

Tugas Kuliah Analisis Regresi

Hesty_G1401221019, Muhammad Firlan Maulana_G1401221042, Laras Suprpti_G1401221044

2024-02-07

Read Data

```
Data_Sleep<- read.csv("D:/firlan/Documents/College/Semester 4/Analisis Regeresi/Kuliah 3/Sleep Health and Lifestyle Dataset.csv")
y<-Data_Sleep$QualityofSleep
x<-Data_Sleep$StressLevel

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

```
##    y x
##  1 6 6
##  2 6 8
##  3 6 8
##  4 4 8
##  5 4 8
##  6 4 8
```

```
n<-nrow(Data_Sleep)
p<-ncol(Data_Sleep)
```

Deskripsi Data

Dataset berjudul Sleep Health and Lifestyle Dataset dari laman web kaggle. Seperti judulnya, data ini berisikan hubungan gaya hidup dengan kualitas tidur. Data diupload di kaggle oleh Laksika Tharmalingam selaku pemilik dataset. Di dalam dataset terdapat beberapa variabel seperti durasi tidur dalam satuan jam, kualitas tidur dalam skala 1 sampai 10, tingkatan stress dalam skala 1 sampai 10, tekanan darah, usia, jenis kelamin, dan lain sebagainya. Dari variabel-variabel tersebut kelompok kami memilih variabel kualitas tidur dan tingkatan stress untuk dilihat hubungannya.

Persamaan Regresi Linear Sederhana (menggunakan fungsi lm)

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

```
##
## Call:
## lm(formula = y ~ x, data = Data_Sleep <- Data_Sleep)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7276 -0.1525  0.2413  0.2724  1.0600
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.57739    0.08693   121.68  <2e-16 ***
## x           -0.60623    0.01533   -39.53  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5255 on 372 degrees of freedom
## Multiple R-squared:  0.8078, Adjusted R-squared:  0.8072
## F-statistic: 1563 on 1 and 372 DF, p-value: < 2.2e-16
```

Dari perhitungan tersebut, didapati dugaan persamaan regresi linear sederhana atau nilai harapan dari peubah respon Y adalah sebagai berikut. $E[Y]^{\wedge} = Y^{\wedge} = 10.57739 - 0.60623X$

Berikut interpretasinya. 1. Rata-rata skala kualitas tidur jika tingkatan stress bernilai 0 adalah berada di skala 10.57739. Namun dalam dataset ini angka 0 tidak terdapat dalam selang variabel X sehingga interpretasi ini tidak praktis karena angka 0 tidak terdapat pada skala tingkatan stress 1 sampai 10. 2. Setiap peningkatan satu skala tingkatan stress akan mengurangi skala kualitas tidur sebesar 0.60623.

Ringkasan Umum variabel StressLevel(X) dan QualityofSleep(Y)

```
summary(x)
```

```
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   3.000   4.000   5.000   5.385   7.000   8.000
```

```
summary(y)
```

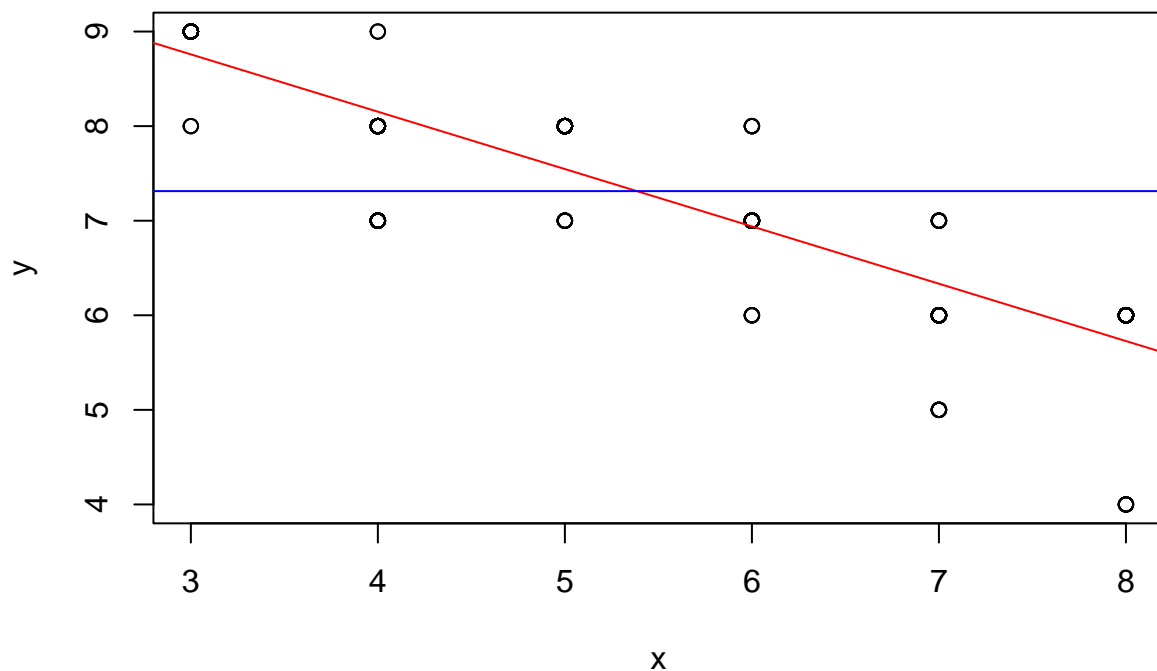
```
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   4.000   6.000   7.000   7.313   8.000   9.000
```

Diketahui nilai minimum dan maksimum dari masing-masing variabel. Diketahui pula nilai dari kuartil 1, median, dan kuartil 3 sekaligus rata-rata dari variabel x dan y.

Scatter Plot

Dalam hal ini kami ingin melihat hubungan antara tingkatan stress dan kualitas tidur sehingga tingkatan stress menjadi sumbu-x dan kualitas tidur sebagai sumbu-y. Kami menduga bahwa semakin tinggi tingkatan stress seseorang akan mengakibatkan penurunan kualitas tidur.

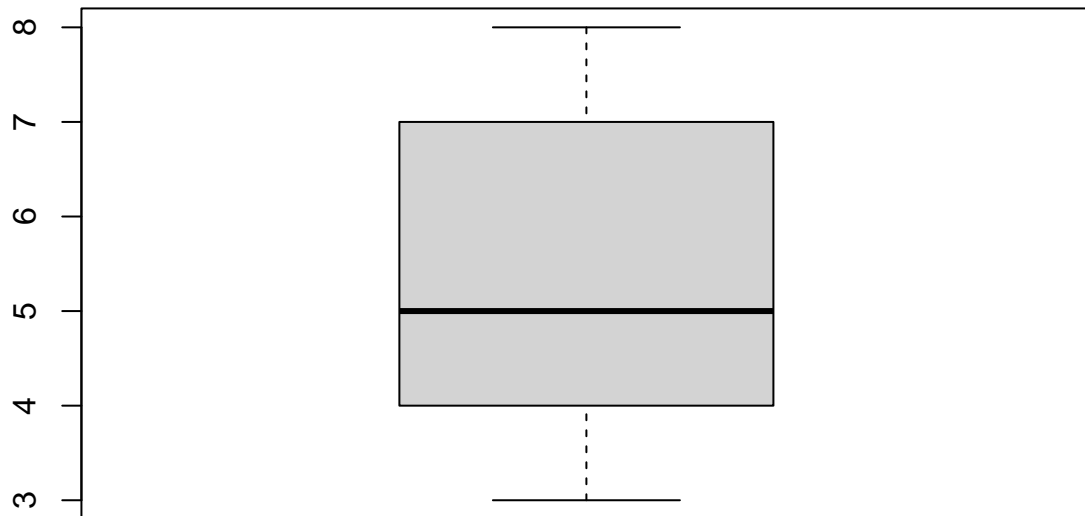
```
y.bar <- mean(y)
plot(x,y)
abline(model, col="red")
text(30, 220, "Y_duga", adj = c(-0.1, 1.5), col = "red", cex = 0.8)
abline(h=y.bar, col="blue")
text(31, 185, "Y_bar", adj = c(-0.1, 1.5), col = "blue", cex = 0.8)
```



Ditunjukkan dua buah garis yakni garis biru dan garis merah dari scatter plot tersebut. Garis biru merupakan garis \bar{Y} atau garis dari rata-rata skala kualitas tidur, sedangkan garis merah adalah garis persamaan regresi. Dapat dilihat bahwa garis merah (garis persamaan regresi linear) mengarah ke bawah kanan sesuai dengan nilai slope (b_1) yang negatif.

Boxplot StressLevel (tingkatan stress)

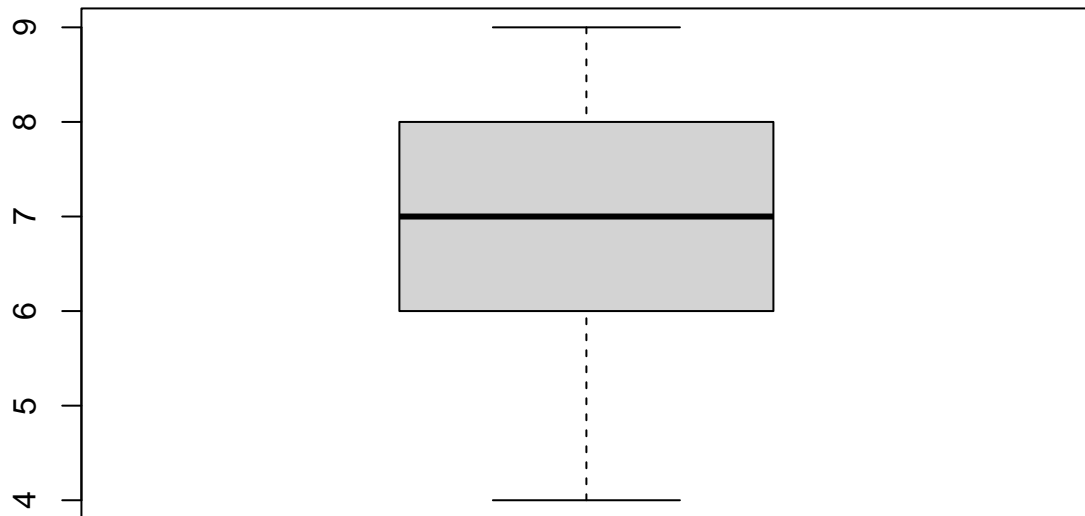
```
boxplot(x)
```



Dari boxplot diketahui bahwa data tingkatan stress relatif menyebar dengan normal, tampak dari jambang boxplot yang sama panjangnya antara kanan dan kiri. Tampak pula bahwa tidak terdapat pencilan dalam data tersebut.

Boxplot QualityofSleep (kualitas tidur)

```
boxplot(y)
```



Dari boxplot tampak bahwa sebaran data dari variabel kualitas tidur relatif menjulur ke kiri dan tidak memiliki pencilan.

KOEFISIEN DETERMINASI DAN PENYESUAIANNYA

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

```
## [1] 0.8077552
```

```
r
```

```
## [1] -0.898752
```

didapatkan bahwa koefisien determinasi sebesar 0.8077552 dan -0.898752 yang menunjukkan bahwa hubungan antara tingkatan stress terhadap kualitas tidur yang sangat baik.

UKURAN KERAGAMAN

```
b1<-(sum(x*y)-sum(x)*sum(y)/n)/(sum(x^2)-(sum(x)^2/n))
b1
```

```
## [1] -0.6062274
```

```
b0<-mean(y)-b1*mean(x)
b0
```

```
## [1] 10.57739
```

```
galat<-y-(b0+b1*x)
```

```
JKG <- sum((y - (b0+b1*x))^2)
JKReg <- sum(((b0+b1*x)- mean(y))^2)
JKT <- sum((y - mean(y))^2)
JKT <- JKReg+JKG
```

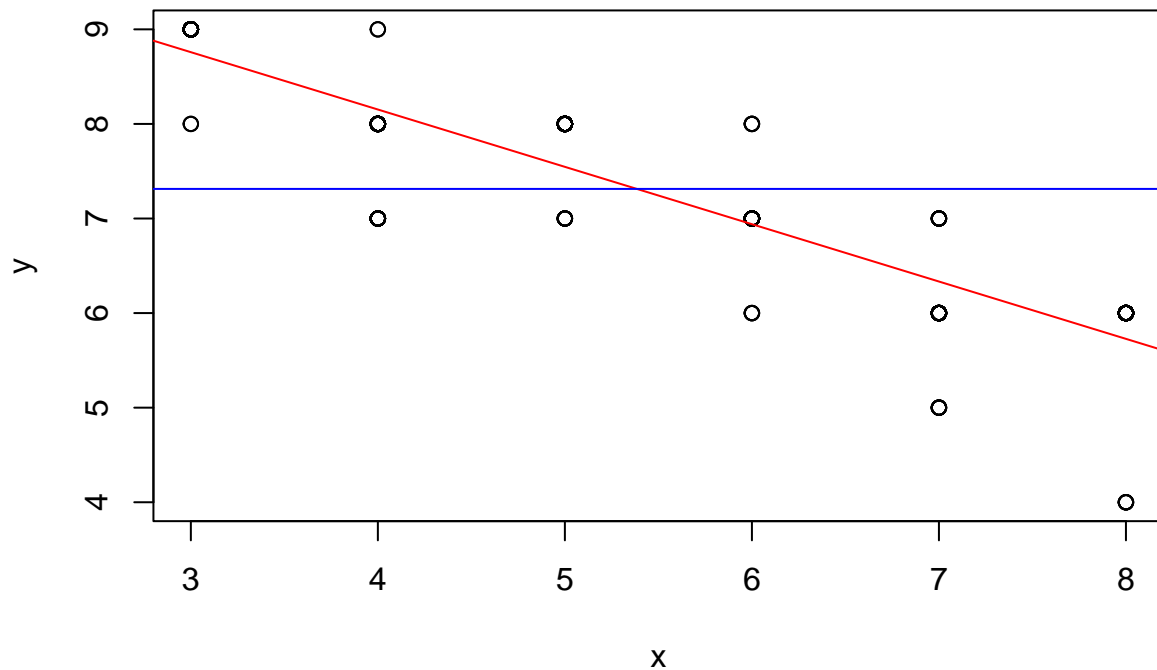
```
dbReg<-1
dbg<-n-2
dbt<-n-1
```

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

```
## [1] 1563.033
```

Penguraian Keragaman

```
y.bar <- mean(y)
plot(x,y)
abline(model, col="red")
text(30, 220, "Y_duga", adj = c(-0.1, 1.5), col = "red", cex = 0.8)
abline(h=y.bar, col="blue")
text(31, 185, "Y_bar", adj = c(-0.1, 1.5), col = "blue", cex = 0.8)
```



Dari scatter plot di atas, teramati adanya amatan yang menyimpang relatif terhadap nilai harapannya. Penyimpangan amatan dari nilai harapannya inilah yang disebut sebagai galat. Keragaman dari galat tiap amatan ini dapat diurai berdasarkan garis dugaan persamaan (\hat{Y}) dan garis rata-rata nilai resposnya (\bar{Y}). Penguraian keragaman ini dapat dituliskan sebagai Jumlah Kuadrat Regresi (JKR), Jumlah Kuadrat Galat (JKG), dan Jumlah Kuadrat Total (JKT). Nilai penguraian keragaman ini dapat diamati pula menggunakan fungsi anova pada bagian Sum Sq (Jumlah Kuadrat) seperti yang ditampilkan di bawah. Pada output sintaks di bawah, JKR model bernilai 431.66 dan JKG model bernilai 102.74.

```
anova(model)
```

```
## Analysis of Variance Table
##
## Response: y
##          Df Sum Sq Mean Sq F value    Pr(>F)
## x           1  431.66   431.66    1563 < 2.2e-16 ***
## Residuals 372  102.74     0.28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Dari tabel ANOVA dan penguraian keragaman, dapat diduga pula ragam galat/error dari model yang kita peroleh. Secara umum, error dari model dapat diduga oleh nilai Kuadrat Tengah Galat (KTG) sehingga berdasarkan output sintaks di atas, diketahui bahwa ragam galat dari model bernilai 0,28. Hubungan antara keduanya dapat dituliskan sebagai berikut.

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

```
## [1] 0.2761702
```

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

```
## [1] 0.525519
```

Dari nilai dugaan ragam galat di atas, dapat diduga pula nilai dugaan dari simpangan bakunya. Dugaan simpangan baku inilah yang disebut sebagai galat baku. Oleh karena simpangan baku merupakan akar kuadrat dari ragam, maka nilai dugaan galat baku model yang kita bentuk adalah $se = 0.525519$

Keragaman Dugaan Parameter

Dugaan Paramater beta0

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

```
## [1] 10.57739
```

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

```
## [1] 0.08692965
```

```
(t_b0 <- b0/se_b0)
```

```
## [1] 121.6775
```

```
(t_tab <- qt(0.025,anova (model)$Df[2],lower.tail=FALSE))
```

```
## [1] 1.966362
```

Didapatkan bahwa $t_{b0} = 121.6775 > t_{tab} = 1.966362$ maka tolak $H_0(b_0=0)$, artinya ada kualitas tidur yang tidak dapat dijelaskan oleh tingkatan stress.

Dugaan Paramater beta1

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

```
## [1] -0.6062274
```

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

```
## [1] 0.01533386
```



```
(t_b1 <- b1/se_b1)
```

```
## [1] -39.53521
```

Didapatkan bahwa $|t_{b1}| = 39.53521 > t_{\text{tab}} = 1.966362$ maka tolak $H_0(b_1=0)$, artinya berpengaruh (hubungan linear) tingkatan stress terhadap kualitas tidur.

Selang Kepercayaan Parameter

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

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

```
## [1] 10.40645
```

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

```
## [1] -0.6363793
```

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

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

```
## [1] 10.74832
```

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

```
## [1] -0.5760755
```

Sehingga dapat disusun suatu selang kepercayaan untuk β_0^{\wedge} dan β_1^{\wedge} sebagai berikut.

$10.40645 < \beta_0^{\wedge} < 10.74832$

Yang dapat dimaknai bahwa dalam taraf kepercayaan 95%, diyakini bahwa dugaan parameter β_0^{\wedge} berada dalam selang 10.40645 hingga 10.74832.

$-0.6363793 < \beta_1^{\wedge} < -0.5760755$

Yang dapat dimaknai bahwa dalam taraf kepercayaan 95%, diyakini bahwa dugaan parameter β_1^{\wedge} berada dalam selang -0.6363793 hingga -0.5760755.

Selang Kepercayaan Rataan (Nilai Harapan) Amatan

```
amatan.diduga <- data.frame(x=8)
predict(model, amatan.diduga, interval = "confidence")
```

```
##           fit           lwr           upr
## 1 5.727566 5.632319 5.822812
```

Berdasarkan output sintaks di atas, diketahui bahwa dugaan rata-rata nilai Y ketika nilai $x_1=8$ adalah 5.727566. Selain itu, diindikasikan juga bahwa dalam taraf kepercayaan $\alpha=0.05$, diyakini bahwa nilai dugaan rata-rata Y ketika nilai $x_1=8$ berada dalam selang 5.632319 hingga 5.822812.

Selang Kepercayaan Individu Amatan

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

```
##           fit           lwr           upr
## 1 5.727566 4.689825 6.765306
```

Berdasarkan output sintaks di atas, diketahui bahwa dugaan nilai individu Y ketika nilai $x_1=8$ adalah 5.727566. Selain itu, diindikasikan juga bahwa dalam taraf kepercayaan $\alpha=0.05$, diyakini bahwa nilai amatan individu Y ketika nilai $x_1=8$ berada dalam selang 4.689825 hingga 6.765306.