

Tugas Kelompok Analisis Regresi

Kelompok 2 - Kelas Paralel 1

2024-02-11

Anggota Kelompok 2

Hariyol G1401221009

Ghonniyu Hiban Saputra G1401221012

Rakesha Putra Antique G1401221056

Persamaan Regresi Linear Sederhana

Membaca data dari luar R dengan menggunakan fungsi read.csv

```
data <- read.csv("C:/Users/Ghonniyu/Downloads/Data Anreg Kelompok 2.csv")
```

```
y <- data$price  
x <- data$sqft_living
```

Membuat Data Frame

```
data_saya <- data.frame(cbind(y,x))
```

```
model_1 <- lm(y~x,data_saya)  
summary(model_1)
```

Pendugaan Parameter menggunakan Metode Kuadrat Terkecil (MKT)

#Menghitung nilai rata-rata x dan y

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

```
x_bar
```

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

```
y_bar
```

Menghitung nilai β_1 (Koefisien Regresi/Slope)

```
beta1 <- sum((x-x_bar)*(y-y_bar))/sum((x-x_bar)^2)  
beta1
```

Menghitung nilai β_0 (Intersep)

```
beta0 <- y_bar - beta1 *x_bar  
beta0
```

```
$$ \hat{\beta}_1 = \frac{\sum(y_i - \bar{y})(x_i - \bar{x})}{\sum(x_i - \bar{x})^2} \\ \hat{\beta}_1 = 221.7486 \\ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_0 = 80895.31 $
```

Berdasarkan perhitungan, didapat dugaan persamaan regresi linear sebagai berikut. Dugaan persamaan tersebut dapat pula disebut sebagai nilai harapan dari peubah respons Y

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X \quad \hat{Y} = 80895.31 + 221.7486X$$

Dari persamaan regresi linear sederhana tersebut dapat diinterpretasikan sebagai berikut:

Intersep adalah 80895.31. ini adalah nilai yang diprediksi dari harga rumah ketika luas tanah (X) sama dengan nol. Dalam konteks ini tidak memiliki interpretasi praktis yang signifikan, karena luas tanah tidak mungkin nol.

Koefisien regresi adalah 221.7486. Ini menunjukkan bahwa untuk setiap peningkatan satu unit dalam luas tanah akan mengakibatkan harga rumah meningkat sebesar 221.7486. Ini menunjukkan adanya hubungan linier positif antara luas tanah dan harga rumah, dengan asumsi semua faktor lainnya tetap konstan.

Eksplorasi Data

Scatter Plot

```
titik <- 10
x_titik <- x[titik]
y_titik <- y[titik]
y_prediksi <- fitted(model_1)[titik]
# Colpal
color_palette <- colorRampPalette(colors = c("black"))

# Base
plot(x, y, pch = 16, xlab = "Luas Rumah", ylab = "Harga Rumah", col =
color_palette(length(y)))
abline(model_1, col = "red", lwd = 1.5)

# xy strip
segments(x_titik, 0, x_titik, y_titik, col = "green", lty = 2, lwd =
1.5)
segments(0, y_titik, x_titik, y_titik, col = "green", lty = 2, lwd =
1.5)

# error line
segments(x_titik - 2, y_titik - 0.3, x_titik + 2, y_prediksi + 0.3, col
= "purple")

# error arrow
arrows(x_titik - 2, y_titik - 0.3, x_titik + 2, y_prediksi + 0.3, length =
0.04, col = "green")
arrows(x_titik - 2, y_prediksi + 0.3, x_titik + 2, y_titik - 0.3, length =
0.04, col = "green")
```

Berdasarkan *scatter plot* tersebut, dapat dilihat dari garis regresi yang terbentuk terdapat hubungan yang positif dan juga terlihat beberapa data pencilan yang menyimpang dari garis regresi tersebut.

Keragaman Dugaan Parameter

```
(anova.model <- anova(model_1))
```

```
(KTG <- anova.model$`Mean Sq`[2])
```

```
(galat.baku <- sqrt(KTG))
```

1.Dugaan Parameter β_0

Dengan nilai intersep(b_0) sebesar 80895.31 ,selanjutnya akan dilakukan uji apakah ada variasi Harga Rumah (y) yang tidak dapat dijelaskan oleh besarnya Luas Rumah (x).

Hipotesis uji:

$H_0: \beta_1 = 0$ (Semua Harga Rumah dapat dijelaskan dengan besarnya Luas Rumah)

$H_1: \beta_1 \neq 0$ (Ada Harga Rumah yang tidak dapat dijelaskan oleh besarnya Luas Rumah)

```
(se_b1 <- sqrt(KTG/sum((x-mean(x))^2)))
```

```
(t_b1 <- beta1/se_b1)
```

$t_{hitung} = 17.72756$

$t_{(db, \alpha/2)} = 2.036933343$

Karena $t_{hitung} > t(366, 0.025)$ maka *tolak H_0* . Oleh karena itu, hal ini menunjukkan adanya Harga Rumah(y) yang tidak dapat dijelaskan oleh besarnya Luas Rumah(x) pada taraf nyata 5%.

2.Dugaan Parameter β_1

Dengan nilai slope(b_1) sebesar 221.7485796 ,selanjutnya akan dilakukan uji apakah luas rumah (x) memiliki hubungan linear atau tidak terhadap harga rumah(y).

Hipotesis uji:

$H_0: \beta_1 = 0$ (Harga Rumah tidak memiliki hubungan linear dengan Luas Rumah)

$H_1: \beta_1 \neq 0$ (Harga Rumah memiliki hubungan linear dengan Luas Rumah)

```
(se_b0 <- sqrt(KTG*(1/368+mean(x)^2/sum((x-mean(x))^2))))
```

```
(t_b0 <- beta0/se_b0)
```

$t_{hitung} = 2.754312$

$$t_{(db, \alpha/2)} = 2.036933343$$

Karena $t_{hitung} > t(366, 0.025)$ maka *tolak H_0* . Oleh karena itu, hal ini menunjukkan adanya hubungan linier antara Luas Rumah(x) dan Harga Rumah(y). Hal tersebut juga memberikan bukti yang cukup untuk menyatakan bahwa Luas Rumah memengaruhi Harga rumah pada taraf nyata 5%.

Nilai Koefisien Determinasi dan Korelasi

```
(Koeff_det<-1-(anova.model$`Sum Sq`[2]/sum(anova.model$`Sum Sq`)))
```

```
Korelasi <- sqrt(Koeff_det)
Korelasi
```

Nilai R^2 sebesar 0.461975343 menunjukkan bahwa sekitar 46.20% dari variasi harga rumah(y) dapat dijelaskan oleh persamaan regresi yang digunakan. Ini berarti terdapat sebagian besar variabilitas yang tidak dapat dijelaskan atau diprediksi oleh persamaan regresi. Sementara nilai korelasi sebesar 0.6796 menunjukkan hubungan positif yang *cukup kuat* antara harga rumah(y) dan luas rumah(x).

Selang Kepercayaan Parameter

Secara umum, kita dapat memahami bahwa selang kepercayaan untuk β_0 dan β_1 terdiri atas batas bawah dan batas atas yang dapat dituliskan sebagai berikut.

Batas bawah untuk kedua dugaan parameter dalam taraf kepercayaan $\alpha=0.05$ dapat dihitung sebagai berikut.

```
#Batas Bawah beta0
(bb.b0 <- beta0 - abs(qt(0.025, df=368-2))*se_b0)
```

```
#Batas Bawah beta1
(bb.b1 <- beta1 - abs(qt(0.025, df=368-2))*se_b1)
```

Sementara itu, batas atas untuk kedua dugaan parameter dalam taraf kepercayaan $\alpha=0.05$ dapat dihitung sebagai berikut.

```
#Batas Atas beta_0
(ba.b0 <- beta0 + abs(qt(0.025, df=368-2))*se_b0)
```

```
#Batas Atas beta_1
(ba.b1 <- beta1 + abs(qt(0.025, df=368-2))*se_b1)
```

Sehingga dapat disusun suatu selang kepercayaan untuk β_0 sebagai berikut.

$$23139.35 < \beta_0 < 138651.3$$

Yang dapat dimaknai bahwa dalam taraf kepercayaan 95%, diyakini bahwa dugaan parameter β_0 berada dalam selang 23139.35 hingga 138651.3.

Sementara untuk selang kepercayaan β_1 adalah sebagai berikut

$$197.1506 < \beta_1 < 246.3465$$

Yang dapat dimaknai bahwa dalam taraf kepercayaan 95%, diyakini bahwa dugaan parameter β_1 berada dalam selang 197.1506 hingga 246.3465.

Selang Kepercayaan Rataan (Nilai Harapan) Amatan

Secara umum, kita dapat memahami bahwa selang kepercayaan untuk $E(Y|x_0)$ terdiri atas batas bawah dan batas atas yang dapat dituliskan sebagai berikut.

Misalkan kita ingin menduga nilai rataan (harapan) amatan ketika nilai $x=3000$. Kita dapat memanfaatkan fungsi predict untuk mendapatkan nilai dugaan dan selang kepercayaannya sebagai berikut.

```
amatan.diduga <- data.frame(x=3000)
predict(model_1, amatan.diduga, interval = "confidence")
```

Berdasarkan output sintaks di atas, diketahui bahwa rataan nilai \hat{y} ketika nilai $x=3000$ adalah 746141.1. Selain itu, diindikasikan juga bahwa dalam taraf kepercayaan 95%, diyakini bahwa nilai rataan \hat{y} ketika nilai $x=3000$ berada dalam selang 712723.4 hingga 779558.7.

Selang Kepercayaan Individu Amatan

Secara umum, kita dapat memahami bahwa selang kepercayaan untuk $\hat{y}(x_i)$ terdiri atas batas bawah dan batas atas yang dapat dituliskan sebagai berikut.

Misalkan kita ingin menduga nilai individu amatan ketika nilai $x=3000$. Kita dapat memanfaatkan fungsi predict untuk mendapatkan nilai dugaan dan selang kepercayaannya sebagai berikut.

```
predict(model_1, amatan.diduga, interval = "prediction")
```

Berdasarkan output sintaks di atas, diketahui bahwa nilai amatan individu $\hat{y}(x_i)$ ketika nilai $x=3000$ adalah 746141.1. Selain itu, diindikasikan juga bahwa dalam taraf kepercayaan 95%, diyakini bahwa nilai amatan individu $\hat{y}(x_i)$ ketika nilai $x=3000$ berada dalam selang 260139.6 hingga 1232142.

Tabel Sidik Ragam

```
dbr <- 1
dbg <- 368-2
```

```
dbt <- 368-1
JKR <- 1.91048E+13
JKG <- 2.22497E+13
JKT <- JKR+JKG
KTR <- JKR/dbr

SK <- c("Regresi", "Galat", "Total")
db <- c(dbr, dbg, dbt)
JK <- c(JKR, JKG, JKT)
KT <- c(KTR, KTG, NA)
data_frame <- data.frame(SK, db, JK, KT)
data_frame
```

Kesimpulan

Berdasarkan analisis yang kami lakukan, terdapat hubungan positif antara harga rumah dan luas rumah, yang mengindikasikan bahwa semakin besar luas suatu rumah, maka kemungkinan tinggi juga semakin naik harga rumah yang ditawarkan. Namun, faktor luas rumah tidak sepenuhnya menjadi faktor utama untuk menentukan harga rumah. Faktor-faktor lain seperti tahun pembuatan, banyaknya ruangan, dan banyak faktor lainnya.