

Tugas Kelompok Analisis Regresi (Regresi Linear Sederhana)

Hubungan antara Jumlah *Profile Visits* dan Jumlah *Likes* pada Media Sosial *Instagram*

Kelompok 6 (P1)

1. Anindya Septyani G1401221074
2. Moch Rizam G1401221028
3. Rose Dwi Aulia Amaradhani G1401221002

Input Data

```
library(readxl)
data <- read_excel("C:/Users/Lenovo/Downloads/Instagram-data-fixed.xlsx")
View(data)
```

Menentukan X dan Y

```
y <- data$Likes
x <- data$`Profile Visits`

data <- data.frame(cbind(y,x))
head(data)

##      y  x
## 1 162 35
## 2 224 48
## 3 131 62
## 4 213 23
## 5 123  8
## 6 144  9

n <- nrow(data)
n

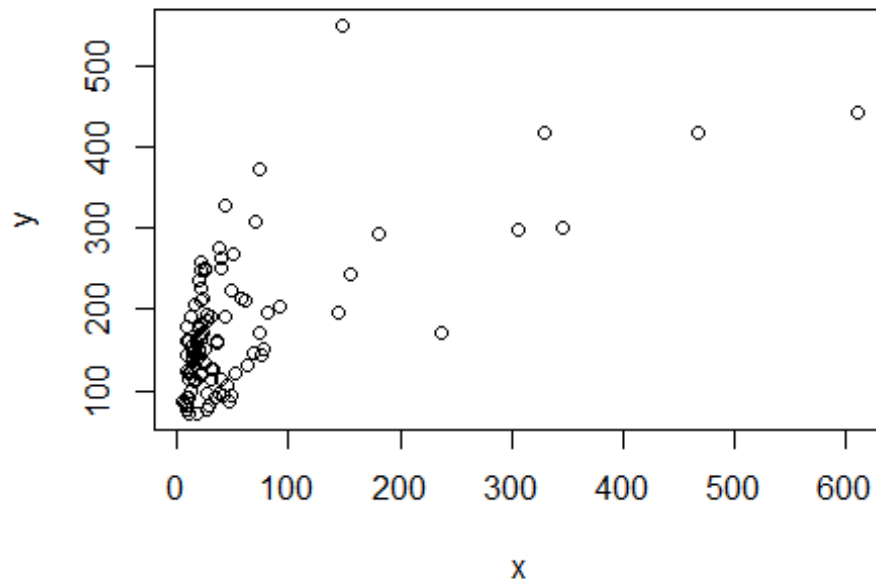
## [1] 119

p <- ncol(data)
p

## [1] 2
```

Analisis Data

```
y.bar <- mean(y)
plot(x,y)
```



X : Profile Visits

Y: Jumlah Suka (*Likes*)

Berikut adalah *scatter plot* hasil sebaran data Instagram. Dapat terlihat dalam *scatter plot* tersebut bahwa *profile visits* berhubungan linear dengan jumlah suka (*likes*) sehingga data tersebut dapat dianalisis menggunakan analisis regresi linear sederhana.

Pembuatan Model

```
model <- lm(y~x, data)
summary(model)
```

```
##
## Call:
## lm(formula = y ~ x, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -112.16  -40.28   -6.12   27.30   317.55
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 143.80076    6.84717  21.001  < 2e-16 ***
## x           0.59225     0.06819   8.685 2.65e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 64.51 on 117 degrees of freedom
## Multiple R-squared:  0.392, Adjusted R-squared:  0.3868
## F-statistic: 75.44 on 1 and 117 DF,  p-value: 2.647e-14

(anova.model <- anova(model))

## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x           1 313915   313915   75.437 2.647e-14 ***
## Residuals 117 486868     4161
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

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

## [1] 4161.262

(galat.baku <- sqrt(KTG))

## [1] 64.50784
```

Persamaan Regresi Linear Sederhana

```
(b1 <- model$coefficients[[2]])

## [1] 0.5922492

(b0 <- model$coefficients[[1]])

## [1] 143.8008
```

Persamaan dari regresi linear sederhana adalah sebagai berikut:

$$\hat{Y} = 143.8008 + 0.5922492X_1$$

Korelasi dan Koefisien Determinasi

```
r<-(sum(x*y)-sum(x)*sum(y)/119)/
  sqrt((sum(x^2)-(sum(x)^2/119))*(sum(y^2)-(sum(y)^2/119)))
r

## [1] 0.626107

Koef_det<-r^2
Koef_det

## [1] 0.39201
```

Kasus ini memiliki korelasi sebesar 0.626107 yang memiliki makna jumlah *Profile Visits* dan jumlah suka (*likes*) memiliki hubungan positif yang cukup besar, dengan koefisien keragaman sebesar 0.39201 yang menjelaskan keragaman pada jumlah *Profile Visits* sebesar 39.201%.

Uji Hipotesis

```
ttabel <- qt(0.025, df= 117, lower.tail = FALSE)
ttabel

## [1] 1.980448
```

Dengan nilai intersep sebesar 143,8008, akan diuji apakah ada jumlah *likes* yang tidak dapat dijelaskan oleh banyaknya jumlah *profile visits* atau tidak.

Hipotesis Uji:

$H_0: \beta_0 = 0$ (Jumlah *likes* dapat dijelaskan oleh banyaknya *profile visits*)

$H_1: \beta_0 \neq 0$ (Ada jumlah *likes* yang tidak dapat dijelaskan oleh banyaknya *profile visits*)

Statistik Uji:

```
(b0 <- model$coefficients[[1]])
## [1] 143.8008

(se_b0 <- sqrt(KTG*(1/n+mean(x)^2/sum((x-mean(x))^2))))
## [1] 6.847168

(t_b0 <- b0/se_b0)
## [1] 21.0015
```

Kaidah Keputusan:

Dapat dilihat bahwa t-hitung sebesar 21,0015 dan t-tabel sebesar 1,980448. Oleh Karena itu, t-hitung > t-tabel maka tolak H_0 yang berarti terdapat cukup bukti untuk menyatakan bahwa ada jumlah *likes* yang tidak dapat dijelaskan oleh banyaknya jumlah *profile visits* pada taraf nyata 5%.

Dengan nilai slope sebesar 0,5922492, selanjutnya akan diuji apakah *profile visits* (x) memengaruhi jumlah *likes* (y) atau tidak.

Hipotesis Uji:

$H_0: \beta_1 = 0$ (Tidak ada hubungan linier antara jumlah *profile visits* dan jumlah *likes*)

$H_1: \beta_1 \neq 0$ (Ada hubungan linier antara jumlah *profile visits* dan jumlah *likes*)

Statistik Uji:

```
(b1 <- model$coefficients[[2]])  
## [1] 0.5922492  
  
(se_b1 <- sqrt(KTG/sum((x-mean(x))^2)))  
## [1] 0.06818851  
  
(t_b1 <- b1/se_b1)  
## [1] 8.68547
```

Kaidah Keputusan:

Dapat dilihat bahwa t-hitung sebesar 8,68547 dan t-tabel sebesar 1,980448. Oleh Karena itu, $t\text{-hitung} > t\text{-tabel}$ maka tolak H_0 yang berarti antara jumlah *profile visits* dan jumlah *likes* terdapat hubungan linear serta cukup bukti untuk menyatakan bahwa banyaknya jumlah *profile visits* memengaruhi jumlah *likes* pada taraf nyata 5%.

Penguraian Keragaman

```
JKG <- sum((y-(b0+b1*x))^2)  
JKG  
## [1] 486867.6
```

JKG atau Jumlah Kuadrat Galat mencerminkan jumlah kesalahan kuadrat antara nilai amatan sebenarnya dan nilai yang diprediksi oleh model regresi. Dalam hal ini, jumlah kesalahan tersebut adalah 486867,6.

```
JKReg <- sum(((b0+b1*x) - mean(y))^2)  
JKReg  
## [1] 313914.7
```

JKReg atau Jumlah Kuadrat Regresi mencerminkan jumlah varians dari variabel dependen (y) yang dapat dijelaskan oleh model regresi. Dengan nilai sebesar 313914,7, ini menunjukkan sejauh mana variasi jumlah *likes* yang dapat dijelaskan oleh variabel *profile visits* dalam model.

```
JKT <- JKReg+JKG  
JKT  
## [1] 800782.3
```

JKT atau Jumlah Kuadrat Total mencerminkan total variasi dalam variabel respons. Dalam hal ini, didapatkan JKT sebesar 800782,3 yang bermaksud menggambarkan tentang seberapa besar variasi yang ada dalam data respons (jumlah *likes*).

```
dbReg <- 1
dbReg

## [1] 1

dbg <- n-2
dbg

## [1] 117

dbt <- n-1
dbt

## [1] 118
```

Derajat bebas Regresi atau dbReg = 1 menunjukkan bahwa model regresi yang digunakan hanya melibatkan satu variabel independen (*Profile Visits*).

Derajat bebas Galat atau dbg = 117, ini artinya ada 117 derajat kebebasan yang terkait dengan kesalahan atau residu dalam model regresi.

Derajat bebas total atau dbt = 118, ini berarti terdapat 118 derajat kebebasan yang terkait dengan total data dalam model regresi.

```
Fhit <- (JKReg/dbReg)/(JKG/dbg)
Fhit

## [1] 75.43738
```

F-hitung sebesar 75,43738 mengindikasikan sebagai ukuran keberagaman variabilitas antara kelompok atau model dengan variabilitas dalam kelompok.

```
P.value <- 1-pf(Fhit,dbReg,dbg,lower.tail<-F)
P.value

## [1] 2.642331e-14
```

Nilai p yang sangat kecil yaitu sebesar 2,642331e-14 menunjukkan bahwa terdapat bukti yang sangat kuat untuk menolak hipotesis nol.

Selang Kepercayaan Parameter

```
#Batas bawah b0
(bb.b0 <- b0 - abs(qt(0.025, df=117))*se_b0)

## [1] 130.2403

#Batas atas b0
(bb.b0 <- b0 + abs(qt(0.025, df=117))*se_b0)

## [1] 157.3612
```

Pada taraf nyata sebesar 5%, dapat diyakini bahwa nilai parameter b_0 berada dalam selang 130,2403 hingga 157,3612.

$$130.2403 < b_0 < 157.3612$$

```
#Batas Bawah b1
(bb.b1 <- b1 - abs(qt(0.025, df=117))*se_b1)

## [1] 0.4572055

#Batas Atas b1
(ba.b1 <- b1 + abs(qt(0.025, df=117))*se_b1)

## [1] 0.727293
```

Pada taraf nyata sebesar 5% dapat diyakini bahwa nilai parameter b_1 berada dalam selang 0,4572055 hingga 0,72793.

$$0.4572055 < b_1 < 0.727293$$