Tugas Mandiri 2 Pengantar Model Linier

Angga Fathan Rofiqy

25 September, 2023

Rpubs : https://rpubs.com/ZenR_Prog/PML-TM2

No 1.

1. Dalam suatu penelitian yang dilakukan terhadap 10 rumah tangga yang dipilih secara acak, diperoleh data pengeluaran untuk pembelian barang-barang tahan lama per minggu (Y), pendapatan per minggu (X1), dan banyaknya anggota rumah tangga (X2) sebagai berikut: (menggunakan software R, syntax/html gabung saja menjadi 1 pdf)

X1 (Ratusan Ribu Rupiah)	X2 (Orang)	Y (Puluhan Ribu Rupiah)
10	7	23
2	3	7
4	2	15
6	4	17
8	6	23
7	5	22
4	3	10
6	3	14
7	4	20
6	3	19

Data tersebut dapat dimodelkan menjadi $Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X2 + \varepsilon_i$.

Input data

```
n <- 10 #10 Rumah tangga
y <- c(23,7,15,17,23,22,10,14,20,19)
#Matriks X
X <- cbind(X0 = rep(1, n),
           X1 = c(10,2,4,6,8,7,4,6,7,6),
           X2 = c(7,3,2,4,6,5,3,3,4,3) )
```

Diketahui banyaknya peubah penjelas yakni X1 dan X2, sehingga k=2.

a. Tunjukan bahwa model tersebut merupakan model penuh.

Model linier disebut model penuh apabila $r(\mathbf{X}) = k + 1$, dalam kasus ini $r(\mathbf{X}) = 2 + 1$, sehingga model dikatakan model penuh apabila rank nya bernilai 3.

Lihat rank dari matriks X:

```
if (qr(X)$rank == ncol(X)){
cat("Matriks X Model Penuh, dengan rank :", qr(X)$rank)
} else {
```

```
cat("Matriks X BUKAN Model Penuh, dengan rank :", qr(X)$rank)
}

## Matriks X Model Penuh, dengan rank : 3
```

Terlihat bahwa Model linier tersebut memiliki rank matriks $\mathbf{X}$ sebesar 3 atau sebesar $k + 1$ sehingga memiliki nilai determinan dari matriks $\mathbf{X}^t\mathbf{X} \neq 0$. Jadi, terbukti bahwa model tersebut **merupakan model penuh.**

b. Dugalah parameter model tersebut.

Mencari penduga tak bias bagi β

```
E <- rnorm(n,0,1) #pembangkitan nilai galat
b <- solve(t(X) %*% X) %*% t(X) %*% y  #mencari matriks beta = (X'X)^-1*X'y
b

##           [,1]
## X0  3.9187279
## X1  2.4911661
## X2 -0.4664311
```

Dari nilai parameter β diatas, didapatkan model :

$$\widehat{Y}_i = 3.9187279 + 2.4911661X_1 - 0.4664311X_2 + \varepsilon_i$$

c. Hitung penduga tak bias bagi ragam

Mencari penduga tak bias bagi ragam

Dengan formula :

$$s^2 = \frac{(y - \mathbf{X}^t\mathbf{b})^t(y - \mathbf{X}^t\mathbf{b})}{n - p}$$

```
p <- nrow(b) #banyak beta pada model
s2 <- (t(y-(X %*% b)) %*%( y-(X %*% b)) ) / (n-p) #mencari nilai s^2
s2

##          [,1]
## [1,] 6.355376
```

Diperoleh nilai penduga tak bias bagi ragam : $s^2 = 6.355376$.

No 2.

2. Dalam pembuatan produk kayu komersial, penting untuk memperkirakan hubungan antara kepadatan produk kayu (density) dan kekakuannya (stiffness). Jenis baru dari papan partikel sedang dipertimbangkan. Tiga puluh papan ini diproduksi dengan kepadatan berkisar antara 8 hingga 26 pon per kaki kubik, dan kekakuan diukur dalam pon per inci persegi. Diperoleh data di bawah ini. (Berdasarkan data yang disajikan dalam "Investigation of Certain Mechanical Properties of a Wood-Foam Composite", Terrance E. Conners, M.S. thesis, Departement of Forestry and Wildlife Management, University of Massachusetts, Amherst, 1979.)

Density (x) lb/ft^3	Stiffness (y) lb/in^2	Density (x) lb/ft^3	Stiffness (y) lb/in^2
15.0	2,622	8.4	17,502
14.5	22,148	9.8	14,007
14.8	26,751	11.0	19,443
13.6	18,036	8.3	7,573
25.6	96,305	9.9	14,191
23.4	104,170	8.6	9,714
24.4	72,594	6.4	8,076
23.3	49,512	7.0	5,304
19.5	32,207	8.2	10,728
21.2	48,218	17.4	43,243
22.8	70,453	15.0	25,319
21.7	47,661	15.2	28,028
19.8	38,138	16.4	41,792
21.3	54,045	16.7	49,499
9.5	14,814	15.4	25,312

For these data,

$$X'X = \begin{bmatrix} 30 & 464.1 \\ 464.1 & 8166.29 \end{bmatrix}, \quad (X'X)^{-1} = \begin{bmatrix} 0.2758892 & -0.0156791 \\ -0.0156791 & 0.001013517 \end{bmatrix},$$

$$X'\mathbf{y} = \begin{bmatrix} 1017405 \\ 19589339 \end{bmatrix}, \quad \text{and} \quad s^2 = 165242295.59$$

Input data

```
n <- 30 #30 Papan

y  <- c(2622, 22148, 26751, 18036, 96305, 104170, 72594, 49512, 32207, 48218,
        70453, 47661, 38138, 54045, 14814, 17502, 14007, 19443, 7573, 14191,
        9714, 8076, 5304, 10728, 43243, 25319, 28028, 41792, 49499, 25312)

#Matriks X
X <- cbind(X0 = rep(1, n),
           X1 = c(15.0, 14.5, 14.8, 13.6, 25.6, 23.4, 24.4, 23.3, 19.5, 21.2,
                  22.8, 21.7, 19.8, 21.3, 9.5, 8.4, 9.8, 11.0, 8.3, 9.9,
                  8.6, 6.4, 7.0, 8.2, 17.4, 15.0, 15.2, 16.4, 16.7, 15.4) )

#Matriks X'X
```

```r
tX.X <- matrix(c(30, 464.1,
                 464.1, 8166.29), nrow = 2)

#Matriks X'X
tX.X.inv <- matrix(c(0.2758892, -0.0156791,
                     -0.0156791, 0.001013517), nrow = 2)

#Matriks X'y
tX.y <- matrix(c(1017405, 19589339), ncol = 1)

#Penduga ragam s^2
s2 <- 165242295.59
```

a. Dugalah β

```r
b <- solve(t(X) %*% X) %*% t(X) %*% y #mencari (X'X)^-1*X'y
b

##           [,1]
## X0 -26452.394
## X1   3902.126
```

Dari nilai parameter β diatas, didapatkan model :

$$\widehat{Y_i} = -26452.394 + 3902.126X_1 + \varepsilon_i$$

b. Tentukan penduga titik untuk pembacaan kekakuan (*stiffness*) rata-rata ketika kepadatan (*density*) papan partikel adalah $10lb/ft^3$. Tentukan selang kepercayaan 95% pada pembacaan rata-rata ini

Dengan formula :

$$ \mathbf{x_*}^t \mathbf{b} \pm t_{\left(n-p;\frac{\alpha}{2} \right)} s \sqrt{\mathbf{x_*}^t $(\mathbf{X}^t\mathbf{X})^-1 \mathbf{x_*}} $$

Penduga titik rata-rata ketika x = 10

```r
xbaru <- c(1,10)
xb <- t(xbaru) %*% b; xb

##           [,1]
## [1,] 12568.87
```

Ini sama saja dengan membuat model linear regresi dan substitusi nilai baru pada matriks $\mathbf{x}_*$.

Simpangan baku

$$s = \sqrt{s^2} = \sqrt{\frac{(y - \mathbf{X}^t\mathbf{b})^t(y - \mathbf{X}^t\mathbf{b})}{n - p}}$$

```
#Ragam
p <- nrow(b) #banyak beta pada model
s <- sqrt( (t(y-(X %*% b)) %*%( y-(X %*% b)) ) / (n-p) )
cat("Simpangan baku (s) :", s)

## Simpangan baku (s) : 12854.66
```

Matriks $(\mathbf{X}^t\mathbf{X})^-1$

```
tX.X.inv

##              [,1]         [,2]
## [1,]  0.2758892 -0.015679100
## [2,] -0.0156791  0.001013517
```

Nilai $t_{\left(n-p;\frac{\alpha}{2}\right)}$

Dengan selang Kepercayaan 95% maka $\alpha = 5\%$.

```
alpha <- 5/100
t <- qt(1 - alpha/2, n-p)
cat('t :', t)

## t : 2.048407
```

Selang kepercayaan 95% bagi rata-rata

```
cat(" Selang kepercayaan 95% bagi rata-rata\n",
    "Batas Bawah :", xb - t *s *sqrt(t(xbaru) %*% tX.X.inv %*% xbaru), "\n",
    "Batas Atas  :", xb + t *s *sqrt(t(xbaru)%*% tX.X.inv %*% xbaru)
    )

##  Selang kepercayaan 95% bagi rata-rata
##  Batas Bawah : 5925.224
##  Batas Atas  : 19212.51
```

$$ \mathbf{x_*}^t \mathbf{b} \pm t_{\left(n-p;\frac{\alpha}{2} \right)} s \sqrt{\mathbf{x_*}^t $(\mathbf{X}^t\mathbf{X})^-1 \mathbf{x_*}}=[5925.224; 19212.51] $$

c. Tentukan penduga titik untuk pembacaan kekakuan (*stiffness*) rata-rata ketika kepadatan (*density*) papan partikel adalah $10lb/ft^3$. Tentukan selang prediksi 95% kekakuan untuk papan seperti itu

Dengan Formula :

$$ \mathbf{x_*}^t \mathbf{b} \pm t_{\left(n-p;\frac{\alpha}{2} \right)} s \sqrt{ 1 + \mathbf{x_*}^t $(\mathbf{X}^t\mathbf{X})^-1 \mathbf{x_*}} $$

Sama seperti bagian b, hanya beda di formulanya saja.

Selang kepercayaan 95% bagi satu amatan

```
cat(" Selang kepercayaan 95% bagi satu amatan\n",
    "Batas Bawah :", xb - t *s *sqrt(1 + t(xbaru) %*% tX.X.inv %*% xbaru),
"\n",
    "Batas Atas  :", xb + t *s *sqrt(1 + t(xbaru )%*% tX.X.inv %*% xbaru)
    )
```

```
##  Selang kepercayaan 95% bagi satu amatan
##  Batas Bawah : -14587.9
##  Batas Atas  : 39725.64
```

$$ \mathbf{x_*}^t \mathbf{b} \pm t_{\left(n-p;\frac{\alpha}{2} \right)} s \sqrt{ 1 + \mathbf{x_*}^t $(\mathbf{X}^t\mathbf{X})^-1 \mathbf{x_*}} =[-14587.9; 39725.63] $$

d. Tentukan selang kepercayaan 95% pada kemiringan garis regresi

Dengan formula :

$$\beta_1 \pm t_{\left(n-p;\frac{\alpha}{2}\right)} \sqrt{C_{11}}$$

Selang kepercayaan 95% bagi kemiringan garis regresi (β_1)

```
cat(" Selang kepercayaan 95% bagi kemiringan garis regresi (b1)\n",
    "Batas Bawah :", b[2] - t *s *sqrt(tX.X.inv[2,2]), "\n",
    "Batas Atas  :", b[2] + t *s *sqrt(tX.X.inv[2,2])
    )
```

```
##  Selang kepercayaan 95% bagi kemiringan garis regresi (b1)
##  Batas Bawah : 3063.84
##  Batas Atas  : 4740.413
```

$$\beta_1 \pm t_{\left(n-p;\frac{\alpha}{2}\right)} \sqrt{C_{11}} = [3063.84; 4740.413]$$

e. Tentukan daerah kepercayaan gabungan 95% pada pasangan parameter (β_0, β_1).

$$P\left[(b-\beta)^t (X^t X)(b-\beta) \le s^2 p F_{p,(n-p)}\right] = 1-\alpha$$

```
cat("Matriks b :\n"); b
```

```
## Matriks b :

##           [,1]
## X0 -26452.394
## X1   3902.126
```

```
cat("\n\nMatriks X'X :\n"); tX.X
```

```
##
##
## Matriks X'X :
```

```
##         [,1]    [,2]
## [1,]  30.0  464.10
## [2,] 464.1 8166.29
```

```
cat("\n\ns^2       :", s2,
    "\np        :", p,
    "\nF(p,n-p) :", qf(1 - alpha, p, n-p))
```

```
##
##
## s^2      : 165242296
## p        : 2
## F(p,n-p) : 3.340386
```

$$\left(\begin{bmatrix}-26452.394\\3902.126\end{bmatrix}-\begin{bmatrix}\beta_0\\\beta_1\end{bmatrix}\right)^t\left(\begin{bmatrix}30 & 464.10\\464.1 & 8166.29\end{bmatrix}\right)\left(\begin{bmatrix}-26452.394\\3902.126\end{bmatrix}-\begin{bmatrix}\beta_0\\\beta_1\end{bmatrix}\right)$$
$$\leq 165242296\times 2\times 3.340386$$

```
cat("\ns^2 * p * F(p,n-p), :", s2 *p *qf(1 - alpha, p, n-p))
```

```
##
## s^2 * p * F(p,n-p), : 1103945956
```

Perhitugan ini akan mengasilkan :

$$c_1\beta_0^2\pm c_2\beta_1^2\pm c_3\beta_0\beta_1\pm c_4\beta_0\pm c_5\beta_1\pm c_6\leq 1103945956$$

Dengan c adalah konstanta.

Agar memudahkan perhitungan di R, maka pisah menjadi :

$$\left(\begin{bmatrix}-26452.394 & 3902.126\end{bmatrix}\begin{bmatrix}30 & 464.10\\464.1 & 8166.29\end{bmatrix}\left(\begin{bmatrix}-26452.394\\3902.126\end{bmatrix}-\begin{bmatrix}\beta_0\\\beta_1\end{bmatrix}\right)\right)$$
$$-\left(\begin{bmatrix}\beta_0 & \beta_1\end{bmatrix}\begin{bmatrix}30 & 464.10\\464.1 & 8166.29\end{bmatrix}\left(\begin{bmatrix}-26452.394\\3902.126\end{bmatrix}-\begin{bmatrix}\beta_0\\\beta_1\end{bmatrix}\right)\right)\leq 1103945956$$

Lalu pisah lagi menjadi :

$$\left(\begin{bmatrix}-26452.394 & 3902.126\end{bmatrix}\begin{bmatrix}30 & 464.10\\464.1 & 8166.29\end{bmatrix}\begin{bmatrix}-26452.394\\3902.126\end{bmatrix}\right)$$
$$-\left(\begin{bmatrix}-26452.394 & 3902.126\end{bmatrix}\begin{bmatrix}30 & 464.10\\464.1 & 8166.29\end{bmatrix}\begin{bmatrix}\beta_0\\\beta_1\end{bmatrix}\right)$$
$$-\left(\begin{bmatrix}\beta_0 & \beta_1\end{bmatrix}\begin{bmatrix}30 & 464.10\\464.1 & 8166.29\end{bmatrix}\begin{bmatrix}-26452.394\\3902.126\end{bmatrix}\right)$$
$$+\left(\begin{bmatrix}\beta_0 & \beta_1\end{bmatrix}\begin{bmatrix}30 & 464.10\\464.1 & 8166.29\end{bmatrix}\begin{bmatrix}\beta_0\\\beta_1\end{bmatrix}\right)$$

$$\leq 1103945956$$

Dapat dilihat bahwa baris pertama akan mengasilkan c_6, baris kedua akan dan ketiga akan menghasilkan c_4 & c_5, terakhir baris keempat akan mengasilkan c_1, c_2, & c_3.

```
cat(  "c6      :", t(b) %*% tX.X %*% b,
    "\nc4 & c5 :", -(t(b) %*% tX.X) -t((tX.X %*% b)),
    "\nc1      :", tX.X[1],
    "\nc2      :", tX.X[4],
    "\nc3      :", tX.X[2] + tX.X[3] )

## c6      : 49527277168
## c4 & c5 : -2034810 -39178678
## c1      : 30
## c2      : 8166.29
## c3      : 928.2
```

$$c_1\beta_0^2 \pm c_2\beta_1^2 \pm c_3\beta_0\beta_1 \pm c_4\beta_0 \pm c_5\beta_1 \pm c_6 \leq 1103945956$$

Sehingga diperoleh :

$$30\beta_0^2 + 8166.29\beta_1^2 + 928.2\beta_0\beta_1 - 2034810\beta_0 - 39178678\beta_1 + 49527277168 \leq 1103945956$$

```
49527277168 - 1103945956

## [1] 48423331212
```

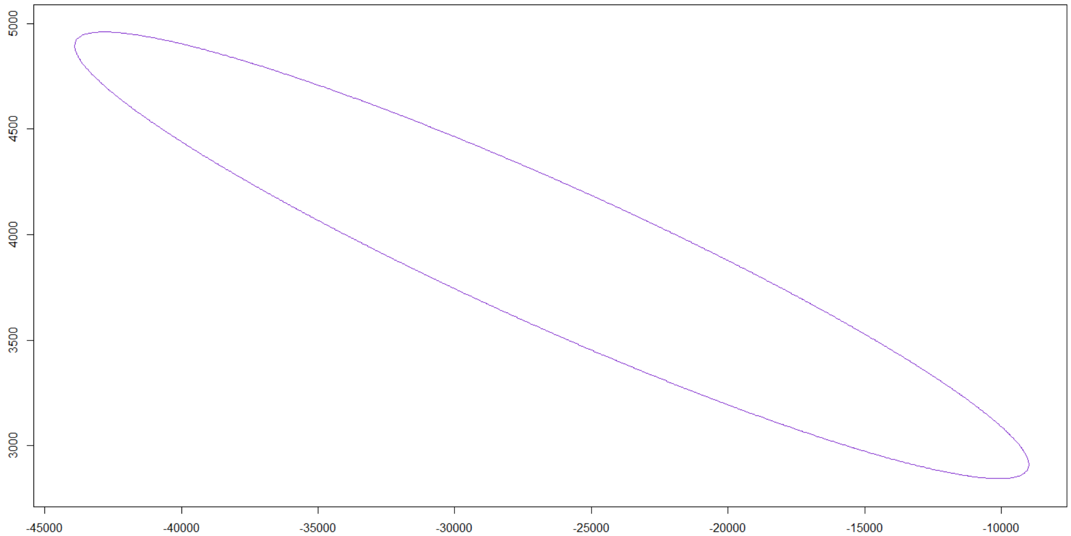
Hasil Akhirnya adalah :

$$30\beta_0^2 + 8166.29\beta_1^2 + 928.2\beta_0\beta_1 - 2034810\beta_0 - 39178678\beta_1 + 48423331212 \leq 0$$

```
# Membuat fungsi dari ketidaksetaraan
f <- function(x, y) return(30 * x^2 + 8166.29 * y^2 +
                           928.2 * x * y - 2034810 * x -
                           39178678 * y + 48423331212)

# Membuat grid untuk x dan y
x <- seq(-44000, -9000, length.out = 500)  # Rentang x dengan 500 titik
y <- seq(2800, 5000, length.out = 500)    # Rentang y dengan 500 titik
z <- outer(x, y, Vectorize(f))  # Menghitung nilai ketidaksetaraan

# Membuat plot kontur
contour(x, y, z, levels = 0, drawlabels = FALSE, xlim = c(-44000, -9000),
        ylim = c(2800, 5000), col = 'purple3')
```



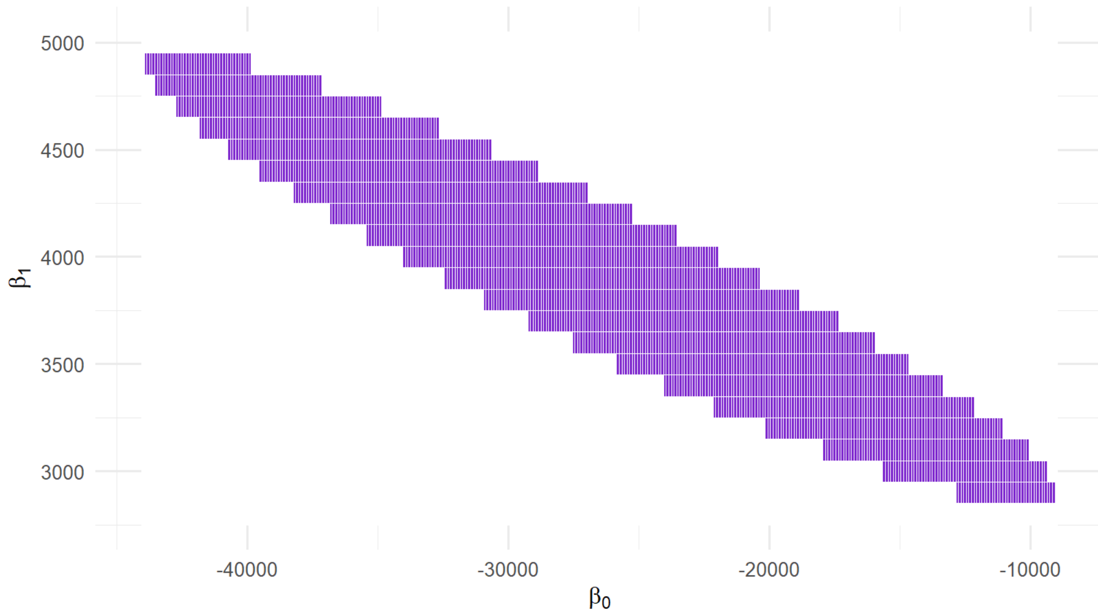
Atau dapat seperti ini, namun tidak terlihat kontinu karena butuh banyak data.

```r
# Memuat pustaka ggplot2
library(ggplot2)

# Membuat data frame dengan berbagai nilai x dan y
x <- seq(-44000, -9000, by = 100)  # Rentang x
y <- seq(2800, 5000, by = 100)    # Rentang y
data <- expand.grid(x = x, y = y)

# Menghitung nilai di sebelah kiri ketidaksetaraan
data$inequality_value <- 30 * data$x^2 + 8166.29 * data$y^2 + 928.2 * data$x
* data$y - 2034810 * data$x - 39178678 * data$y + 48423331212

# Membuat plot daerah yang memenuhi ketidaksetaraan
ggplot(data, aes(x = x, y = y)) +
  geom_tile(aes(fill = inequality_value <= 0), color = "white") +
  scale_fill_manual(values = c("TRUE" = "purple3", "FALSE" = "white")) +
  theme_minimal(base_size = 25) +
  labs(x = expression(beta[0]), y = expression(beta[1])) +
  theme(legend.position = "none")
```


5000
4500
4000
3500
3000
β_1
-40000
-30000
-20000
-10000
β_0