

Nama : Fabyan Riza Kiram
Kelas : Medstat A - 1
NIM : 164221068

1. Apa itu regresi linier? Variabel apa saja yang dibutuhkan dalam melakukan pemodelan regresi linier?

Jawab: Regresi linier adalah suatu metode dalam statistik yang berfungsi dalam memprediksi suatu variable (variable dependen) berdasarkan variable lain (variable independent) dengan menggunakan sebuah persamaan

2. Sebutkan Langkah awal sebelum regresi linier sederhana dapat dilakukan?

Jawab: Sebelum dapat melakukan regresi linier sederhana, kita perlu menyusun tabel yang berisi variable x (variable independent) dan variable y (variable dependen). Dan setelah itu kita dapat menggambar scatter plot/diagram pencar agar kita dapat meng-estimasi apakah data tersebut memiliki pola yang linier.

3. Apakah regresi linear sederhana dapat digunakan untuk memprediksi? Sebutkan alasannya!

Jawab: Regresi linear sederhana dapat digunakan untuk memprediksi suatu variabel dikarenakan terdapat persamaan yang menentukan hasil dari variabel yang ingin kita prediksi. Kekuatan suatu persamaan regresi dalam memprediksi suatu variabel bergantung pada faktor r-squared. Makin besar nilai r-squared maka makin besar juga keakuratan hasil untuk memprediksi suatu variabel

4. Berdasarkan dataset ini. Lakukan pemodelan regresi linier dengan menggunakan fungsi perhitungan manual di R dan interpretasikan hasilnya beserta residual yang didapat!

Jawab:

- a. Syntax

```

setwd("D:/Materi/SEMESTER 2/Medstat/Kumpulan data")
data1 = read.csv("possum.csv", header = TRUE) # Memanggil data
data1

## MENGGUNAKAN FUNGSI BAWAAN R

x = data1$footlgth # Variabel Independen
y = data1$earconch # Variabel dependen

# Menampilkan scatter plot dari variabel telinga dan panjang kaki
plot(x,y)
abline(lm(y ~ x))

# Mencari hasil dengan rumus regresi
hasil <- lm(y ~ x, data = data1)

# Menampilkan hasil dari perhitungan regresi
summary(hasil)

# Estimasi Parameter
hasil$coefficients

# R - Squared
summary(hasil)$r.squared

# Residual
resid(hasil)

```

b. Output Syntax

```

Call:
lm(formula = y ~ x, data = data1)

Residuals:
    Min       1Q   Median       3Q      Max
-7.1319 -1.4162  0.0594  1.5542  7.1541

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.2535     3.9170  -0.32    0.75
x              0.7225     0.0571  12.65 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.535 on 101 degrees of freedom
(1 observation deleted due to missingness)
Multiple R-squared:  0.6132,    Adjusted R-squared:  0.6093
F-statistic: 160.1 on 1 and 101 DF,  p-value: < 2.2e-16

> # Estimasi Parameter
> hasil$coefficients
(Intercept)          x
-1.2534634    0.7224777
> # R - Squared
> summary(hasil)$r.squared
[1] 0.6131671

```

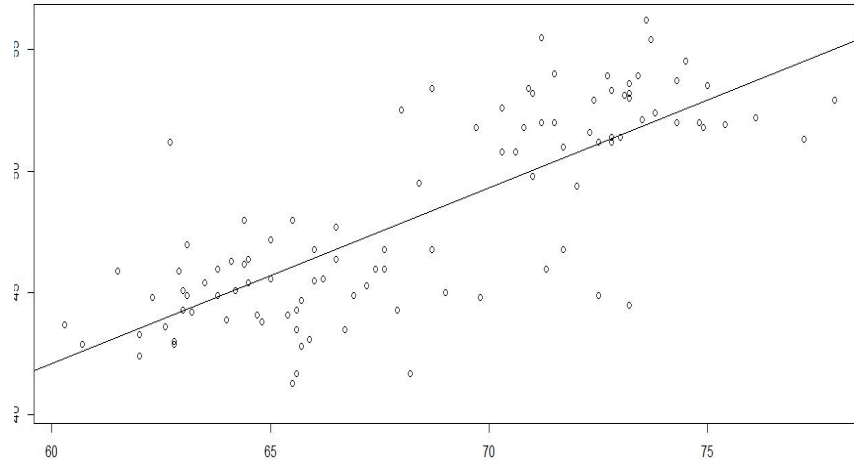
c. Residual

```

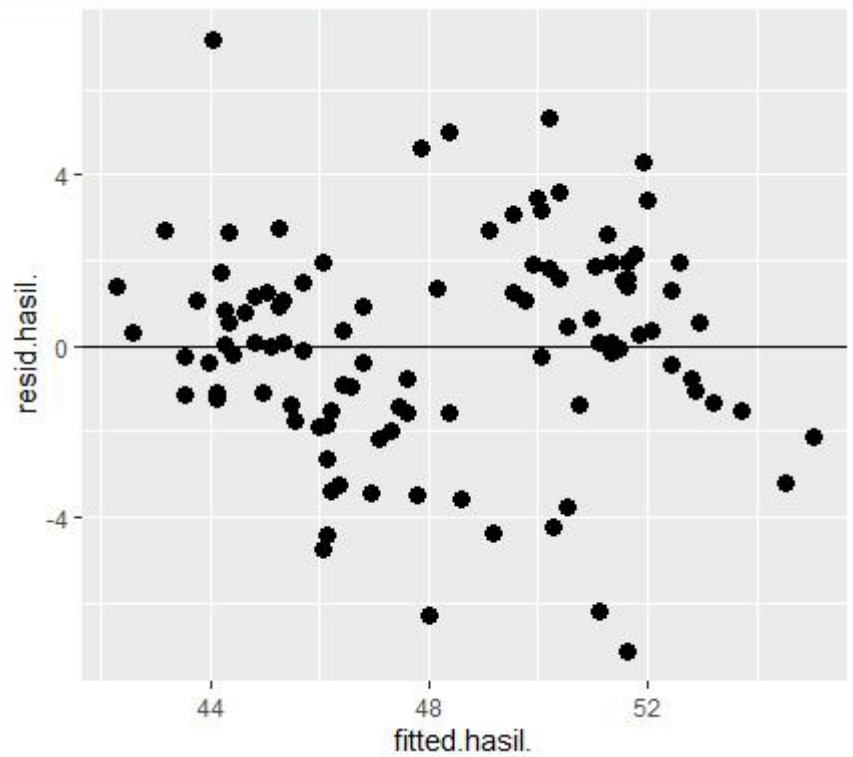
> resid(hasil)
      1      2      3      4      5      6      7
1.92887766 0.07383299 -1.32135223 -1.52708659 3.15754948 1.96809862 1.59631065
      8      9     10     11     12     13     14
2.62933745 1.84608075 3.42979724 -3.22181202 0.45181511 -0.24245052 1.27337320
     15     16     17     18     19     20     21
2.69677044 -0.08740584 1.56809862 3.40685979 2.12360309 0.33461203 -2.12754638
     22     23     24     25     26     27     28
0.25135532 1.95708969 0.56763883 0.61832852 4.27910756 1.26328384 1.36809862
     29     30     31     32     33     34     35
-1.06011340 1.04654054 4.62498246 -0.78786563 1.90204501 1.54034639 5.31305394
     36     37     38     39     40     42     43
-0.42662680 1.81305394 1.33599140 5.01924810 0.05708969 7.15411406 -1.36492818
     44     45     46     47     48     49     50
3.06328384 3.59631065 -0.14291031 -0.93006222 -2.18029211 -3.59749520 1.49241544
     51     52     53     54     55     56     57
-1.08510690 -3.50276977 -1.44153094 -4.25919382 -6.22616701 -1.58075190 -1.99703541
     58     59     60     61     62     63     64
-0.39130105 -7.13190138 2.66512300 0.56512300 -4.76882339 1.24264534 0.83737076
     65     66     67     68     69     70     71
-6.31951307 -0.20712477 -1.39084126 -3.25781445 -0.10758456 -1.89657562 -3.41331892
     72     73     74     75     76     77     78
-1.14015157 -1.51331892 2.72108726 -0.97455775 1.05365427 -1.21813370 0.03737076
     79     80     81     82     83     84     85
-1.11813370 1.04310513 -4.44107115 0.29906938 -0.24015157 -0.37363817 -2.64107115
     86     87     88     89     90     91     92
1.38806045 -3.74818489 -3.43579658 0.92590204 -4.37547733 1.93117661 -0.78602648
     93     94     95     96     97     98     99
0.05938864 -0.02960243 0.05365427 0.90869895 1.15938864 -1.84107115 -1.76308903
    100    101    102    103    104
0.36993778 2.72590204 0.77613193 1.70961853 -1.58602648

```

d. Diagram Pencar
Regresi Linier:



Residual:



e. Interpretasi

- Berdasarkan hasil output diatas, dapat diketahui bahwa data memiliki persamaan regresi:

$$y = -1.2535 + 0.7225x$$

- Dikarenakan nilai x yang positif, maka terdapat hubungan searah diantara daun telinga posum dan panjang kakinya
- Nilai R^2 yang dihasilkan sebesar 61.32%, yang artinya variabel independen mampu menjelaskan variabel dependen sebesar 61.32% dan sisanya berasal dari model lain. Hal ini menunjukkan bahwa model regresi yang telah di hitung adalah model yang cukup baik
- Dikarenakan P – value dari koefisien x lebih kecil dibanding alpha(0.05), maka dapat disimpulkan bahwa panjang kaki memiliki pengaruh yang signifikan terhadap besar daun telinga possum

5. Berdasarkan dataset ini. Lakukan pemodelan regresi linier dengan menggunakan perhitungan manual di R dan interpretasikan hasilnya beserta residual yang didapat

Jawab:

- a. Syntax

```
## MENGGUNAKAN CARA MANUAL
slope = function(x,y){
  mean_x = mean(x)
  mean_y = mean(y)
  sxy = sum((x - mean_x)*(y - mean_y))
  sxx = sum((x - mean_x)^2)
  b1 = sxy / sxx
  return (b1)
}

intercept <- function(x, y, b1){
  b0 <- mean(y) - (b1 * mean(x))
  return(b0)
}

b1 <- slope(x, y)
b0 <- intercept(x, y, b1)
# pembuatan model
model = lm(y~x, data=data1)
summary(model)
```

- b. Output

```

Call:
lm(formula = y ~ x, data = data1)

Residuals:
    Min       1Q   Median       3Q      Max
-7.1319 -1.4162  0.0594  1.5542  7.1541

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.2535     3.9170   -0.32    0.75
x              0.7225     0.0571   12.65 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.535 on 101 degrees of freedom
(1 observation deleted due to missingness)
Multiple R-squared:  0.6132,    Adjusted R-squared:  0.6093
F-statistic: 160.1 on 1 and 101 DF,  p-value: < 2.2e-16

> # Estimasi Parameter
> hasil$coefficients
(Intercept)          x
-1.2534634    0.7224777
> # R - Squared
> summary(hasil)$r.squared
[1] 0.6131671

```

c. Residual

```

> resid(hasil)
      1      2      3      4      5      6      7
1.92887766 0.07383299 -1.32135223 -1.52708659 3.15754948 1.96809862 1.59631065
8      9     10     11     12     13     14
2.62933745 1.84608075 3.42979724 -3.22181202 0.45181511 -0.24245052 1.27337320
15     16     17     18     19     20     21
2.69677044 -0.08740584 1.56809862 3.40685979 2.12360309 0.33461203 -2.12754638
22     23     24     25     26     27     28
0.25135532 1.95708969 0.56763883 0.61832852 4.27910756 1.26328384 1.36809862
29     30     31     32     33     34     35
-1.06011340 1.04654054 4.62498246 -0.78786563 1.90204501 1.54034639 5.31305394
36     37     38     39     40     42     43
-0.42662680 1.81305394 1.33599140 5.01924810 0.05708969 7.15411406 -1.36492818
44     45     46     47     48     49     50
3.06328384 3.59631065 -0.14291031 -0.93006222 -2.18029211 -3.59749520 1.49241544
51     52     53     54     55     56     57
-1.08510690 -3.50276977 -1.44153094 -4.25919382 -6.22616701 -1.58075190 -1.99703541
58     59     60     61     62     63     64
-0.39130105 -7.13190138 2.66512300 0.56512300 -4.76882339 1.24264534 0.83737076
65     66     67     68     69     70     71
-6.31951307 -0.20712477 -1.39084126 -3.25781445 -0.10758456 -1.89657562 -3.41331892
72     73     74     75     76     77     78
-1.14015157 -1.51331892 2.72108726 -0.97455775 1.05365427 -1.21813370 0.03737076
79     80     81     82     83     84     85
-1.11813370 1.04310513 -4.44107115 0.29906938 -0.24015157 -0.37363817 -2.64107115
86     87     88     89     90     91     92
1.38806045 -3.74818489 -3.43579658 0.92590204 -4.37547733 1.93117661 -0.78602648
93     94     95     96     97     98     99
0.05938864 -0.02960243 0.05365427 0.90869895 1.15938864 -1.84107115 -1.76308903
100     101     102     103     104
0.36993778 2.72590204 0.77613193 1.70961853 -1.58602648

```

d. Interpretasi:

- Berdasarkan hasil output diatas, dapat diketahui bahwa data memiliki persamaan regresi:

$$y = -1.2535 + 0.7225x$$

- Dikarenakan nilai x yang positif, maka terdapat hubungan searah diantar daun telinga posum dan panjang kakinya
- Nilai R^2 yang dihasilkan sebesar 61.32%, yang artinya variable independen mampu menjelaskan variabel dependen sebesar 61.32% dan sisanya berasal dari model lain. Hal ini menunjukkan bahwa model regresi yang telah di hitung adalah model yang cukup baik
- Dikarenakan P – value dari koefisien x lebih kecil dibanding alpha(0.05), maka dapat disimpulkan bahwa panjang kaki memiliki pengaruh yang signifikan terhadap besar daun telinga possum