

# Tugas 2 Praktikum PSD

Zafyra Nur Rizqi

2025-05-31

## Data

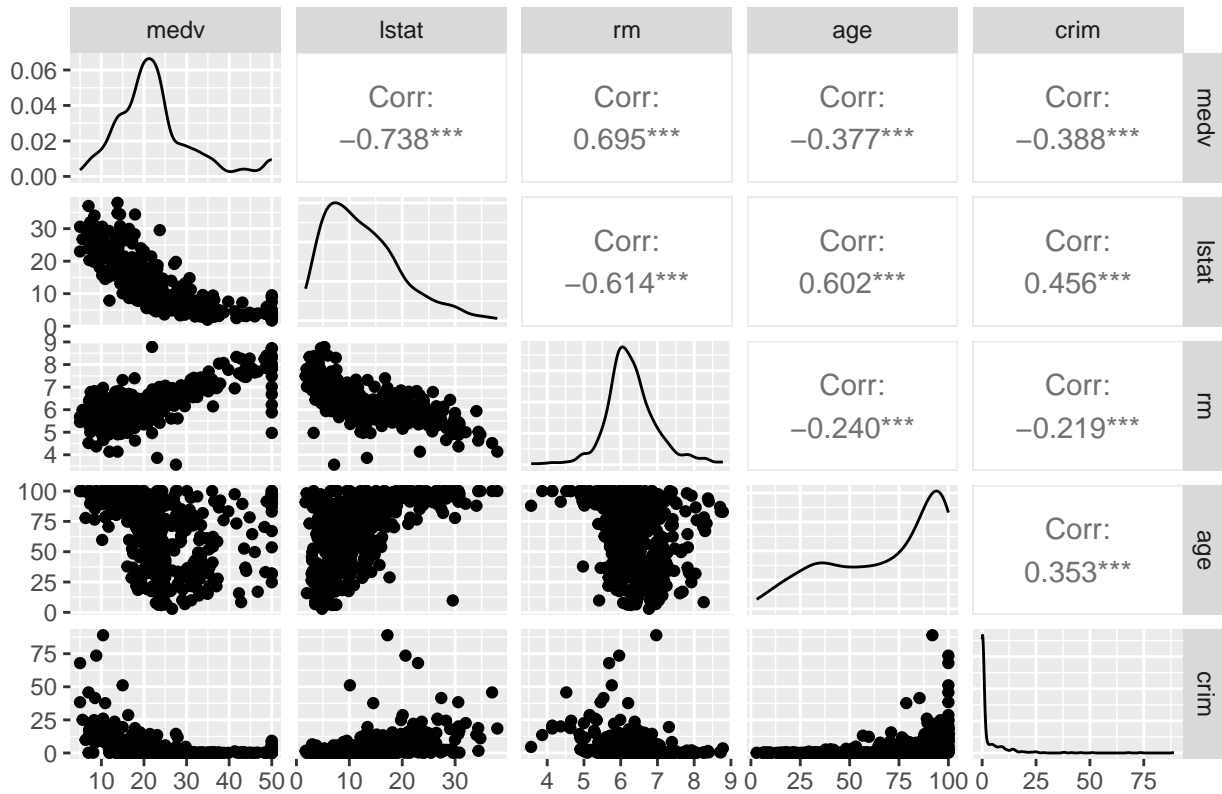
---

```
##      crim zn indus chas   nox    rm age    dis rad tax ptratio  black lstat
## 1 0.00632 18  2.31    0 0.538 6.575 65.2 4.0900   1 296    15.3 396.90  4.98
## 2 0.02731  0  7.07    0 0.469 6.421 78.9 4.9671   2 242    17.8 396.90  9.14
## 3 0.02729  0  7.07    0 0.469 7.185 61.1 4.9671   2 242    17.8 392.83  4.03
## 4 0.03237  0  2.18    0 0.458 6.998 45.8 6.0622   3 222    18.7 394.63  2.94
## 5 0.06905  0  2.18    0 0.458 7.147 54.2 6.0622   3 222    18.7 396.90  5.33
## 6 0.02985  0  2.18    0 0.458 6.430 58.7 6.0622   3 222    18.7 394.12  5.21
##   medv
## 1 24.0
## 2 21.6
## 3 34.7
## 4 33.4
## 5 36.2
## 6 28.7
```

```
##      crim          zn          indus          chas
## Min.   : 0.00632   Min.   : 0.00   Min.   : 0.46   Min.   :0.00000
## 1st Qu.: 0.08205   1st Qu.: 0.00   1st Qu.: 5.19   1st Qu.:0.00000
## Median : 0.25651   Median : 0.00   Median : 9.69   Median :0.00000
## Mean   : 3.61352   Mean   : 11.36   Mean   :11.14   Mean   :0.06917
## 3rd Qu.: 3.67708   3rd Qu.: 12.50   3rd Qu.:18.10   3rd Qu.:0.00000
## Max.   :88.97620   Max.   :100.00   Max.   :27.74   Max.   :1.00000
##      nox          rm          age          dis
## Min.   :0.3850   Min.   :3.561   Min.   : 2.90   Min.   : 1.130
## 1st Qu.:0.4490   1st Qu.:5.886   1st Qu.: 45.02   1st Qu.: 2.100
## Median :0.5380   Median :6.208   Median : 77.50   Median : 3.207
## Mean   :0.5547   Mean   :6.285   Mean   : 68.57   Mean   : 3.795
## 3rd Qu.:0.6240   3rd Qu.:6.623   3rd Qu.: 94.08   3rd Qu.: 5.188
## Max.   :0.8710   Max.   :8.780   Max.   :100.00   Max.   :12.127
##      rad          tax          ptratio          black
## Min.   : 1.000   Min.   :187.0   Min.   :12.60   Min.   : 0.32
## 1st Qu.: 4.000   1st Qu.:279.0   1st Qu.:17.40   1st Qu.:375.38
## Median : 5.000   Median :330.0   Median :19.05   Median :391.44
## Mean   : 9.549   Mean   :408.2   Mean   :18.46   Mean   :356.67
## 3rd Qu.:24.000   3rd Qu.:666.0   3rd Qu.:20.20   3rd Qu.:396.23
## Max.   :24.000   Max.   :711.0   Max.   :22.00   Max.   :396.90
##      lstat          medv
```

```
## Min.    : 1.73    Min.    : 5.00
## 1st Qu.: 6.95    1st Qu.:17.02
## Median :11.36    Median :21.20
## Mean   :12.65    Mean   :22.53
## 3rd Qu.:16.95    3rd Qu.:25.00
## Max.    :37.97    Max.    :50.00
```

### Korelasi Antar Variabel Penting



```
##          medv      lstat      rm      age      crim
## medv    1.0000000 -0.7376627  0.6953599 -0.3769546 -0.3883046
## lstat  -0.7376627  1.0000000 -0.6138083  0.6023385  0.4556215
## rm      0.6953599 -0.6138083  1.0000000 -0.2402649 -0.2192467
## age    -0.3769546  0.6023385 -0.2402649  1.0000000  0.3527343
## crim   -0.3883046  0.4556215 -0.2192467  0.3527343  1.0000000
```

Pengertian: Saya memakai dataset Boston yang berasal dari package R yaitu MASS. Dataset ini berisi data properti di wilayah Boston. Beberapa variabel penting dalam dataset ini berupa medv (median harga rumah), lstat (persentase penduduk dengan status sosial rendah), rm (rata-rata jumlah kamar/rumah), age (proporsi rumah tua), crim (tingkat kejahatan per kapita), dan indus (proporsi area bisnis non retail). Adapula package library yang saya gunakan guna menunjang analisis ini, yaitu berupa MASS (dataset Boston), ggplot2 (visualisasi), dplyr (manipulasi data), psych (statistik deskriptif), car (uji asumsi klasik), lmtest (uji homoskedastisitas).

## Model Regresi

```
model <- lm(medv ~ lstat + rm + age, data = Boston)
summary(model)
```

```
##
## Call:
## lm(formula = medv ~ lstat + rm + age, data = Boston)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.210  -3.467  -1.053   1.957  27.500
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.175311   3.181924  -0.369   0.712
## lstat       -0.668513   0.054357 -12.298 <2e-16 ***
## rm          5.019133   0.454306  11.048 <2e-16 ***
## age         0.009091   0.011215   0.811   0.418
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.542 on 502 degrees of freedom
## Multiple R-squared:  0.639, Adjusted R-squared:  0.6369
## F-statistic: 296.2 on 3 and 502 DF, p-value: < 2.2e-16
```

Pengertian: Analisis ini saya buat untuk mengetahui faktor-faktor apa saja yang memengaruhi harga rumah (medv) di Boston maka dari itu digunakanlah regresi berganda untuk memodelkan hubungan antara harga rumah dan beberapa variabel prediktor, yaitu lstat, rm dan age sehingga dari hasil model regresi ini, dapat dilihat variabel mana yang paling berpengaruh, memprediksi harga rumah, dan menguji apakah model ini valid melalui uji asumsi klasik.

---

## Uji Asumsi Klasik

### a. Multikolineritas

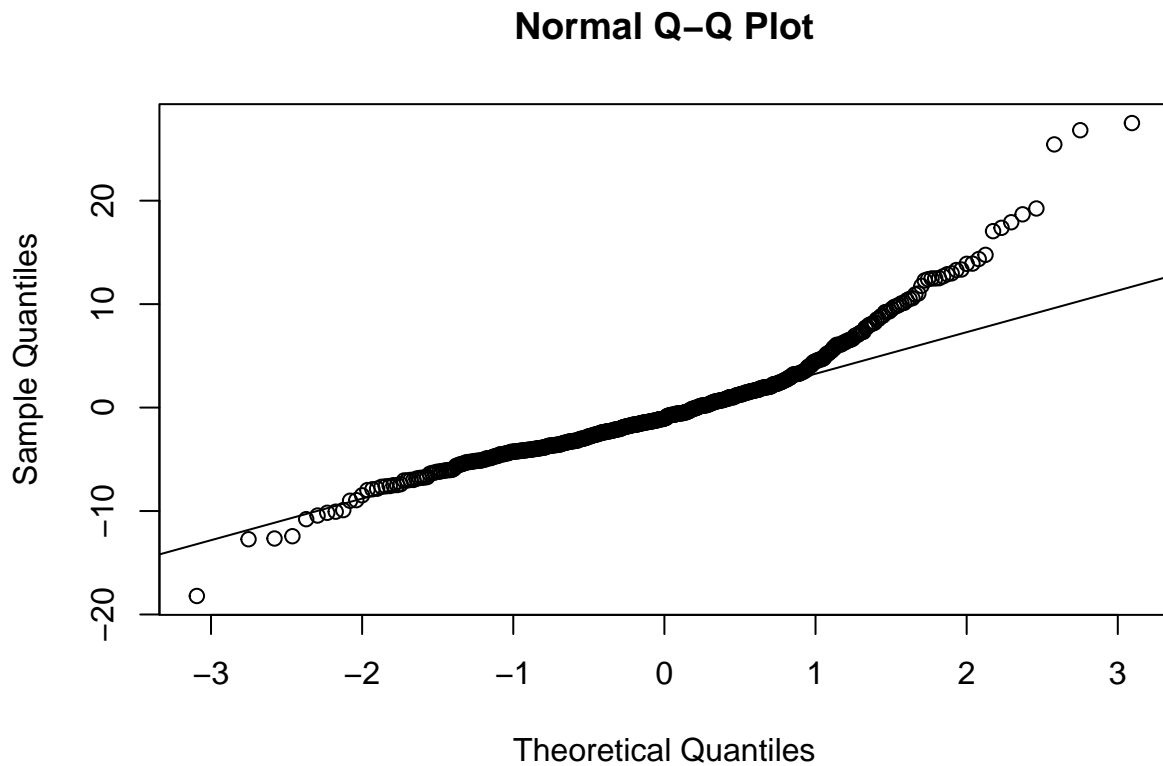
```
vif(model)
```

```
##      lstat      rm      age
## 2.477304 1.675215 1.638542
```

Pengertian: Menghitung Variance Inflation Factor.  $VIF > 10$  artinya terjadi multikolinearitas. Jika semua  $< 10$ , aman.

### b. Normalitas Residual

```
qqnorm(resid(model))
qqline(resid(model))
```



```
shapiro.test(resid(model))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(model)
## W = 0.91406, p-value = 2.385e-16
```

Pengertian: Fungsi ini untuk mengetahui apakah residual berdistribusi normal, dibantu dengan visualisasi datanya. Dilakukan juga uji statistik Shapiro-Wilk, yang dimana jika hasil nilai  $p > 0.05$  = data normal.

### c. Homoskedastisitas

```
bptest(model)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model
## BP = 19.771, df = 3, p-value = 0.0001894
```

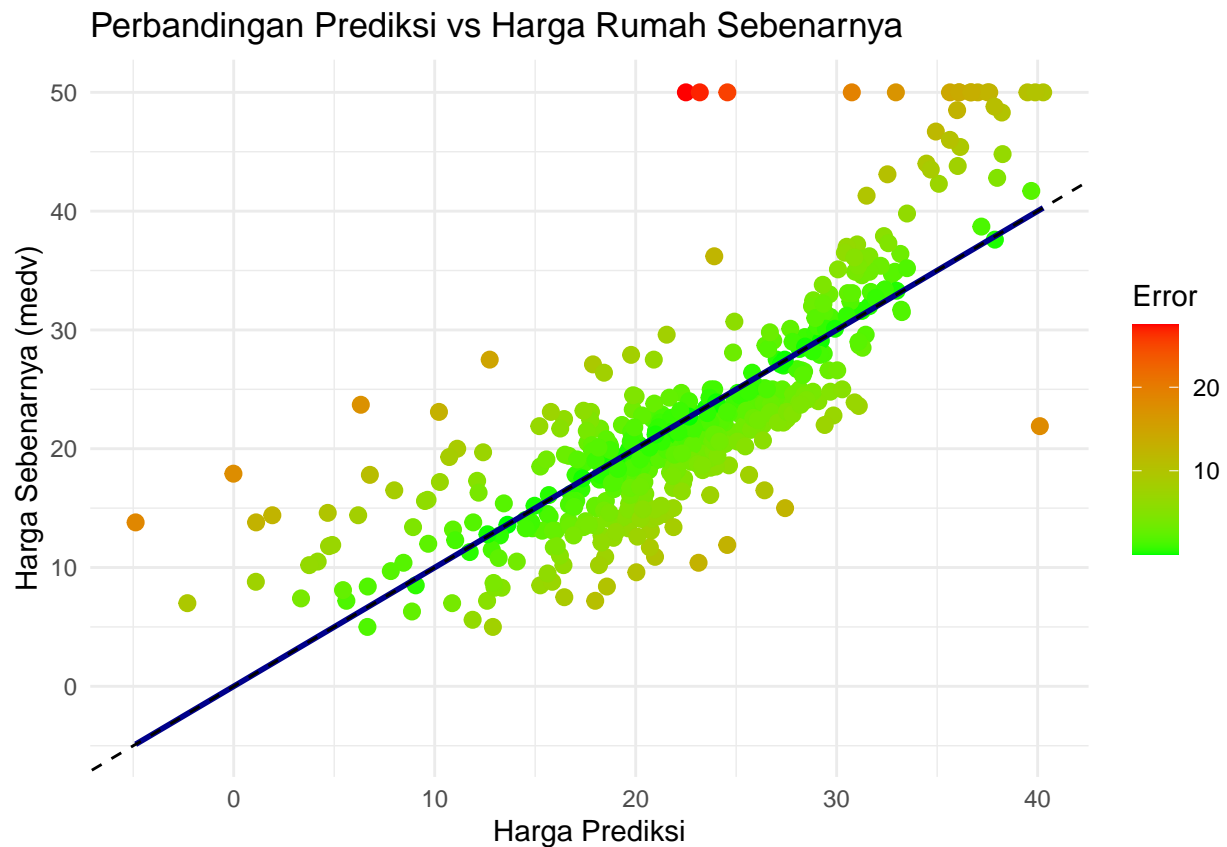
Pengertian: Dilakukan uji Breusch-Pagan yang dimana jika hasil p-value  $> 0.05$  menunjukkan residual memiliki varians konstan (tidak heteroskedastis).

## Visualisasi

### a. Plot Prediksi vs Realisasi

```
prediksi <- predict(model)
ggplot(data = Boston, aes(x = prediksi, y = medv)) +
  geom_point(aes(color = abs(prediksi - medv)), size = 2.5) +
  scale_color_gradient(low = "green", high = "red") +
  geom_smooth(method = "lm", se = FALSE, color = "darkblue") +
  geom_abline(slope = 1, intercept = 0, color = "black", linetype = "dashed") +
  labs(
    title = "Perbandingan Prediksi vs Harga Rumah Sebenarnya",
    x = "Harga Prediksi",
    y = "Harga Sebenarnya (medv)",
    color = "Error"
  ) +
  theme_minimal()
```

## 'geom\_smooth()' using formula = 'y ~ x'



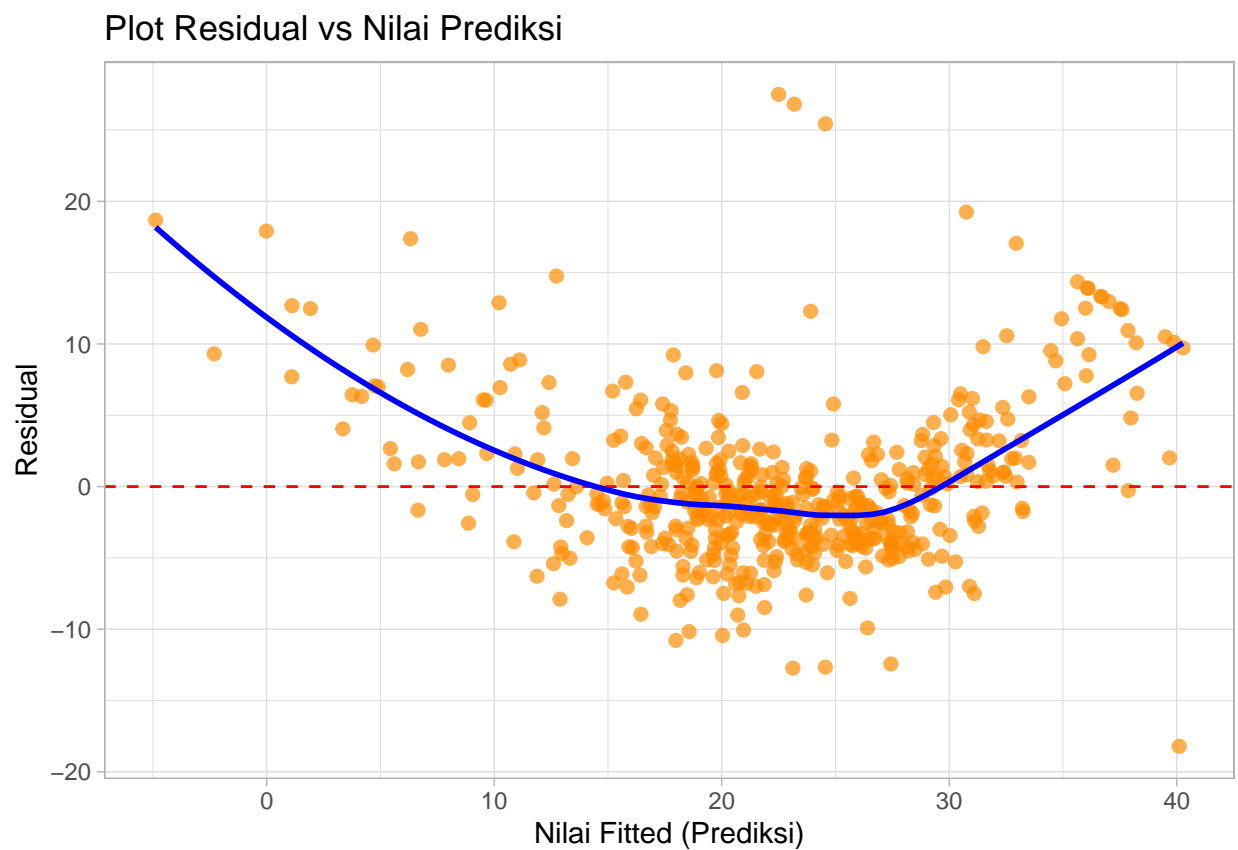
Pengertian:

## b. Plot Residual vs Fitted

```
residuals <- resid(model)
fitted <- fitted(model)

ggplot(data = NULL, aes(x = fitted, y = residuals)) +
  geom_point(color = "darkorange", alpha = 0.7, size = 2) +
  geom_hline(yintercept = 0, color = "red", linetype = "dashed") +
  geom_smooth(method = "loess", se = FALSE, color = "blue") +
  labs(
    title = "Plot Residual vs Nilai Prediksi",
    x = "Nilai Fitted (Prediksi)",
    y = "Residual"
  ) +
  theme_light()
```

## 'geom\_smooth()' using formula = 'y ~ x'



Pengertian:

### c. Plot Koefisien Model

```
tidy(model) %>%  
  filter(term != "(Intercept)") %>%  
  ggplot(aes(x = reorder(term, estimate), y = estimate)) +  
  geom_col(fill = "steelblue") +  
  coord_flip() +  
  labs(title = "Koefisien Model Regresi", x = "Variabel", y = "Estimasi Koefisien") +  
  theme_minimal()
```

