Tugas Kelompok Analisis Regresi Pekan 3

Elke Frida Rahmawati - G1401221025; Bulan Cahyani Suhaeri - G1401221030; Shabrina Shafwah Al-Rahmah - G1401221083

2024-02-09

Analisis Regresi Linier Sederhana
Kadar Glukosa dan Insulin pada Penderita Diabetes

Data

```
diabetes <- read.csv("D:/Shabrina/College/Semester 4/Analisis Regresi/anreg k
el.csv", sep = ";")

y<-diabetes$Glucose..y.
x<-diabetes$Insulin..x.

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

n<-nrow(diabetes)
n

## [1] 130

DT::datatable(diabetes)
```

Show 10 entries

Search:

	y	x
1	137	168
2	78	88
3	197	543
4	189	846
5	166	175
6	118	230
7	115	96
8	143	146
9	125	115
10	158	245

Showing 1 to 10 of 130 entries

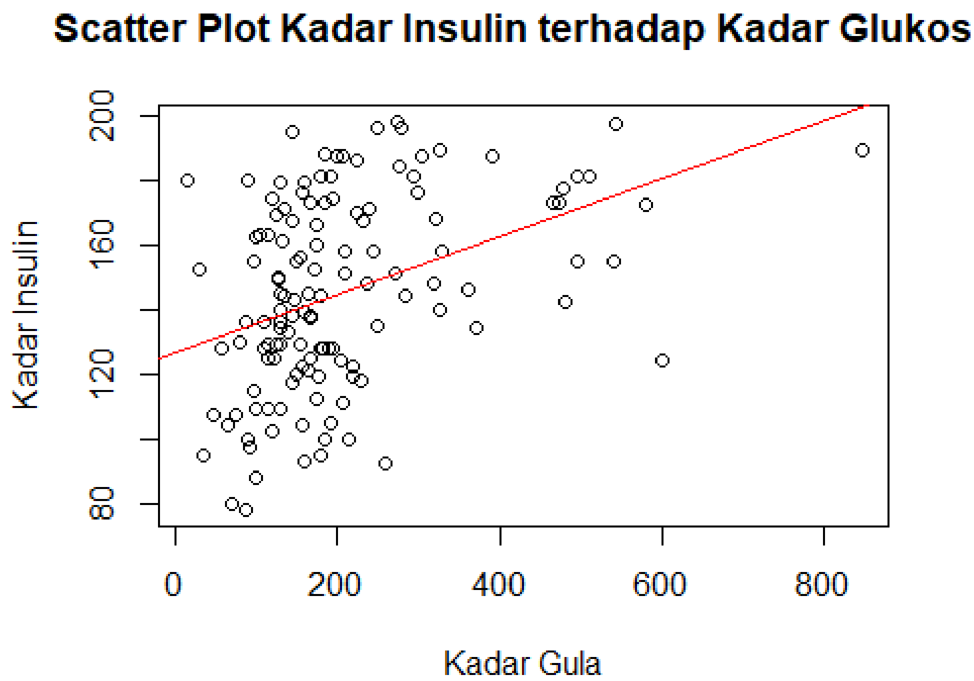
Previous 1 2 3 4 5 … 13 Next

Data yang digunakan berupa kadar insulin sebagai peubah X atau peubah bebas serta kadar glukosa sebagai peubah Y atau peubah tak bebas yang masing-masing berjumlah 130 data.

Eksplorasi Data

```
plot(x,y,
     main = "Scatter Plot Kadar Insulin terhadap Kadar Glukosa",
       xlab = "Kadar Gula",
```

```
        ylab = "Kadar Insulin")
abline(lm(y~x), col="red")
```



Hasil *scatter plot* menggambarkan Y sebagai peubah tak bebas dan X sebagai peubah bebas. Pola titik yang cenderung membentuk garis lurus dari kiri bawah ke kanan atas menunjukkan adanya hubungan positif secara linier antara kedua peubah tersebut. Akan tetapi, hubungan tersebut cenderung lemah karena sebaran titik titik jauh dari garis regresi.

Berikut ringkasan atau statistik lima serangkai dari peubah tak bebas.

```
summary(y)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    78.0   124.2   144.5   145.2   171.8   198.0
```

Berikut ringkasan atau statistik lima serangkai dari peubah bebas.

```
summary(x)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    14.0   127.5   169.5   206.8   239.2   846.0
```

Pendugaan Parameter Regresi

Seperti yang diketahui bahwa dugaan persamaan regresi linier sederhana sebagai berikut.

$$\hat{Y} = \widehat{\beta_0} + \widehat{\beta_1}X$$

1. Dugaan Parameter Regresi $\widehat{\beta_1}$

Parameter regresi $\widehat{\beta_1}$ dapat diduga salah satunya dengan Metode Kuadrat Terkecil sebagai berikut.

$$\hat{\beta}_1 = \frac{\sum(x_i y_i) - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}$$

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

## [1] 0.08950445
```

2. Dugaan Parameter Regresi $\widehat{\beta_0}$

Parameter regresi $\widehat{\beta_0}$ dapat diduga salah satunya dengan Metode Kuadrat Terkecil sebagai berikut.

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

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

## [1] 126.6787
```

3. Dugaan Persamaan Regresi

Hasil perhitungan sebelumnya akan menghasilkan dugaan persamaan regresi sebagai berikut.

$$\hat{Y} = 126.6787 + 0.08950445X$$

Dari dugaan persamaan regresi tersebut, dapat diinterpretasikan bahwa $\widehat{\beta_0}$ merupakan nilai dugaan rataan y (glukosa) pada saat X (insulin) bernilai 0, yakni sebesar 126,6787. Selain itu, $\widehat{\beta_1}$ merupakan nilai dugaan rataan kadar glukosa yang akan meningkat sebesar 0.08950445 setiap kadar insulin bertambah satu satuan. Secara garis besar, dari persamaan tersebut dapat dikatakan bahwa nilai dugaan rataan bagi glukosa adalah 126,6787 dan setiap penambahan satu satuan akan bertambah sebesar 0.08950445.

Koefisien Determinasi dan Penyesuaiannya

```
Sxy <- (sum(x*y)-sum(x)*sum(y)/n)
Sxy

## [1] 203317.8
```

```
Sxx<- sum(x^2)- (sum(x)^2/n)
Sxx

## [1] 2271595

Syy<- sum(y^2)- (sum(y)^2/n)
Syy

## [1] 114860.2

r<- Sxy/ sqrt(Sxx*Syy)
r

## [1] 0.3980387

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] 0.1584348

Adj_R2<-1-((1-Koef_det)*(n-1)/(n-1-1))
Adj_R2

## [1] 0.1518601
```

Dari perhitungan tersebut didapat bahwa nilai korelasi antara glukosa dan insulin bernilai 0.3980387. Angka ini dapat dikatakan cukup rendah yang berarti ada hubungan linear positif yang lemah antara glukosa dan insulin.

R^2 atau koefisien determinasi berguna untuk mengukur proporsi keragaman atau variasi total di sekitar nilai tengah (y) yang dapat dijelaskan oleh garis regresi. Dengan kata lain, koefisien determinasi mengukur kebaikan model regresi. R^2 bernilai 0.1584348 berarti sekitar 15,84% keragaman dalam kadar glukosa dapat dijelaskan oleh model regresi. Angka tersebut menunjukkan bahwa model regresi yang telah dibuat tidak begitu kuat untuk menjelaskan hubungan antara x dan y mengingat nilai koefisien determinasi yang paling baik adalah 1.

Signifikansi Parameter Regresi (uji-t)

1. Uji Hipotesis bagi $\widehat{\beta_1}$

Uji-t terhadap slope/kemiringan ($\widehat{\beta_1}$) dilakukan untuk mengetahui adakah hubungan linier antara peubah Y (glukosa) dan X (insulin) dengan hipotesis sebagai berikut.

$H_0: \beta_1 = 0$ (Tidak ada hubungan linier antara Y dan X)

$H_1: \beta_1 \neq 0$ (Ada hubungan linier antara Y dan X)

```
#Standard Error
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] 0.018233
```

Dengan dugaan β_1 sebesar 0.08950445, terdapat ketidakpastian sekitar 0.018233.

```
#Nilai t-hitung
t_b1<-b1/se_b1
t_b1

## [1] 4.908925

#Nilai t-tabel
dbg<-n-2
(nilai_kritis_t <- qt(0.975, dbg))

## [1] 1.978671
```

Didapat nilai t-hitung > t-tabel, dimana 4.908925 > 1.978671 sehingga Tolak H0. Artinya pada taraf nyata 5%, cukup bukti untuk mengatakan bahwa terdapat hubungan linear antara x dan y (Insulin dan Glukosa).

2. Uji Hipotesis bagi $\widehat{\beta_0}$

Uji-t terhadap intersep ($\widehat{\beta_0}$) dilakukan untuk mengetahui adakah nilai Y (glukosa) yang tidak dapat dijelaskan oleh X (insulin) dengan hipotesis sebagai berikut.

$H_0: \beta_0 = 0$ (Semua nilai Y dapat dijelaskan oleh X)

$H_1: \beta_0 \neq 0$ (Ada nilai Y yang tidak bisa dijelaskan oleh X)

```
#Standard Error
se_b0<-sqrt(ragam_galat*(1/n+mean(x)^2/sum((x-mean(x))^2)))
se_b0

## [1] 4.47579
```

Dengan dugaan β_0 sebesar 126.6787, terdapat ketidakpastian sekitar 4.47579.

```
#Nilai t-hitung
t_b0<-b0/se_b0
t_b0

## [1] 28.30308

#Nilai t-tabel
dbg<-n-2
(nilai_kritis_t <- qt(0.975, dbg))
```

```
## [1] 1.978671
```

Didapat Nilai t-hitung > t-tabel, dimana 28.30308 > 1.978671 sehingga Tolak H0. Artinya pada taraf nyata 5%, cukup bukti untuk menyatakan bahwa ada nilai Y yang tidak dapat dijelaskan oleh X.

Pembentukan Model dengan Fungsi lm

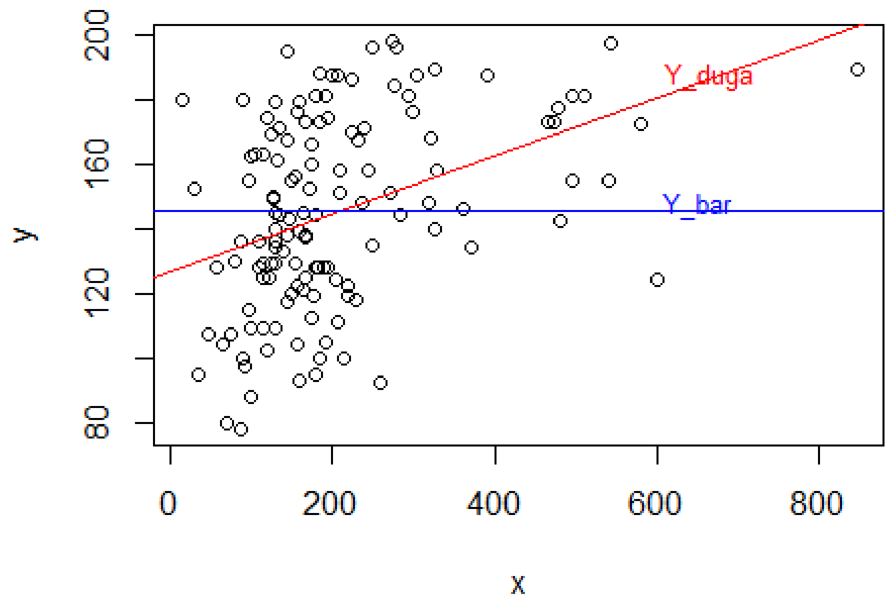
Model regresi dapat dibuat dengan fungsi lm sebagai berikut.

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

##
## Call:
## lm(formula = y ~ x, data = data <- diabetes)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -57.771 -20.229  -1.046  22.812  55.343
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 126.67866    4.47579  28.303  < 2e-16 ***
## x             0.08950    0.01823   4.909 2.73e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27.48 on 128 degrees of freedom
## Multiple R-squared:  0.1584, Adjusted R-squared:  0.1519
## F-statistic:  24.1 on 1 and 128 DF,  p-value: 2.735e-06
```

Penguraian Keragaman

```
y.bar <- mean(y)
plot(x,y)
abline(model, col="red")
text(600, 195, "Y_duga", adj = c(-0.1, 1.5), col = "red", cex = 0.8)
abline(h=y.bar, col="blue")
text(600, 155, "Y_bar", adj = c(-0.1, 1.5), col = "blue", cex = 0.8)
```



Pada *scatter plot* tersebut, dapat dilihat ada beberapa data yang menyimpang relatif terhadap nilai harapannya. Penyimpangan tersebut yang disebut galat. Keragaman galat dapat diperoleh dengan menguraikan persamaan $\hat{y}$ dan garis rataan $\bar{y}$. Penguraian ini yang sering disebut sebagai Jumlah Kuadrat Regresi (JKR), Jumlah Kuadrat Galat (JKG), dan Jumlah Kuadrat Total (JKT). Secara umum, error pada model anova dapat diduga dengan Kuadrat Tengah Galat (KTG)

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

## Analysis of Variance Table
##
## Response: y
##            Df Sum Sq Mean Sq F value    Pr(>F)
## x           1  18198 18197.9  24.098 2.735e-06 ***
## Residuals 128  96662   755.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$\widehat{\sigma^2} = s_e^2 = \mathrm{KTG} = \frac{\mathrm{JKG}}{n-2}$$

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

## [1] 755.1745
```

Nilai simpangan baku juga dapat dicari dengan nilai dugaan ragamnya. Dugaan simpangan baku ini yang sering disebut sebagai galat baku. Galat baku dapat dituliskan sebagai berikut.

$$s_e = \sqrt{s_e^2}$$

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

## [1] 27.48044
```

Tabel Sidik Ragam

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

(JKG <- sum((y - (b0+b1*x))^2))

## [1] 96662.34

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

## [1] 18197.85

JKT <- sum((y - mean(y))^2)
(JKT <- JKReg+JKG)

## [1] 114860.2

(dbReg<-1)

## [1] 1

(dbg<-n-2)

## [1] 128

(dbt<-n-1)

## [1] 129

(KTReg<-JKReg/dbReg)

## [1] 18197.85

(KTT<-JKT/dbt)

## [1] 890.3891

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

## [1] 24.09754
```

Dari perhitungan di atas, didapat tabel sidik ragam sebagai berikut.

Sumber Keragaman (SK)	Derajat Bebas (db)	Jumlah Kuadrat (JK)	Kuadrat Tengah (KT)	F-hitung
Regresi	1	18197.85	18197.85	24.09754
Galat	128	96662.34	755.1745	
Total	129	114860.2	890.09754	

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

## [1] 2.734883e-06
```

Selang Kepercayaan Parameter Regresi

```
## Warning in knitr::include_graphics("C:/Users/ASUS/Downloads/WhatsApp Image
## 2024-02-11 at 21.46.33.jpeg"): It is highly recommended to use relative pa
ths
## for images. You had absolute paths: "C:/Users/ASUS/Downloads/WhatsApp Imag
e
## 2024-02-11 at 21.46.33.jpeg"
```

$$\hat{\beta}_0 - t_{(n-2;\frac{\alpha}{2})} s_{\hat{\beta}_0} < \hat{\beta}_0 < \hat{\beta}_0 + t_{(n-2;\frac{\alpha}{2})} s_{\hat{\beta}_0}$$

$$\hat{\beta}_1 - t_{(n-2;\frac{\alpha}{2})} s_{\hat{\beta}_1} < \hat{\beta}_1 < \hat{\beta}_1 + t_{(n-2;\frac{\alpha}{2})} s_{\hat{\beta}_1}$$

1. Selang Kepercayaan bagi $\widehat{\beta_0}$

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

## [1] 117.8225

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

## [1] 135.5348
```

Sehingga dapat disusun suatu selang kepercayaan untuk $\widehat{\beta_0}$ sebagai berikut.

$$117.8225 < \widehat{\beta_0} < 135.5348$$

Dengan selang kepercayaan 95%, dapat diyakini bahwa dugaan parameter $\widehat{\beta_0}$ berada dalam selang 117.8225 hingga 135.5348.

2. Selang Kepercayaan bagi $\widehat{\beta_1}$

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

## [1] 0.05342733

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

## [1] 0.1255816
```

Sehingga dapat disusun suatu selang kepercayaan untuk $\widehat{\beta_1}$ sebagai berikut.

$$0.05342733 < \widehat{\beta_1} < 0.1255816$$

Dengan selang kepercayaan 95%, dapat diyakini bahwa dugaan parameter $\widehat{\beta_1}$ berada dalam selang 0.05342733 hingga 0.1255816.

3. Selang Kepercayaan Rataan (Nilai Harapan) Amatan

Misal ingin diduga rataan (nilai harapan) amatan ketika x = 350

$$E\left(\hat{Y}|x_0\right) \pm t_{\left(n-2;\frac{\alpha}{2}\right)} s_e \sqrt{\left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]}$$

```
amatan.diduga <- data.frame(x=350)
predict(model, amatan.diduga, interval = "confidence")

##        fit      lwr      upr
## 1 158.0052 150.9756 165.0349
```

Dengan dugaan rataan nilai Y ketika nilai x = 350 adalah 158.0052 dan dalam selang kepercayaan 95%, dapat diyakini bahwa nilai dugaan rataan Y ketika nilai x = 350 berada dalam selang 150.9756 hingga 165.0349.

4. Selang Kepercayaan Individu Amatan

Misal ingin diduga individu amatan ketika x = 350

$$\hat{y}(x_i) \pm t_{\left(n-2;\frac{\alpha}{2}\right)} s_e \sqrt{\left[1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]}$$

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

##        fit      lwr      upr
## 1 158.0052 103.1779 212.8325
```

Dengan dugaan rataan nilai Y ketika nilai x = 350 adalah 158.0052 dan dalam selang kepercayaan 95%, dapat diyakini bahwa nilai amatan individu Y ketika nilai x = 350 berada dalam selang 103.1779 hingga 212.8325.