau bentuk yang lain tergantung dari bentuk kurva normalnya. Transformasi ini bertujuan untuk mengurangi skewness (kemencengan) dari bentuk distribusi data sehingga membentuk distribusi yang simetris (normal). Sebelum melakukan transformasi, terlebih dahulu melihat bentuk grafik dari histogram apakah lebih condong ke kiri atau kanan untuk menentukan bentuk transformasi apa yang akan digunakan.
2. Menghapus data pengamatan yang memiliki nilai *outliers* pada data residualnya. *Outlier* adalah data yang memiliki skor ekstrem, baik ekstrem tinggi maupun ekstrem rendah. Adanya *outliers* dapat membuat distribusi skor condong ke kiri atau ke kanan. Beberapa ahli menilai data outliers ini lebih baik dibuang, karena ada kemungkinan subjek mengerjakan dengan asal-asalan, selain itu adanya data *outliers* juga mengacaukan pengujian statistik.
3. Menambah jumlah data baru (ditambah hingga hasil menunjukkan data berdistribusi normal). Usahakan jumlah sampel data lebih dari 30. Bila jumlah sampel observasi data lebih dari 30 dan ternyata data tidak normal maka biarkan saja. Gunakan pendapat *central limit theorem* yang menyatakan bahwa data yang memiliki jumlah sampel lebih

dari 30 maka dianggap normal. Karena uji normalitas pada dasarnya diperuntukkan untuk data yang memiliki sampel kecil. Untuk data dengan jumlah sampel besar dianggap normal.

4. Mengubah ke analisis regresi non-parametrik. Analisis non-parametrik tidak memerlukan asumsi normalitas seperti yang diperlukan pada analisis parametrik. Meski demikian, power test analisis non-parametrik ini tentu lebih lemah jika dibandingkan dengan analisis parametrik. Analisis regresi non-parametrik adalah cara untuk mencari tahu apakah ada hubungan antara dua variabel tanpa harus mengasumsikan bagaimana data terdistribusi atau memiliki bentuk seperti apa.

Untuk mengatasi pelanggaran terhadap setiap asumsi, dapat menggunakan Metode *Bootstrap*. Metode bootstrap pertama kali dipelajari oleh Efron (1979). Metode *bootstrap* merupakan metode penaksiran non-parametrik yang digunakan untuk mengestimasi suatu distribusi empiris yang diperoleh dari proses penyampelan ulang. Teknik penarikan sampel metode *bootstrap* adalah dengan pengembalian dari sebuah sampel asli. Istilah sampel asli digunakan untuk menyebut himpunan bagian yang pertama diambil dari populasi, sebelum dilakuukan *resampling*, yaitu proses pengambilan sampel kembali dari sampel yang telah diambil dari populasi, sedangkan istilah sampel *bootstrap* (*resample*) digunakan untuk menyebut sampel yang telah di *resampling* dari sampel asli. Dalam prakteknya resampling pada *bootstrap* dilakukan sebanyakan 3.000 atau lebih.

Adapun tujuan dari *bootstrap* salah satunya yaitu untuk mencoba dan mempelajari tentang parameter statistik dari sebuah distribusi, misalnya *mean* dan *standard error* ketika distribusi yang sesungguhnya tidak diketahui dan hanya mempunyai sekumpulan observasi. Langkah-langkah *bootstrap* dapat dinyatakan sebagai berikut:

1) Sampel data x didefinisikan sebagai data sampel yang berukuran n yang terdiri dari $x_i = x_1, x_2, \ldots, x_n$ dengan x_n sebagai vektor data pengamatan.
2) Sampel data x diambil secara acak dengan pengembalian sebanyak n kali. Data sampel baru yang didefinisikan sebagai X^*. Sampel data X^* terdiri dari anggota data asli, akan tetapi mungkin beberapa data asli tidak akan muncul, atau muncul hanya satu kali atau dua kali, tergantung dari randomisasinya.
3) Langkah (2) dilakukan secara berulang sebanyak B sehingga didapatkkan himpunan data *bootstrap* dengan ($x_1, x_2, \ldots, x_B$). Setiap sampel bootstrap merupakan sampel acak yang independen.

3.6 Studi Kasus

Studi kasus akan dilakukan dengan menggunakan data kualitas udara di Kota Hyderabad, India berdasarkan parameter Indeks Kualitas Udara (AQI), Amonia, dan O_3 pada beberapa sample udara yang bersumber dari website Kaggle.

Sumber: [*https://www.kaggle.com/datasets/manavgupta92/airqualityindexdata*](https://www.kaggle.com/datasets/manavgupta92/airqualityindexdata)

Kualitas Udara di Kota Hyderabad Berdasarkan Indeks Kualitas Udara (AQI), Amonia, dan O_3			
Id	**AQI**	**Ammonia (æg/m3)**	**O_3 (æg/m3)**
1	149	25	107.6
2	129	31	103
3	128	26	80.7
4	111	36	79.5
5		28	70
6		26	38.8
7	106.7	30	24.4
8	78.3	23	22.5
9	86	20	35.2
10	130	23	52.8
11		17	57.4
12		19	63.5
13	156	26	71.7
14	118	23	54.1
15	129	22	48.3
16	109.4	20	38.7
17	106.9	16.9	39.5
18	79.6	27.9	18
19	65.6	33	7.9
20	75	24	12.6
21	93	22	9.9

22	127	14	18
23	110.5	11	20.1
24	135.9	11	19.3
25	103.5	21	17.6
26	103.5	16	18.9
27	100.6	19	15.1
28	97.3	14	31.6
29	105.2	25	34.2
30	101.5	26	24.9
31	94	25	18
32	99	34	21.4
33	92.4	27	21.6
34	90.6	22	29.7
35	97.1	22	41.7
36	105	20	33.5
37	123	23	31.5
38	120.7	21	26
39	120	22	23.5
40		23	26.4
41	115	32	24
42	123	37	12
43	119	23	7.8
44	108	28	6.9
45	78	32	25.9
46	84	33	45.2
47	101	17	47.2
48	104	17	36.6
49	102	22	21.7
50	93	22	10.2

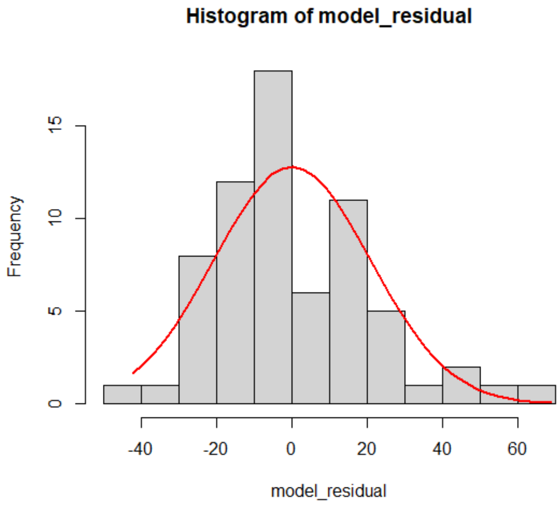
51	103	24	2.4
52	107	29	8.7
53	81	21	6.1
54	85	20	7.8
55	83	20	8.7
56	82	25	7.4
57	85	20	4.7
58	92	23	15.4
59	96	21	15.5
60	90	23	13.6
61	146	16	14.6
62	119	26	18.6
63	137	35	12.6
64	118	31	22.6
65	123	29	5.3
66	92	31	11.1
67	58	26	15.7
68	67	26	15.2
69	76	31	10
70	101	28	9.7
71	166	30	12.1
72	162	15	16.5

Tabel 5 Studi Kasus

3.6.1 Langkah Uji Asumsi Sisaan Menyebar Normal

1. Melakukan Eksplorasi Data dengan Histogram

Output

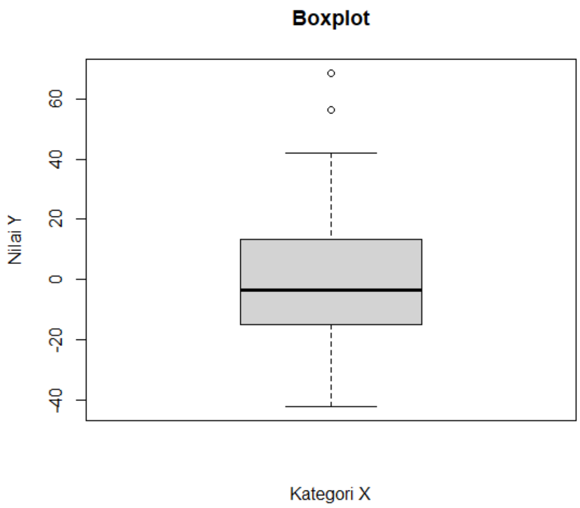


Gambar 4 Histogram Output Model Residual

Hasil diagnosa dari histogram di atas: sebaran sisaan agak menjulur ke kanan yang berarti memiliki skewness positif, dan dapat diperhatikan garis fit berwarna merah yang menyerupai histogram yang berdistribusi normal.

2. Melakukan Eksplorasi Data dengan Boxplot

Output



Gambar 5 Boxplot Outliers

Hasil diagnosa dari boxplot di atas: dapat diperhatikan terdapat beberapa outliers di atas nilai max pada boxplot, outliers tersebut mengindikasikan adanya pelanggaran pada asumsi normalitas.

3. Membuat Model Regresi Linear Berganda

Output

```
> # Import data excel menggunakan library 'readlxl'
> library(readxl)
> library(nortest)
> DL <- read_excel("C:/Users/deffi/Downloads/Dataset_habib exl.xlt")
>
> Y= as.integer(DL$AQI)
> X1= as.integer(DL$`Ammonia - NH3  in æg/m3`)
> X2= as.integer(DL$`O3   in æg/m3`)
>
> # Membuat model regresi linear berganda
> model=lm(Y~X1+X2)
> summary(model)

Call:
lm(formula = Y ~ X1 + X2)

Residuals:
    Min      1Q  Median      3Q     Max 
-42.322 -14.900  -3.408  13.249  68.649 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 105.7407    11.1496   9.484 8.16e-14 ***
X1           -0.4407     0.4420  -0.997  0.32249    
X2            0.4027     0.1191   3.383  0.00123 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.23 on 64 degrees of freedom
  (5 observations deleted due to missingness)
Multiple R-squared:  0.1578,	Adjusted R-squared:  0.1315 
F-statistic: 5.996 on 2 and 64 DF,  p-value: 0.004103
```

Dapat dilihat dari informasi model regresi yang telah dibuat, terdapat nilai dugaan $\boldsymbol{b_0} = \mathbf{105.7407}$, $\boldsymbol{b_1} = \mathbf{-0.4407}$, $\boldsymbol{b_2} = \mathbf{0.4027}$. Sehingga kita mendapatkan bentuk persamaan regresi berganda yaitu $\mathbf{Y = 105.7407 - 0.4407X_1 + 0.4027X_2}$.

4. Melakukan Uji Asumsi Klasik

Jenis Uji	Hipotesis	Output	Kesimpulan

Uji *Kolmogorov-Smirnov*	H_0: Residual berdistribusi normal
	H_1: Residual tidak berdistribusi normal
		Output:
		`Exact one-sample Kolmogorov-Smirnov test`
		`data: model_residual_new`
		`D = 0.014925, p-value = 1`
		`alternative hypothesis: two-sided`
			Dengan melakukan Uji Kolmogorov-Smirnov, kita melihat bahwa terdapat nilai $\boldsymbol{p-value} > 0.05$, menandakan bahwa data yang kita miliki berdistribusi normal.
			
			Karena $\boldsymbol{p-value} > 0.05$ maka H_0 diterima.

Uji *Shapiro-Wilk*	H_0: Residual berdistribusi normal
	H_1: Residual tidak berdistribusi normal
		Output:
		`Shapiro-Wilk normality test`
		`data: model_residual`
		`W = 0.95497, p-value = 0.01641`
			Dengan melakukan Uji *Shapiro-Wilk*, kita melihat bahwa terdapat nilai $\boldsymbol{p-value} < 0.05$, menandakan bahwa data yang kita miliki tidak berdistribusi normal.
			
			Karena $\boldsymbol{p-value} < 0.05$ maka H_0 ditolak sehingga terjadi pelanggaran asumsi.

Uji *Lillie Fors*	H_0: Residual berdistribusi normal
	H_1: Residual tidak berdistribusi normal
		Output:
		`Lilliefors (Kolmogorov-Smirnov) normality test`
		`data:  model_residual`
		`D = 0.11479, p-value = 0.02864`
			Dengan melakukan Uji *Lillie Fors*, kita melihat bahwa terdapat nilai $\boldsymbol{p-value} < 0.05$, menandakan bahwa data yang kita miliki tidak berdistribusi normal.
			
			Karena $\boldsymbol{p-value} < 0.05$ maka H_0 ditolak sehingga terjadi pelanggaran asumsi.

Tabel 6 Uji Asumsi Sisaan Menyebar Normal

3.6.2 Mengatasi Pelanggaran Asumsi Sisaan Menyebar Normal

1. Dengan Transformasi

Output

```
> # Transformasi data karena terdapat pelanggaran asumsi pada uji Chi-square
> Y_trans=log10(Y)
> X1_trans=log10(X4)
> X2_trans=log10(X5)
>
> DL_new=cbind(Y_trans,X1_trans,X2_trans) # Data baru
> model_new=lm(Y_trans~X1_trans+X2_trans) # Model baru
> model_new

Call:
lm(formula = Y_trans ~ X1_trans + X2_trans)

Coefficients:
(Intercept)     X1_trans     X2_trans
    2.04175     -0.10453      0.08948

> summary(model_new)

Call:
lm(formula = Y_trans ~ X1_trans + X2_trans)

Residuals:
     Min       1Q   Median       3Q      Max
-0.23566 -0.05470 -0.01148  0.05920  0.23619

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.04175    0.13864  14.727  < 2e-16 ***
X1_trans    -0.10453    0.09549  -1.095  0.27777
X2_trans     0.08948    0.03190   2.805  0.00665 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.08908 on 64 degrees of freedom
  (5 observations deleted due to missingness)
Multiple R-squared:  0.127,	Adjusted R-squared:  0.09969
F-statistic: 4.654 on 2 and 64 DF,  p-value: 0.01297

> model_residual_new=resid(model_new) # Model residual baru
```

Dapat dilihat dari informasi model regresi yang telah dibuat dari hasil transformasi residual data ke bentuk log10, terdapat nilai dugaan ***b₀_trans*** **= 2.04175,** ***b₁_trans*** **= -0.10453,** ***b₂*_trans = 0.08948**. Sehingga kita mendapatkan bentuk persamaan regresi berganda yaitu **Y_trans = 2.04175 - 0.10453X₁_trans+ 0.08948X₂_trans**.

2. Dengan Menghapus Outliers

Output

```
> # Menghapus outliers
> outliers <- boxplot(model_residual)$out # Memeriksa adanya outliers
> outliers
      71       72 
68.64876 56.42667 
> boxplot_stats <- boxplot.stats(model_residual)
> erase_outliers <- boxplot_stats$out
> data_clean <- model_residual[!model_residual %in% erase_outliers] # Menghapus
outliers pada model residual
> data_clean
          1           2           3           4           7           8 
 11.1848834  -4.5597058   1.4995674 -10.6902665   3.8158979 -26.4638242 
          9          10          13          14          15          16 
-25.0216538  13.4540189  33.1242144   0.6485417  13.6242304  -3.2298695 
         17          18          19          20          21          22 
 -8.3955791 -22.0898989 -29.0153182 -24.9956958  -6.6689657  20.1804452 
         23          24          25          26          27          28 
  1.0527398  26.4554784  -0.3316169  -2.9380693  -3.4076253 -15.0551561 
         29          30          31          32          33          34 
 -3.4152015  -2.9470731  -7.9713845  -0.2129153 -10.2981146 -17.7237369 
         35          36          37          38          39          41 
-15.5565997  -5.2161767  14.9115287  13.0437360  14.6926945  13.6973835 
         42          43          44          45          46          47 
 28.7339600  20.5772542  12.1837066 -23.7053551 -25.3193836 -16.1767449 
         48          49          50          51          52          53 
 -8.7466207  -2.5018284  -7.0717042   7.0316898  10.8189722 -17.9014927 
         54          55          56          57          58          59 
-14.7449740 -17.1477126 -15.5412603 -13.5367584  -9.6446543  -6.5261398 
         60          61          62          63          64          65 
-10.8391772  41.6728850  17.4693583  41.8524745  17.0621178  28.0271879 
         66          67          68          69          70 
 -4.5077580 -42.3224260 -33.3224260 -20.1050194   3.9754909 
```

Dapat dilihat bahwa di dalam data terdapat outliers yaitu pada baris 71 dan 72 sebesar 68.64876 dan 56.42667. Karena adanya outliers mengindikasikan adanya pelanggaran pada uji asumsi normalitas, maka outliers tersebut dihapus sehingga pada baris 71 dan 72 dihilangkan.

```
> boxplot(data_clean, main = "Boxplot",xlab = "Kategori X", ylab = "Nilai Y") # Memeriksa kembali outliers pada data yang sudah dibersihkan
> new_outliers <- boxplot(data_clean)$out
> new_outliers
numeric(0)
> model_clean <- lm(Y ~ X1 + X2, data = DL, subset = !(model_residual %in% erase_outliers))
> summary(model_clean)

Call:
lm(formula = Y ~ X1 + X2, data = DL, subset = !(model_residual %in%
    erase_outliers))

Residuals:
    Min      1Q  Median      3Q     Max
-34.364 -15.099  -3.825  12.497  67.317

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 105.5274    11.0025   9.591  7.3e-14 ***
X1           -0.3819     0.4400  -0.868  0.38873
X2            0.3845     0.1178   3.265  0.00179 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 20.86 on 62 degrees of freedom
  (5 observations deleted due to missingness)
Multiple R-squared:  0.1505,    Adjusted R-squared:  0.1231
F-statistic: 5.491 on 2 and 62 DF,  p-value: 0.006373
```

Dapat dilihat dari informasi model regresi dari residual data yang telah dihapus outliersnya, terdapat nilai dugaan ***b_0_new*** **= 105.5274,** ***b_1_new*** **= -0.3819,** ***b_2*****_new = 0.3845**. Sehingga kita mendapatkan bentuk persamaan regresi berganda yaitu **Y = 105.5274 - 0.3819 X_1_new + 0.3845X_2_new**.

3.6.3 Pengujian Ulang Semua Uji Asumsi Klasik

1. Dengan Transformasi

Output

```
> # Uji kedua asumsi klasik normalitas dengan basis transformasi
> ks.test(model_residual_new, ecdf(model_residual_new)) # Uji kolmogorov-smirnov

        Exact one-sample Kolmogorov-Smirnov test

data:  model_residual_new
D = 0.014925, p-value = 1
alternative hypothesis: two-sided

> shapiro.test(model_residual_new) # Uji saphiro-wilk

        Shapiro-Wilk normality test

data:  model_residual_new
W = 0.99069, p-value = 0.9007

> lillie.test(model_residual_new) # Uji lilliefors

        Lilliefors (Kolmogorov-Smirnov) normality test

data:  model_residual_new
D = 0.10068, p-value = 0.08904
```

Kita lakukan uji asumsi sekali lagi dengan residual data yang sudah ditransformasi dan dapat kita lihat bahwa untuk Uji Kolmogorov-Smirnov memiliki $\boldsymbol{p-value}$ yang sama dengan sebelum ditransformasi yaitu sebesar 1 dan termasuk di dalamnya Uji Saprhiro-Wilk serta Uji Lilliefors kita mendapatkan nilai $\boldsymbol{p-value} > \mathbf{0.05}$ maka H_0 diterima (data berdistribusi normal).

2. Dengan Menghapus Outliers

Output

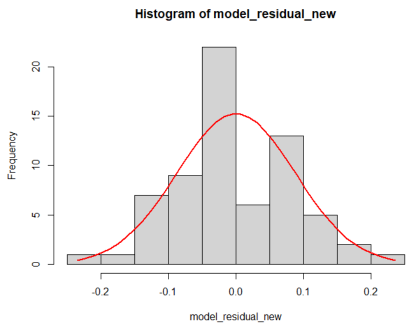
```
> # Uji kedua asumsi klasik normalitas dengan basis menghapus outliers
> ks.test(data_clean, ecdf(data_clean)) # uji kolmogorov-smirnov

        Exact one-sample Kolmogorov-Smirnov test

data:  data_clean
D = 0.015385, p-value = 1
alternative hypothesis: two-sided

> shapiro.test(data_clean) # uji saphiro-wilk

        Shapiro-Wilk normality test

data:  data_clean
W = 0.98171, p-value = 0.4497

> lillie.test(data_clean) # uji lilliefors

        Lilliefors (Kolmogorov-Smirnov) normality test

data:  data_clean
D = 0.097421, p-value = 0.1313
```

Kemudian kita lakukan uji asumsi sekali lagi dengan residual data yang sudah bersih dari outlier dan dapat kita lihat bahwa untuk Uji Kolmogorov-Smirnov memiliki $\boldsymbol{p-value}$ yang sama dengan saat sebelum dibersihkan outliernya yaitu sebesar 1 dan termasuk di dalamnya Uji Saprhiro-Wilk serta Uji Lilliefors kita mendapatkan nilai $\boldsymbol{p-value} > \mathbf{0.05}$ maka H_0 diterima (data berdistribusi normal).

3.6.4 Membuat Histogram Setelah Penanganan Pelanggaran

1. Dengan Transformasi

Output

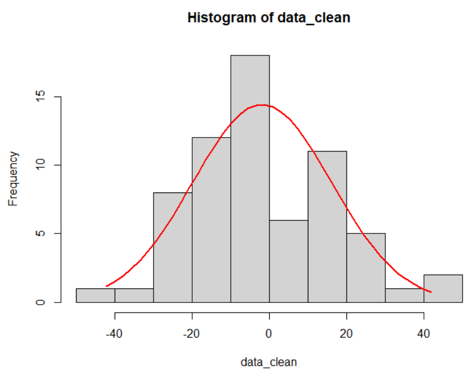


Gambar 6 Histogram Output Model Residual Baru

Hasil diagnosa dari histogram di atas: residual data yang semula memiliki skewneess sisi kanan histogram maka setelah ditransformasi log, data terdistribusi lebih merata dan terkonsentrasi di sekitar nilai-nilai tengah. Dapat diperhatikan garis fit berwarna merah yang menyerupai histogram yang berdistribusi normal.

2. Dengan Menghapus Outliers

Output



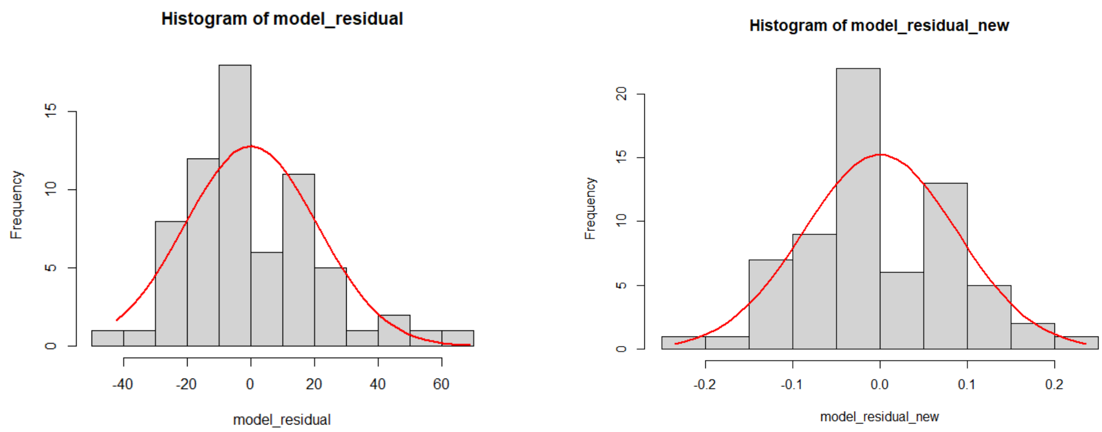
Gambar 7 Histogram Output Data Clean

Hasil diagnosa dari histogram di atas: residual data yang semula memiliki skewness sisi kanan histogram maka setelah beberapa outlier dihapus, data terdistribusi lebih merata dan terkonsentrasi di sekitar nilai-nilai tengah. Dapat diperhatikan garis fit berwarna merah yang menyerupai histogram yang berdistribusi normal.

3.6.5 Membandingkan Histogram Sebelum dan Setelah Pelanggaran Diatasi

1. Dengan Transformasi

Sebelum Diatasi: **Setelah Diatasi:**

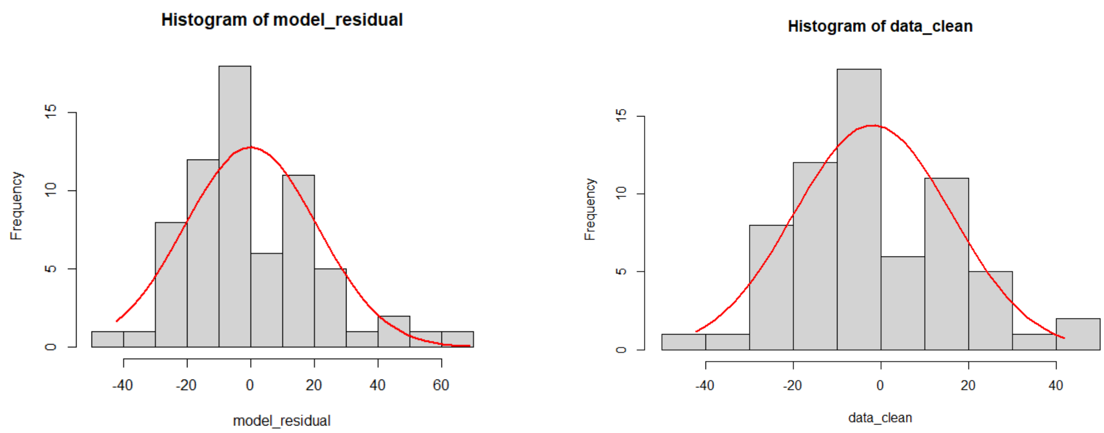


Gambar 8 Histogram Sebelum dan Setelah Transformasi

Berdasarkan pengamatan kedua histogram diatas dapat dilihat bahwa setelah melakukan transformasi data, histogram menjadi lebih simetris dan tidak memiliki skewness seperti histogram sebelum di transformasi. Selain itu, terdapat garis fit berwarna merah yang menyerupai histogram yang berdistribusi normal.

2. Dengan Menghapus Otliers

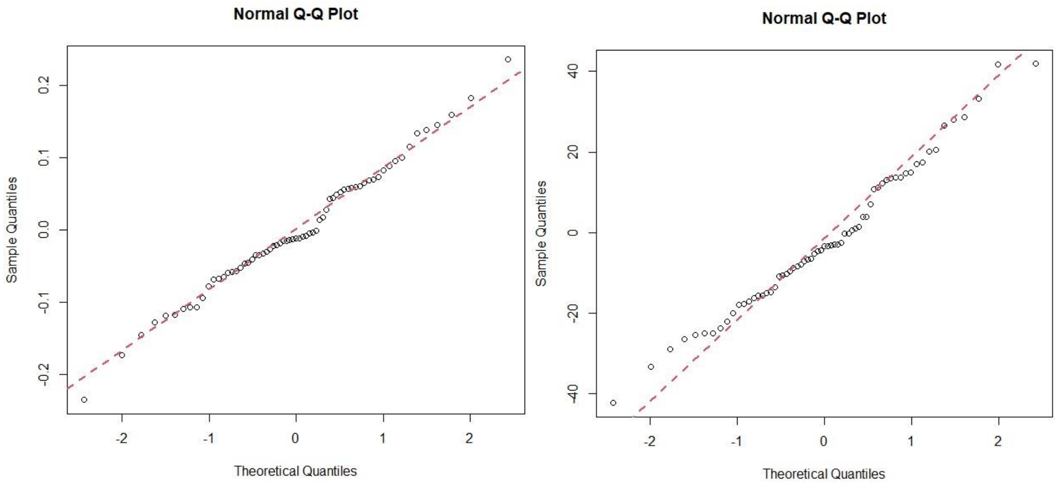
Sebelum Diatasi: **Setelah Diatasi:**



Gambar 9 Histogram Sebelum dan Setelah Menghapus Outliers

Berdasarkan pengamatan kedua histogram diatas dapat dilihat bahwa setelah menghapus outliers, histogram menjadi lebih simetris dan tidak memiliki skewness seperti histogram sebelumnya. Selain itu, terdapat garis fit berwarna merah yang menyerupai histogram yang berdistribusi normal.

3.6.6 Perbandingan Menggunakan QQ Plot



Gambar 10 Perbandingan Metode Transformasi dan Menghapus Outlier dengan Menggunakan QQ Plot

Berdasarkan pengamatan kedua QQ Plot residual data terlihat bahwa sebaran titik dari hasil transformasi cenderung tersebar ke tengah mengikuti garis referensi diagonal dibandingkan dengan hasil dari menghapus outliers. Hal itu menandakan bahwa metode penanganan yang lebih baik adalah dengan menggunakan metode transformasi daripada metode menghapus outliers.

BAB IV
PENUTUP

4.1 Kesimpulan

Berdasarkan penggunaan 3 uji asumsi klasik, dapat disimpulkan bahwa hanya uji Kolmogorov-Smirnov yang menunjukkan nilai p-value melebihi angka 0.05, yang mengindikasikan bahwa residual data tersebut memiliki distribusi normal. Sedangkan untuk uji Shapiro-Wilk dan uji Lilliefors menunjukkan nilai p-value lebih kecil dari 0.05 yang berarti residual data tersebut tidak berdistribusi normal. Oleh karena itu, dilakukan dua metode dalam penanganan pelanggaran tersebut dengan transformasi data dan penghapusan outliers hingga pada uji Shapiro-Wilk dan uji Lilliefors menunjukkan nilai p-value lebih besar dari 0.05 yang artinya residual data sudah berdistribusi normal. Setelah pelanggaran pada residual data ditangani maka metode penanganan yang lebih baik adalah dengan menggunakan metode transformasi ketimbang metode menghapus outliers dilihat dari titik-titik pada QQ plot.

DAFTAR PUSTAKA

Akhtar, Hanif. (2017). Cara Mengatasi Data Berdistribusi Tidak Normal. Semesta Psikometrika. Diakses dari https://www.semestapsikometrika.com/2017/12/mengatasi-data-tidak-normal.html?m=1 tanggal 14 Juni 2023.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2004). Applied Linear Statistical Models. McGraw-Hill Education.

Muzaki, Lubis. (2021). Cara Membuat Histogram dan Keunggulannya. Diakses dari https://www.pengadaanbarang.co.id/2021/03/histogram-adalah.html?m=1 tanggal 26 Juni 2023.

Prasetyo, R. A. & Helma. (2021). Analisis Regresi Linear Berganda Untuk Melihat Faktor Yang Berpengaruh Terhadap Kemiskinan di Provinsi Sumatera Barat. *Journal Of Mathematics UNP*, 20(1), 1-7. http://ejournal.unp.ac.id/students/index.php/mat.

Quraisy, Andi. (2020). Normalitas Data Menggunakan Uji Kolmogorov-Smirnov dan Saphiro-Wilk. *J-HEST: Journal of Health, Education, Economics, Science, and Technology*, 3(1), 7-11.

Rachman, F. P. A. P., Goejantoro, R. & Hayati, M. N. (2018). Penentuan Jumlah Replikasi Bootstrap Menggunakan Metode Pretest Pada Independent Sampel T Test. *Jurnal EKSPONENSIAL*, 9(1), 35-40.

Razali, N.M & Wah, Y.B. (2011). Power Comparisons Saphiro Wilk, Kolmogorov – Smirnov, Lilliefors and Anderson Darling Test. *Journal of Statistical Modeling and Analytics*, 2(1), 21 -33.

Sari, A. Q., Sukestiyarno, Y. L. & Agoestanto, A. (2017). Batasan Prasayarat Uji Normalitas dan Uji Homogenitas Pada Model Regresi Linear. *UNNES Journal of Mathematics*, 6(2), 168-177. http://journal.unnes.ac.id/sju/index.php/ujm.

Statistikian. (2013). Pengertian dan Rumus Uji Saphiro Wilk – Cara Hitung. Diakses 2023. https://www.statistikian.com/2013/01/ saphiro-wilk.html.