

# Praktikum Analisis Regresi Linier Sederhana

## Tujuan Praktikum

1. Memahami konsep dasar regresi sederhana, yaitu hubungan linear antara satu variabel independen (X) dan satu variabel dependen (Y).
2. Menganalisis pengaruh variabel bebas terhadap variabel terikat menggunakan model regresi linear sederhana.
3. Mengestimasi parameter model (koefisien regresi) dan menafsirkan maknanya secara statistik dan ekonometrik.
4. Menggunakan perangkat lunak R untuk melakukan estimasi model regresi sederhana serta menampilkan hasil outputnya.
5. Menguji signifikansi model dan koefisien regresi melalui uji t untuk mengetahui apakah variabel independen berpengaruh signifikan terhadap variabel dependen.

## Pendahuluan

Analisis regresi linier sederhana merupakan salah satu metode statistik inferensia yang digunakan untuk mempelajari dan memodelkan hubungan linier antara dua variabel, yaitu satu variabel independen (X) dan satu variabel dependen (Y). Tujuan utama dari metode ini adalah untuk menjelaskan arah dan kekuatan hubungan, memprediksi nilai variabel dependen berdasarkan variabel independen, serta mengukur besar pengaruh perubahan X terhadap Y dalam konteks linier. Secara konseptual, regresi linier sederhana mengasumsikan bahwa terdapat hubungan linier atau proporsional antara X dan Y. Artinya, setiap perubahan satu satuan pada X akan menyebabkan perubahan tertentu yang konstan pada nilai Y. Hubungan ini dinyatakan dalam bentuk model sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Keterangan:

- $Y_i$ : variabel dependen (nilai yang dijelaskan atau diprediksi)
- $X_i$ : variabel independen (faktor penjelas atau prediktor)
- $\beta_0$ : konstanta (intercept), nilai Y ketika X = 0
- $\beta_1$ : koefisien regresi, menunjukkan besar dan arah pengaruh X terhadap Y
- $\varepsilon_i$ : error atau residu, yaitu selisih antara nilai aktual dan nilai yang diprediksi oleh model

## Pendekatan Ordinary Least Squares (OLS)

Secara umum, analisis regresi linier sederhana menggunakan pendekatan metode kuadrat terkecil (Ordinary Least Squares / OLS). Metode ini bertujuan untuk meminimalkan jumlah kuadrat selisih antara nilai aktual (Y) dan nilai yang diprediksi oleh model ( $\hat{Y}$ ). Prinsip dasar metode OLS dikembangkan oleh Carl Friedrich Gauss dan Adrien-Marie Legendre, dan merupakan fondasi dari analisis ekonometrika modern. Fungsi objektif OLS dapat dituliskan sebagai:

$$\min_{\beta_0, \beta_1} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

Dengan meminimalkan fungsi tersebut, diperoleh estimasi parameter  $\hat{\beta}_0$  dan  $\hat{\beta}_1$  yang menghasilkan garis regresi terbaik (*best-fitting line*) untuk data yang diamati. Nilai estimasi parameter dapat dihitung dengan rumus:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

### Makna Regresi Linier Sederhana

Regresi linier sederhana tidak hanya digunakan untuk memprediksi nilai variabel dependen, tetapi juga untuk **memahami arah dan kekuatan pengaruh antarvariabel**. Nilai  $\beta_1$  memiliki interpretasi penting:

- Jika  $\beta_1 > 0$ , maka hubungan antara X dan Y bersifat **positif**, artinya peningkatan X menyebabkan peningkatan Y.
- Jika  $\beta_1 < 0$ , maka hubungan bersifat **negatif**, artinya peningkatan X menyebabkan penurunan Y.

Dalam praktik akademik, model ini banyak diterapkan dalam berbagai bidang:

- **Ekonomi:** untuk mengkaji pengaruh pendapatan terhadap konsumsi.
- **Sosial:** untuk menganalisis hubungan antara pendidikan dan produktivitas.
- **Teknik dan sains:** untuk memodelkan hubungan antara waktu dan laju reaksi, suhu dan tekanan, atau variabel pengamatan lainnya.

### Estimasi parameter menggunakan manual dan menggunakan fungsi lm

Dataset **mtcars** merupakan dataset bawaan R yang sangat populer dan sering digunakan dalam pembelajaran statistik. Dataset ini berasal dari publikasi majalah *Motor Trend* (1974) dan berisi data teknis 32 jenis mobil yang diproduksi pada tahun 1970-an.

Tujuan utama penggunaan dataset ini dalam analisis adalah untuk **mempelajari hubungan antara karakteristik teknis mobil**, seperti efisiensi bahan bakar, berat kendaraan, tenaga mesin, dan sistem transmisi. Tabel dibawah ini menjelaskan setiap variabel dalam dataset **mtcars**.

Variabel	Keterangan	Tipe Data	Satuan
<b>mpg</b>	Miles per gallon, efisiensi bahan bakar	Numerik	mil per galon
<b>cyl</b>	Jumlah silinder mesin	Numerik (diskret)	-
<b>disp</b>	Displacement atau kapasitas mesin	Numerik	cubic inches

Variabel	Keterangan	Tipe Data	Satuan
<b>hp</b>	Horsepower (tenaga mesin)	Numerik	HP
<b>drat</b>	Rear axle ratio (rasio poros belakang)	Numerik	-
<b>wt</b>	Berat mobil	Numerik	1000 lbs
<b>qsec</b>	Waktu tempuh 1/4 mil	Numerik	detik
<b>vs</b>	Tipe mesin (0 = V-shaped, 1 = straight)	Kategorik (biner)	-
<b>am</b>	Jenis transmisi (0 = otomatis, 1 = manual)	Kategorik (biner)	-
<b>gear</b>	Jumlah gigi transmisi	Numerik (diskret)	-
<b>carb</b>	Jumlah karburator	Numerik (diskret)	-

## **Eksplorasi Data**

Sebelum melakukan analisis regresi linier sederhana, terlebih dahulu dilakukan tahap eksplorasi data untuk memperoleh pemahaman awal mengenai karakteristik, pola hubungan, serta distribusi variabel yang akan dianalisis.

```
# Memanggil dataset
data(mtcars)
```

Kode diatas digunakan untuk memanggil dataset bawaan R bernama mtcars agar dapat digunakan dalam sesi analisis. Fungsi data() memastikan bahwa dataset tersebut dimuat ke dalam memori kerja (workspace) R.

```
# Menampilkan 6 observasi pertama
head(mtcars)

##          mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4     21.0   6 160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag 21.0   6 160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710    22.8   4 108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive 21.4   6 258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8 360 175 3.15 3.440 17.02  0  0    3    2
## Valiant       18.1   6 225 105 2.76 3.460 20.22  1  0    3    1
```

Fungsi head() digunakan untuk menampilkan beberapa baris awal dari dataset, secara default sebanyak 6 baris pertama. Tujuannya adalah untuk memberikan gambaran umum struktur data, seperti nama kolom, tipe data, dan nilai awal dari setiap variabel.

```
# Memilih variabel yang diperlukan
df <- mtcars[, c("mpg", "wt")]
summary(df)

##          mpg             wt
##  Min.   :10.40   Min.   :1.513
##  1st Qu.:15.43  1st Qu.:2.581
##  Median :19.20  Median :3.325
##  Mean   :20.09  Mean   :3.217
```

```
## 3rd Qu.:22.80 3rd Qu.:3.610
## Max. :33.90 Max. :5.424
```

Kode diatas melakukan dua hal:

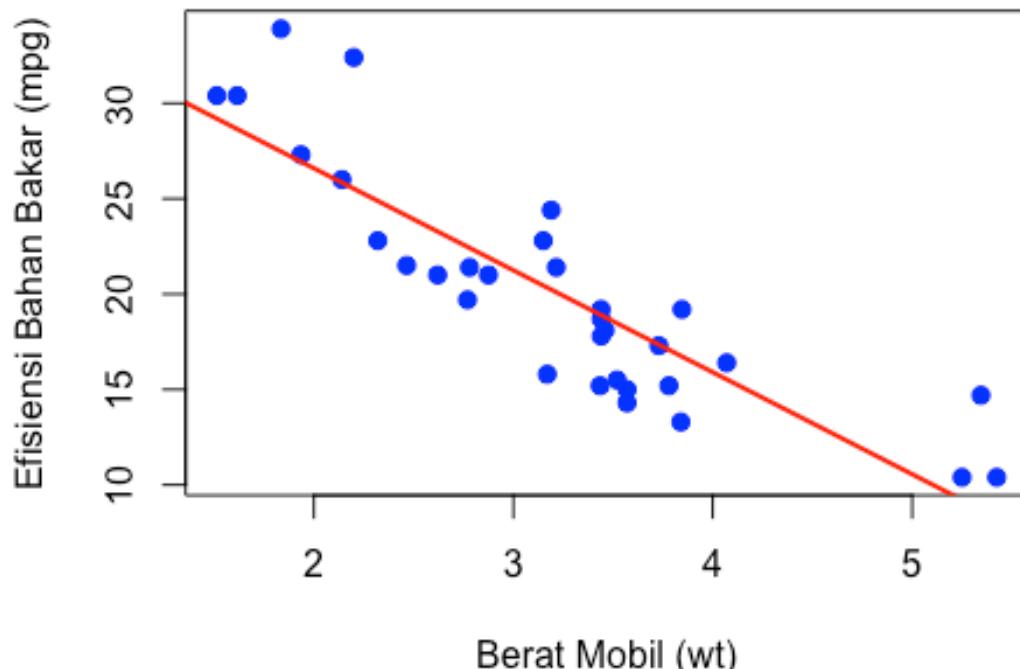
`df <- mtcars[, c("mpg", "wt")]` berfungsi membuat subset data baru bernama `df` yang hanya berisi dua variabel yaitu `mpg` (Miles per gallon) sebagai variabel dependen (Y) dan `wt`(berat mobil) sebagai variabel independen (X). Subset ini dibuat untuk memfokuskan analisis pada hubungan antara berat kendaraan dan efisiensi bahan bakar.

`Summary(df)` berfungsi memberikan statistik deskriptif untuk kedua variabel, termasuk nilai minimum, maksimum, kuartil, median, dan rata-rata. Statistik ini membantu memahami distribusi data dan potensi outlier sebelum melakukan analisis regresi.

```
# Scatter plot mpg vs wt

plot(mtcars$wt, mtcars$mpg,
xlab = "Berat Mobil (wt)",
ylab = "Efisiensi Bahan Bakar (mpg)",
main = "Hubungan antara Berat Mobil dan Efisiensi Bahan Bakar",
col = "blue", pch = 19)
abline(lm(mpg ~ wt, data = mtcars), col = "red", lwd = 2)
```

## Hubungan antara Berat Mobil dan Efisiensi Bahan Bakar



Berdasarkan kode diatas `plot(mtcars$wt, mtcars$mpg)` diatas berfungsi dalam membuat diagram pencar (scatter plot) untuk menggambarkan hubungan antara dua variabel numerik dengan

- Sumbu-X → wt (berat mobil)
- Sumbu-Y → mpg (efisiensi bahan bakar)
- xlab, ylab → memberikan label sumbu agar lebih informatif.
- main → memberikan judul grafik.
- col = "blue" → menentukan warna titik data.
- pch = 19 → menentukan bentuk titik (solid circle).

abline(lm(mpg ~ wt, data = mtcars)) berfungsi untuk menambahkan garis regresi linier ke dalam scatter plot berdasarkan model lm(mpg ~ wt). Fungsi lm() menghitung persamaan garis regresi dengan metode Ordinary Least Squares (OLS), sedangkan abline() menggambarkan garis hasil regresi tersebut di atas plot.

### Estimasi Parameter Analisis Regresi secara manual

```
# Menyimpan variabel X dan Y

x <- df$wt
y <- df$mpg
```

Kode diatas digunakan untuk memisahkan dua variabel penting:

x → variabel independen, yaitu berat mobil (wt) dalam 1000 lbs.

y → variabel dependen, yaitu efisiensi bahan bakar (mpg) dalam mil per galon.

```
# Menghitung nilai rata-rata

x_bar <- mean(x)
y_bar <- mean(y)
```

Fungsi mean() menghitung nilai rata-rata (mean) dari masing-masing variabel. Nilai rata-rata digunakan untuk menentukan posisi pusat distribusi data, serta menjadi komponen penting dalam rumus estimasi  $\beta_1$  dan  $\beta_0$

```
# Menghitung beta1 dan beta0 secara manual

num <- sum((x - x_bar) * (y - y_bar)) # pembilang
den <- sum((x - x_bar)^2) # penyebut

beta1_hat <- num / den
beta0_hat <- y_bar - beta1_hat * x_bar
```

Variabel num mengukur sejauh mana setiap pasangan data (X, Y) menyimpang bersama-sama dari nilai rata-ratanya.

Variabel den mengukur variasi total dari variabel X.

Variabel beta1\_hat menunjukkan arah dan kekuatan hubungan antara X dan Y. Jika  $\beta_1 < 0$ , maka hubungan bersifat negatif.

Variabel beta0\_hat menunjukkan perkiraan nilai Y ketika X = 0, yang secara matematis adalah titik potong garis regresi dengan sumbu Y.

```
# Menampilkan hasil estimasi manual

beta0_hat
## [1] 37.28513

beta1_hat
## [1] -5.344472
```

Baris ini menampilkan nilai hasil estimasi:

beta0\_hat → nilai intercept ( $\beta_0$ )  
beta1\_hat → nilai slope ( $\beta_1$ )

Hasil yang diperoleh akan identik dengan output dari fungsi `lm(mpg ~ wt, data = df)`, karena perhitungan manual menggunakan prinsip dasar yang sama dengan metode Ordinary Least Squares (OLS).

### Estimasi Parameter Analisis Regresi dengan menggunakan library `lm`

```
# Membuat model regresi linier sederhana
model <- lm(mpg ~ wt, data = df)
```

Fungsi `lm()` digunakan untuk membentuk model regresi linier.

Sintaks `mpg ~ wt` yaitu “Variabel mpg diregresikan terhadap variabel wt.”

Argumen `data = df` menunjukkan bahwa kedua variabel (mpg dan wt) diambil dari dataset bernama `df`.

```
# Menampilkan ringkasan hasil regresi
summary(model)

##
## Call:
## lm(formula = mpg ~ wt, data = df)
##
## Residuals:
##      Min      1Q  Median      3Q      Max
## -4.5432 -2.3647 -0.1252  1.4096  6.8727
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 37.2851    1.8776 19.858 < 2e-16 ***
## wt          -5.3445    0.5591 -9.559 1.29e-10 ***
## ---
## Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Residual standard error: 3.046 on 30 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7446
## F-statistic: 91.38 on 1 and 30 DF,  p-value: 1.294e-10
```

Fungsi `summary()` digunakan untuk menampilkan ringkasan lengkap hasil analisis regresi

Output dari fungsi `summary(model)` dalam R terdiri atas beberapa bagian utama yang memberikan informasi penting mengenai hasil estimasi model regresi linier sederhana. Berikut penjelasan tiap komponennya:

### 1. Call

Menunjukkan formula model yang digunakan.

Sebagai contoh:

$$mpg \sim wt$$

Artinya variabel `mpg` diregresikan terhadap variabel `wt`.

### 2. Residuals

Menampilkan ringkasan nilai **residu**, yaitu selisih antara nilai aktual dengan nilai prediksi. Secara matematis dituliskan sebagai:

$$e_i = Y_i - \hat{Y}_i$$

Residu menunjukkan seberapa jauh hasil prediksi model dari nilai sebenarnya. Nilai residu yang kecil menandakan bahwa model memiliki tingkat kesalahan yang rendah.

### 3. Coefficients

Bagian ini berisi informasi penting mengenai estimasi parameter model:

Komponen	Simbol	Keterangan
(Intercept)	$\hat{\beta}_0$	Konstanta model, yaitu nilai prediksi $Y$ ketika $X = 0$
wt	$\hat{\beta}_1$	Koefisien slope yang menunjukkan arah dan besar pengaruh variabel <code>wt</code> terhadap <code>mpg</code>

Selain itu, tabel *Coefficients* juga menampilkan beberapa ukuran statistik: - **Std. Error** → simpangan baku dari estimasi koefisien.

- **t value** → nilai uji t untuk menilai signifikansi masing-masing koefisien.

- **Pr(>|t|)** → nilai *p-value* dari uji t; jika nilai ini < 0.05, maka koefisien dianggap signifikan secara statistik.

### 4. Residual Standard Error (RSE)

Menunjukkan rata-rata kesalahan model dalam satuan yang sama dengan variabel dependen (`mpg`).

Rumusnya adalah:

$$RSE = \sqrt{\frac{\sum(Y_i - \hat{Y}_i)^2}{n - k - 1}}$$

- di mana: -  $n$  = jumlah observasi  
 -  $k$  = jumlah variabel independen

Semakin kecil nilai RSE, semakin baik model dalam menyesuaikan data.

## 5. Multiple R-squared dan Adjusted R-squared

Bagian ini mengukur seberapa besar proporsi variasi pada variabel dependen (mpg) yang dapat dijelaskan oleh variabel independen (wt).

Rumus umum:

$$R^2 = 1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y})^2}$$

- **Multiple R-squared (R<sup>2</sup>)** menunjukkan proporsi variasi mpg yang dijelaskan oleh model.
- **Adjusted R-squared** menyesuaikan nilai R<sup>2</sup> berdasarkan jumlah variabel independen dan ukuran sampel, sehingga lebih akurat untuk model dengan banyak variabel.

Nilai R<sup>2</sup> yang tinggi menunjukkan model mampu menjelaskan sebagian besar variasi data.

## 6. F-statistic

F-statistic digunakan untuk **menguji signifikansi model secara keseluruhan**, dengan hipotesis:

$$H_0: \beta_1 = 0 \quad (\text{tidak ada pengaruh signifikan antara wt terhadap mpg})$$

$$H_1: \beta_1 \neq 0$$

Jika nilai *p-value* untuk uji F lebih kecil dari 0.05, maka model regresi dianggap **signifikan secara statistik**, artinya wt berpengaruh terhadap mpg.

## 7. Kesimpulan

Secara keseluruhan, fungsi `summary(model)` memberikan: - Ringkasan lengkap hasil estimasi koefisien regresi.

- Ukuran signifikansi statistik untuk setiap parameter.
- Nilai kesalahan model (RSE) dan kemampuan model menjelaskan variasi data (R<sup>2</sup>).

Hasil ini menjadi dasar dalam menilai apakah hubungan antara berat mobil (wt) dan efisiensi bahan bakar (mpg) bersifat signifikan dan linier sesuai dengan asumsi model regresi.

**Perbandingan Estimasi Parameter Analisis Regresi dengan menggunakan masnual dan library lm**

```
cat("Hasil Manual: \n")
## Hasil Manual:
cat("Beta0 (Intercept):", round(beta0_hat, 4), "\n")
## Beta0 (Intercept): 37.2851
```

```

cat("Beta1 (Slope):", round(beta1_hat, 4), "\n\n")
## Beta1 (Slope): -5.3445

cat("Hasil Fungsi lm(): \n")
## Hasil Fungsi lm():

coef(model)

## (Intercept)          wt
## 37.285126 -5.344472

```

Berdasarkan hasil perbandingan antara perhitungan manual dan fungsi `lm()` di R, diperoleh bahwa nilai koefisien regresi linier sederhana untuk model hubungan antara **berat mobil (wt)** dan **efisiensi bahan bakar (mpg)** menunjukkan hasil yang identik. Nilai **intercept ( $\beta_0$ )** yang diperoleh adalah sebesar **37.2851**, sedangkan nilai **slope ( $\beta_1$ )** adalah **-5.3445**. Kesamaan hasil ini menunjukkan bahwa fungsi `lm()` di R menggunakan metode **Ordinary Least Squares (OLS)** yang sama dengan perhitungan manual, yaitu dengan meminimalkan jumlah kuadrat selisih antara nilai aktual dan nilai prediksi:

$$\min_{\beta_0, \beta_1} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Secara interpretatif, model regresi linier sederhana yang diperoleh dapat dituliskan sebagai:

$$\widehat{mpg} = 37.2851 - 5.3445 \times wt$$

Model ini menjelaskan bahwa setiap kenaikan berat mobil sebesar **1 satuan (1000 lbs)** akan menurunkan efisiensi bahan bakar rata-rata sebesar **5.34 mil per galon (mpg)**. Nilai **intercept** sebesar **37.2851** menunjukkan nilai prediksi efisiensi bahan bakar ketika berat mobil mendekati nol, yang secara matematis merupakan **titik potong garis regresi dengan sumbu Y**.

Secara keseluruhan, hasil ini memperkuat pemahaman bahwa hubungan antara **berat mobil dan efisiensi bahan bakar bersifat negatif dan signifikan secara statistik**, di mana mobil yang lebih berat cenderung memiliki konsumsi bahan bakar lebih tinggi. Kesamaan hasil antara perhitungan manual dan fungsi `lm()` juga menunjukkan **validitas metode estimasi OLS** yang digunakan dalam analisis regresi linier sederhana.

**Uji hipotesis terhadap estimasi parameter  $b_1$**

Selain estimasi parameter, analisis regresi linier sederhana memungkinkan dilakukannya **pengujian hipotesis** terhadap parameter model. Tujuannya adalah untuk menentukan apakah hubungan antara X dan Y signifikan secara statistik. Uji hipotesis yang digunakan adalah uji t (parsial). Jika  $p\text{-value} < 0.05$ , maka  $H_0$  ditolak, artinya X berpengaruh signifikan terhadap Y.

Berdasarkan kasus diatas, maka nilai dari uji t adalah  $-9.56$  dengan  $p\text{-value}$  sebesar  $1.29 \times 10^{-10}$ . Karena  $p\text{-value} < 0.05$  maka tolak  $H_0$  yang berarti terdapat pengaruh signifikan antara berat mobil (`wt`) terhadap efisiensi bahan bakar (`mpg`)

## Uji hipotesis terhadap estimasi parameter $b_0$

Uji hipotesis yang selanjutnya dilakukan terhadap  $b_0$  dilakukan untuk melihat apakah semua nilai Y dapat dijelaskan oleh x atau tidak. Jika  $p\text{-value} < 0.05$ , maka  $H_0$  ditolak, artinya ada nilai Y yang tidak dapat dijelaskan oleh x.

Berdasarkan kasus diatas, maka nilai dari uji t untuk intercept adalah 19.85 dengan p-value sebesar  $< 2e - 16$ . Karena p-value  $< 0.05$  maka tolak  $H_0$  yang berarti ada efisiensi bahan bakar (*mpg*) yang tidak dapat dijelaskan oleh berat mobil (*wt*).

## Asumsi Klasik Regresi Linier

Dalam analisis regresi linier, terdapat beberapa asumsi dasar atau **asumsi klasik** yang harus dipenuhi agar hasil estimasi menggunakan metode **Ordinary Least Squares (OLS)** menjadi **BLUE (Best Linear Unbiased Estimator)**. Artinya, estimator yang dihasilkan memiliki sifat **terbaik (efisien), linier, dan tidak bias** sebagai berikut.

### 1. Linearitas Hubungan

Asumsi ini menyatakan bahwa hubungan antara variabel independen (X) dan variabel dependen (Y) bersifat **linier**. Model regresi mengasumsikan bahwa perubahan pada X akan menghasilkan perubahan proporsional pada Y. Jika hubungan sebenarnya bersifat non-linier, maka model linier akan memberikan estimasi yang bias.

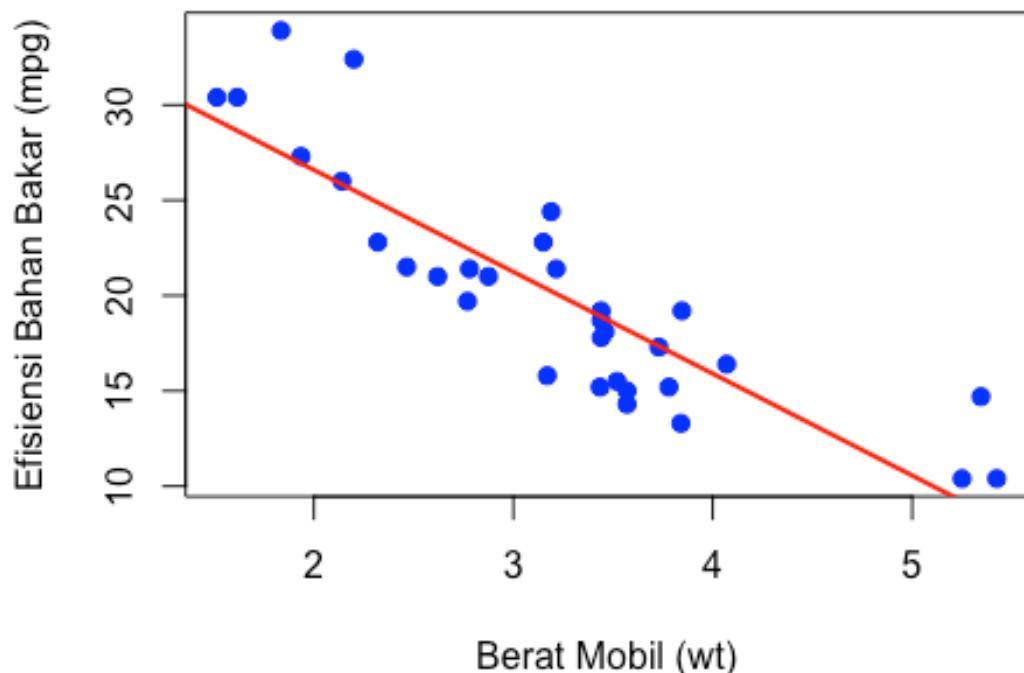
$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Berdasarkan konsep diatas, maka dengan kasus yang sama akan dicari apakah efisiensi bahan bakar (*mpg*) dan berat mobil (*wt*) memiliki hubungan yang linier.

```
# Scatter plot dengan garis regresi

plot(df$wt, df$mpg,
      xlab = "Berat Mobil (wt)",
      ylab = "Efisiensi Bahan Bakar (mpg)",
      main = "Uji Linearitas: Hubungan antara wt dan mpg",
      pch = 19, col = "blue")
abline(model, col = "red", lwd = 2)
```

## Uji Linearitas: Hubungan antara wt dan mpg



Interpretasi dari grafik diatas:

- Titik-titik data (biru) menunjukkan hubungan negatif antara wt dan mpg.
- Garis merah menunjukkan garis regresi linier.
- Pola titik yang menyebar di sekitar garis secara relatif merata mengindikasikan bahwa hubungan antarvariabel bersifat linier.

### 2. Ekspektasi Nilai Error = 0

Asumsi ini menyatakan bahwa nilai rata-rata dari error ( $\varepsilon_i$ ) untuk seluruh observasi adalah **nol**.

$$E(\varepsilon_i) = 0$$

Artinya, model tidak secara sistematis melebihikan atau meremehkan nilai Y. Jika asumsi ini tidak terpenuhi, model memiliki bias dalam prediksi.

Berdasarkan konsep diatas, maka dengan kasus yang sama akan dicari apakah Ekspektasi Nilai Error = 0.

```
# Mengambil nilai residual
residuals_model <- resid(model)
# Melihat beberapa nilai residual pertama
head(residuals_model)
```

```

##          Mazda RX4      Mazda RX4 Wag      Datsun 710      Hornet 4 Drive
##          -2.2826106     -0.9197704     -2.0859521     1.2973499
## Hornet Sportabout      Valiant
##          -0.2001440     -0.6932545

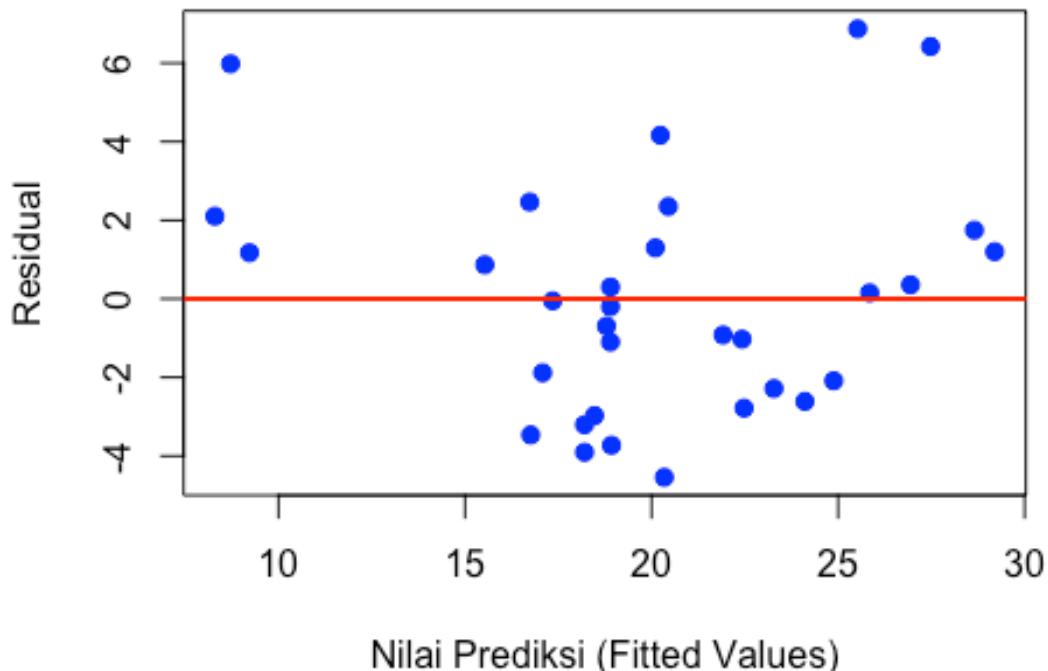
# Menghitung nilai rata-rata residual
mean_residual <- mean(residuals_model)
mean_residual

## [1] 2.220446e-16

# Membuat plot residual terhadap nilai prediksi
plot(fitted(model), residuals_model,
xlab = "Nilai Prediksi (Fitted Values)",
ylab = "Residual",
main = "Pemeriksaan Asumsi  $E(\epsilon) = 0$ ",
pch = 19, col = "blue")
abline(h = 0, col = "red", lwd = 2)

```

## Pemeriksaan Asumsi $E(\epsilon) = 0$



Interpretasi:

- Jika `mean_residual` mendekati nol (misalnya antara  $-1e-15$  hingga  $1e-15$ ), maka asumsi ini terpenuhi. Nilai yang mendekati nol menunjukkan bahwa model tidak bias secara sistematis dalam melakukan prediksi.

- Garis merah menunjukkan garis horizontal di nol. Titik residual (biru) menyebar di sekitar garis nol secara acak tanpa pola tertentu. Hal ini menunjukkan bahwa rata-rata error mendekati nol dan tidak terjadi bias sistematis.

### 3. Homoskedastisitas (Varians Error Konstan)

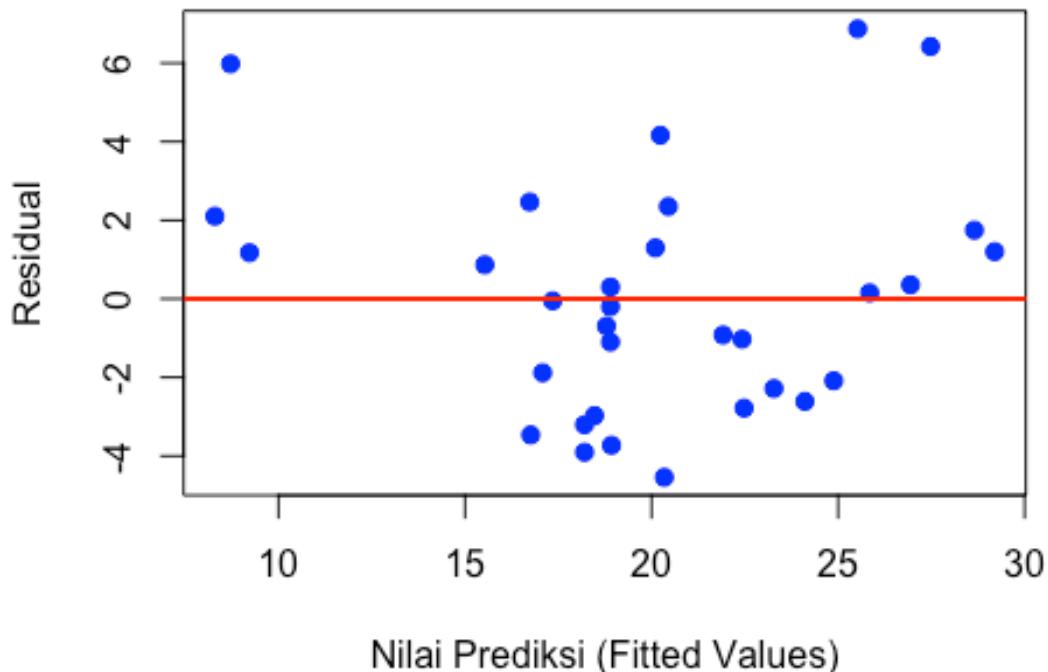
Asumsi ini menyatakan bahwa varians error untuk setiap nilai X adalah **konstan**.

$$Var(\varepsilon_i) = \sigma^2$$

Jika varians error berubah-ubah (heteroskedastisitas), maka estimasi tetap tidak bias tetapi **tidak efisien**, dan uji statistik seperti uji t menjadi tidak valid. Berdasarkan konsep diatas, maka dengan kasus yang sama akan dicari apakah variabel memenuhi asumsi homoskedastisitas.

```
# Simpan residual dan fitted value
residuals_model <- resid(model)
fitted_model <- fitted(model)
# Pengecekan homoskedastisitas menggunakan Plot residual vs fitted
plot(fitted_model, residuals_model,
xlab = "Nilai Prediksi (Fitted Values)",
ylab = "Residual",
main = "Uji Visual Homoskedastisitas",
pch = 19, col = "blue")
abline(h = 0, col = "red", lwd = 2)
```

## Uji Visual Homoskedastisitas



```

# Pengecekan homoskedasitas menggunakan Paket untuk uji Breusch-Pagan
library(lmtest)

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

# Uji Breusch-Pagan
bptest(model)

##
## studentized Breusch-Pagan test
##
## data: model
## BP = 0.040438, df = 1, p-value = 0.8406

```

Interpretasi:

- Jika titik menyebar acak di sekitar garis 0 dan lebar sebarannya relatif sama, asumsi homoskedastisitas cenderung terpenuhi. Jika terlihat pola “kipas” (fan shape) atau corong (semakin melebar / menyempit), itu indikasi heteroskedastisitas.
- Jika p-value > 0.05  $\rightarrow$  gagal tolak  $H_0$   $\rightarrow$  tidak ada bukti heteroskedastisitas  $\rightarrow$  asumsi homoskedastisitas terpenuhi.

## 6. Error Terdistribusi Normal

Asumsi ini diperlukan terutama untuk **pengujian hipotesis statistik (uji t)**.

Error ( $\varepsilon_i$ ) diasumsikan mengikuti distribusi normal dengan rata-rata 0 dan varians konstan.

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

Jika asumsi ini tidak terpenuhi, maka uji signifikansi koefisien bisa menjadi tidak valid, terutama pada ukuran sampel kecil.

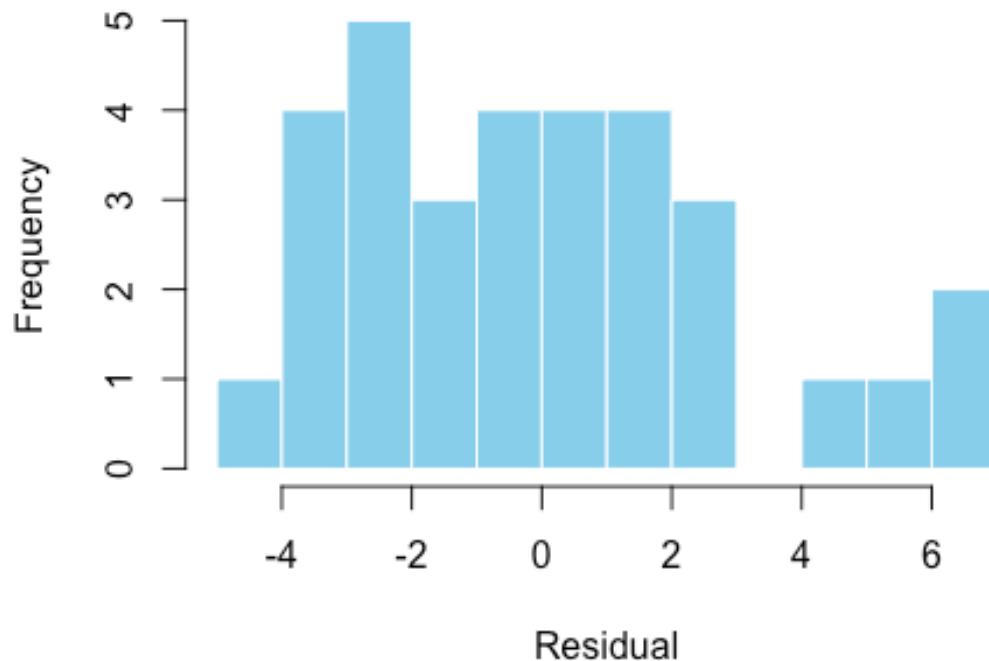
# Histogram residual

```

hist(residuals_model,
main = "Distribusi Residual",
xlab = "Residual",
col = "skyblue",
border = "white",
breaks = 10)

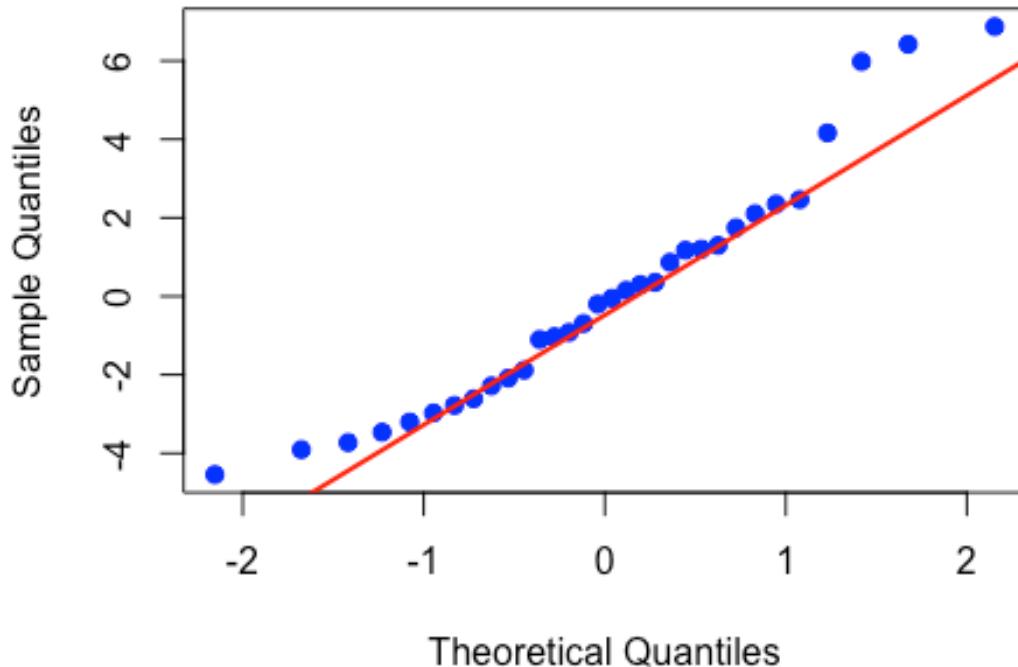
```

## Distribusi Residual



```
# Membuat Q-Q plot
qqnorm(residuals_model,
main = "Q-Q Plot Residual",
pch = 19, col = "blue")
qqline(residuals_model, col = "red", lwd = 2)
```

## Q-Q Plot Residual



```
# Uji Shapiro-Wilk
shapiro.test(residuals_model)

##
##  Shapiro-Wilk normality test
##
## data: residuals_model
## W = 0.94508, p-value = 0.1044
```

Interpretasi :

- Jika bentuk histogram mendekati lonceng (bell-shaped), maka residual dapat dikatakan mendekati distribusi normal. Jika distribusi terlihat miring (skewed) atau memiliki puncak ganda, maka asumsi normalitas mungkin tidak terpenuhi.
- Titik-titik yang mengikuti garis merah menunjukkan bahwa residual berdistribusi mendekati normal. Jika banyak titik menyimpang jauh dari garis, terutama di ekor distribusi, maka indikasi ketidaknormalan muncul.
- Jika p-value > 0.05  $\rightarrow$  gagal tolak  $H_0 \rightarrow$  residual berdistribusi normal.

Sehingga dapat disimpulkan bahwa pemenuhan asumsi klasik sangat penting dalam analisis regresi linier karena memengaruhi:

1. Keakuratan estimasi parameter ( $\hat{\beta}$ )
2. Validitas pengujian statistik (uji-t)
3. Kelayakan model untuk prediksi

Jika salah satu asumsi dilanggar, maka analisis harus dilanjutkan dengan **pengujian diagnostik** dan **perbaikan model**, seperti transformasi data, penggunaan robust standard error, atau model alternatif.

Berdasarkan hasil serangkaian pengujian terhadap asumsi klasik pada model regresi linier sederhana antara **berat mobil (wt)** dan **efisiensi bahan bakar (mpg)** menggunakan dataset *mtcars*, diperoleh hasil bahwa seluruh asumsi utama regresi linier terpenuhi dengan baik.

1. Uji **linearitas hubungan** menunjukkan bahwa hubungan antara wt dan mpg bersifat linier. Hal ini terlihat dari pola sebaran data yang mengikuti garis regresi serta hasil *component + residual plot* yang tidak menunjukkan deviasi berarti dari garis lurus.
2. Asumsi **ekspektasi nilai error = 0** terpenuhi karena nilai rata-rata residual mendekati nol dan tidak menunjukkan pola sistematis terhadap nilai prediksi. Hal ini menandakan bahwa model tidak memiliki bias dalam memprediksi variabel dependen.
3. Hasil uji **homoskedastisitas** (baik secara visual maupun dengan uji Breusch-Pagan) menunjukkan bahwa varians residual relatif konstan di seluruh rentang nilai prediksi, dengan *p-value* yang lebih besar dari 0.05. Ini berarti tidak terdapat indikasi heteroskedastisitas dalam model.
4. Hasil uji **normalitas residual** melalui histogram, Q-Q plot, dan uji Shapiro-Wilk memperlihatkan bahwa residual mengikuti distribusi normal, ditandai dengan pola titik yang hampir mengikuti garis lurus pada Q-Q plot dan nilai *p-value* uji normalitas lebih besar dari 0.05.

Secara keseluruhan, hasil tersebut mengindikasikan bahwa model regresi linier sederhana:

$$\widehat{mpg} = 37.2851 - 5.3445 \times wt$$

telah memenuhi seluruh asumsi klasik OLS (Ordinary Least Squares), sehingga estimator yang dihasilkan bersifat **BLUE (Best Linear Unbiased Estimator)**. Dengan demikian, model ini dapat digunakan untuk melakukan estimasi dan inferensi statistik secara valid dalam konteks hubungan antara berat kendaraan dan efisiensi bahan bakar.

## Korelasi

Korelasi merupakan ukuran statistik yang menunjukkan **arah dan kekuatan hubungan linier** antara dua variabel.

Dalam konteks ini, kita ingin mengetahui seberapa kuat hubungan antara **berat mobil (wt)** dan **efisiensi bahan bakar (mpg)**.

Nilai koefisien korelasi Pearson ( $r$ ) berada dalam rentang:

$$-1 \leq r \leq 1$$

```

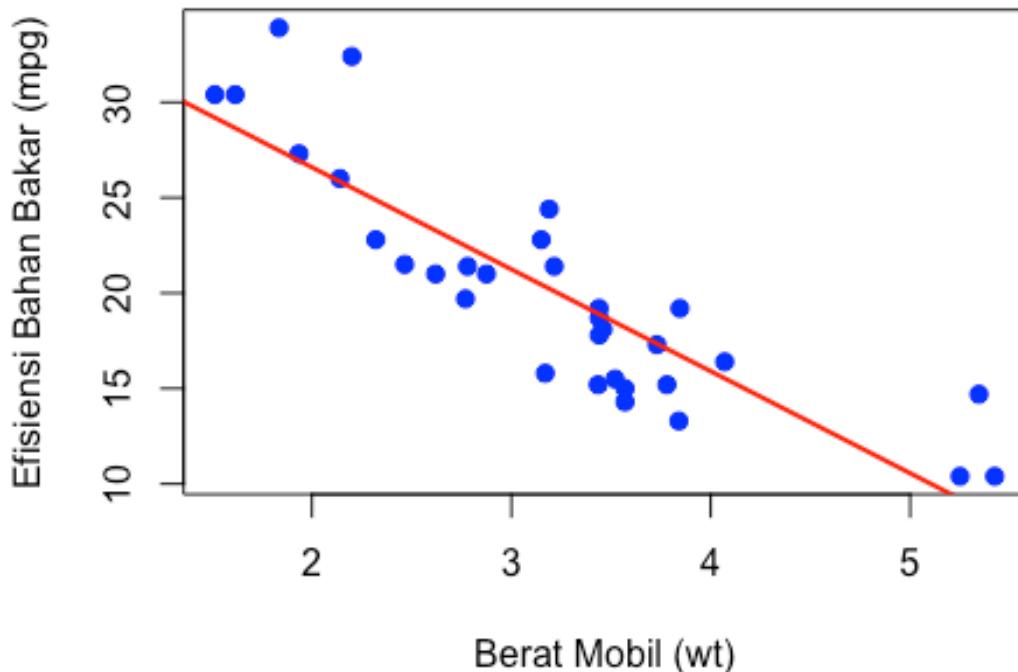
# Menghitung korelasi Pearson
correlation <- cor(df$mpg, df$wt, method = "pearson")
correlation

## [1] -0.8676594

# Visualisasi scatter plot dengan garis regresi
plot(df$wt, df$mpg,
xlab = "Berat Mobil (wt)",
ylab = "Efisiensi Bahan Bakar (mpg)",
main = paste("Korelasi antara wt dan mpg (r =", round(correlation, 3), ")"),
pch = 19, col = "blue")
abline(lm(mpg ~ wt, data = df), col = "red", lwd = 2)

```

## Korelasi antara wt dan mpg (r = -0.868 )



```

# Uji signifikansi korelasi
cor_test <- cor.test(df$mpg, df$wt, method = "pearson")
cor_test

##
## Pearson's product-moment correlation
##
## data: df$mpg and df$wt
## t = -9.559, df = 30, p-value = 1.294e-10
## alternative hypothesis: true correlation is not equal to 0

```

```
## 95 percent confidence interval:  
## -0.9338264 -0.7440872  
## sample estimates:  
## cor  
## -0.8676594
```

Interpretasi:

1. Nilai korelasi negatif menunjukkan bahwa semakin berat mobil, efisiensi bahan bakarnya semakin rendah.
2. Garis merah menggambarkan arah hubungan negatif antara wt dan mpg. Titik-titik data tersebar cukup dekat dengan garis regresi, menandakan hubungan linier yang kuat dan negatif.
3. Jika p-value < 0.05 → tolak  $H_0$  → ada korelasi signifikan.

Berdasarkan hasil perhitungan diperoleh bahwa nilai korelasi Pearson antara wt dan mpg adalah sekitar -0.87 dengan p-value yang jauh lebih kecil dari 0.05. Hal ini menunjukkan bahwa terdapat hubungan negatif yang sangat kuat dan signifikan antara berat mobil dan efisiensi bahan bakar. Yang berarti semakin berat mobil, semakin rendah efisiensi bahan bakar yang dihasilkan. Hasil ini juga konsisten dengan model regresi linier sederhana yang menunjukkan koefisien slope negatif (-5.3445).