Tugas Kelompok 8 Analisis Regresi

Tubagus Achmad Aditya - G1401221006, Qonita Husnia Rahmah - G1401221008, Sintong M.N Purba -

2024-02-10

```
library(rmarkdown)
```

```
## Warning: package 'rmarkdown' was built under R version 4.3.2
```

Membaca data csv

```
data_tugas <- read.csv("C:/Users/Poerba/Downloads/Cities1.csv", sep = "," )
head(data_tugas)
```

```
##               City               Region                  Country AirQuality
## 1    New York City             New York United States of America   46.81604
## 2 Washington, D.C. District of Columbia United States of America   66.12903
## 3    San Francisco           California United States of America   60.51402
## 4           Berlin                                      Germany   62.36413
## 5      Los Angeles           California United States of America   36.62162
## 6             Bern       Canton of Bern              Switzerland   94.31818
##   WaterPollution
## 1       49.50495
## 2       49.10714
## 3       43.00000
## 4       28.61272
## 5       61.29944
## 6       12.50000
```

Pendefinisian peubah yang digunakan

Peubah yang digunakan adalah kualitas udara (*air Quality*) sebagai peubah penjelas X dan polusi air (*water pollution*) sebagai peubah respon Y, sehingga hubungan antara keduanya dapat dinyatakan dalam sebuah persamaan garis linear

```
X<- data_tugas$AirQuality
Y<- data_tugas$WaterPollution
```

Pembentukan model regresi secara manual

Parameter regresi

```
n<-nrow(data_tugas)
n
```

```
## [1] 3963
```

```r
b1<-(sum(X*Y)-(sum(X)*sum(Y)/n))/(sum(X^2)-(sum(X)^2/n))
b0<-mean(Y)-b1*mean(X)
b1
```

```
## [1] -0.3766663
```

```r
b0
```

```
## [1] 68.08415
```

Maka persamaan garis regresinya adalah y = 68.08415 -0.3766663 x . yang artinya jika kualitas udara meningkat 1 poin maka dugaan nilai polusi air akan turun sebesar b1 yaitu 0.3766663, dan saat kualitas udara bernilai 0 maka besar dugaan nilai polusi airnya adalah sebesar b0 yaitu 68.08415

Korelasi dan Koefisien determinasi

```r
r<-(sum(X*Y)-sum(X)*sum(Y)/n)/sqrt((sum(X^2)-(sum(X)^2/n))*(sum(Y^2)-(sum(Y)^2/n)))
r
```

```
## [1] -0.4541726
```

```r
Koef_det<-r^2
Koef_det
```

```
## [1] 0.2062728
```

Didapatkan korelasi sebesar -0.4541726 yang artinya kualitas udara dan polusi air memiliki hubungan negatif yang cukup kecil, sedangkan koefisien keragamannya sebesar 0.2062728 yang menunjukkan bahwa kualitas udara hanya mampu menjelaskan keragaman pada nilai polusi air sebesar 0.2062728 atau 20.62728 %.

Uji hipotesis parameter regresi

Standar eror parameter regresi

```r
galat<-Y-(b0+b1*X)
ragam_galat<-sum(galat^2)/(n-2)

se_b1<-sqrt(ragam_galat/sum((X-mean(X))^2))
se_b1
```

```
## [1] 0.01174002
```

```r
se_b0<-sqrt(ragam_galat*(1/n+mean(X)^2/sum((X-mean(X))^2)))
se_b0
```

```
## [1] 0.8161495
```

Hipotesis

H0: b1=0 (tidak ada hubungan linear antara kualitas udara dan polusi air) H1: b1!=0 (terdapat hubungan linear antara kualitas udara dan polusi air)

dan

H0: b0=0 (Semua nilai polusi air dapat dijelaskan oleh kualitas udara) H1: b0!=0 (terdapat nilai polusi air yang tidak dapat dijelaskan oleh kualitas udara)

Uji t

```
t_b1<-b1/se_b1
t_b1
```

```
## [1] -32.08394
```

```
t_b0<-b0/se_b0
t_b0
```

```
## [1] 83.42118
```

```
qt(0.025, df = n-2, lower.tail = FALSE)
```

```
## [1] 1.960563
```

Untuk b1 : karena |t-hit(b1)|=|-32.08394|>= t tabel= 1.960563, maka tolak H0 sehingga cukup bukti untuk menyatakan terdapat hubungan linear antara kualitas udara dan polusi air.

Untuk b0 : karena |t-hit(b0)|=|83.42118|>= t tabel= 1.960563, maka tolak H0 sehingga cukup bukti untuk menyatakan terdapat nilai polusi air yang tidak dapat dijelaskan oleh kualitas udara.

Ukuran keragaman

```
galat<-Y-(b0+b1*X)

JKG <- sum((Y - (b0+b1*X))^2)
JKG
```

```
## [1] 2071245
```

```
JKReg <- sum(((b0+b1*X)- mean(Y))^2)
JKReg
```

```
## [1] 538272.3
```

```
JKT <- sum((Y - mean(Y))^2)
JKT
```

```
## [1] 2609517
```

```
JKT2 <- JKReg+JKG
JKT2
```

```
## [1] 2609517
```

```
dbReg<-1
dbg<-n-2
dbt<-n-1

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

```
## [1] 1029.379
```

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

```
## [1] 0
```

Pembentukan model regresi menggunakan fungsi lm

model regresi juga dapat dibentuk secara langsung menggunakan fungsi lm

```
model<-lm(Y~X,data_tugas<-data_tugas)
summary(model)
```

```
##
## Call:
## lm(formula = Y ~ X, data = data_tugas <- data_tugas)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -64.317 -17.525   0.749  15.812  69.582
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 68.08415    0.81615   83.42   <2e-16 ***
## X           -0.37667    0.01174  -32.08   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.87 on 3961 degrees of freedom
## Multiple R-squared:  0.2063, Adjusted R-squared:  0.2061
## F-statistic:  1029 on 1 and 3961 DF,  p-value: < 2.2e-16
```

Anova dari model

```
anova(model)
```

```
## Analysis of Variance Table
##
## Response: Y
##             Df  Sum Sq Mean Sq F value    Pr(>F)
## X            1  538272  538272  1029.4 < 2.2e-16 ***
## Residuals 3961 2071245     523
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Penentuan selang kepercayaan parameter model regresi

Selang kepercayaan b0

```
#Batas bawah b0, batas atas b0
Sk_b0<-c(b0 - abs(qt(0.025, df=n-2))*se_b0,b0 + abs(qt(0.025, df=n-2))*se_b0)
Sk_b0
```

```
## [1] 66.48404 69.68426
```

Maka nilai b0 pada taraf kepercayaan 0.05 akan jatuh pada selang [66.48404 , 69.68426]

Selang kepercayaan b1

```
#Batas bawah b1, batas atas b1
Sk_b1<-c(b1 - abs(qt(0.025, df=n-2))*se_b1,b1 + abs(qt(0.025, df=n-2))*se_b1)
```

```
Sk_b1
```

```
## [1] -0.3996833 -0.3536492
```

Maka nilai b1 pada taraf kepercayaan 0.05 akan jatuh pada selang [-0.3996833 , -0.3536492]