



IPB University
— Bogor Indonesia —

Pengantar Regresi dengan RStudio

Prof. Muhammad Firdaus, PhD

**Departemen Ilmu Ekonomi
Fakultas Ekonomi dan Manajemen
Institut Pertanian Bogor**



Apa itu Ekonometrika?

- **Verbeek:** ekonometrika adalah interaksi antara teori ekonomi, data dan metode statistika.
- Dalam hal ini statistika mengacu pada situasi yang **stokastik**
- Ekonometrika: tradisional vs modern. Perkembangan pesat sejak Sims menulis di *Econometrica* (1980)
- Kegunaan: ekonometrika
 1. Pengujian teori ekonomi
 2. Peramalan
 3. Perumusan kebijakan ekonomi

Econometric models are wrong, but some are useful (George E.P Box)

GARIS REGRESI

Garis regresi Gauss (1809): garis yang memplotkan hubungan variabel yang dipengaruhi (variabel dependen, **regressand**, respon, tidak bebas) dengan yang mempengaruhi (variabel independen, **regressor**, prediktor, bebas)

Contoh: hubungan **konsumsi** (tidak bebas) dan **pendapatan** (bebas)

Kegunaan:

- Menghitung dampak perubahan variabel bebas terhadap variabel tidak bebas (elastisitas)
- Memprediksi nilai variabel tidak bebas (peramalan)

Dari sampel, pengaruh tsb ditarik untuk populasi melalui **uji hipotesis**

Garis Regresi Populasi

$$C_i = \alpha + \beta Y_i + \varepsilon_i$$

Diagram illustrating the components of the population regression line equation:

- C_i : Variabel dependen (konsumsi)
- α : konstanta
- β : Koefisien
- Y_i : Variabel independen (pendapatan)
- ε_i : Error-term/**Galat**

The equation is divided into two components:

- Linear component**: $\alpha + \beta Y_i$
- Random Error component**: ε_i

- Sumber galat a.l:**
1. Tidak semua faktor dimasukkan ke dalam model
 2. Kesalahan dalam spesifikasi model
 3. Adanya kesalahan dan disagregasi dalam pengukuran

Garis Regresi Sampel

Variabel dependen (konsumsi) → $C_i = b_0 + b_1 Y_i + e_i$

konstanta → b_0

Koefisien → b_1

Variabel independen (pendapatan) → Y_i

Residual/sisaan → e_i

OLS atau *ordinary least squares* adalah teknik estimasi regresi sampel dengan meminimumkan jumlah kuadrat dari sisaan atau:

$$\min \Sigma (c_i - b_0 - b_1 y_i)^2$$

disebut *ordinary* karena ada **GLS**, **2SLS**, **3SLS** dll.

Estimasi dalam matriks

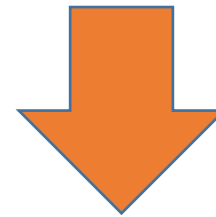
Misal ada data konsumsi (c) dan pendapatan (y) 60 RT:

$$\begin{bmatrix} c_1 \\ c_2 \\ \cdot \\ \cdot \\ \cdot \\ c_{60} \end{bmatrix}$$

matriks 60 x 1

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_{60} \end{bmatrix}$$

60 x 1



$$\widehat{b}_1 = (y'y)^{-1} \cdot y'c \quad \text{untuk koefisien}$$

$$\widehat{b}_0 = \bar{c} - \widehat{b}_1 \bar{y} \quad \text{untuk konstanta}$$

matriks 1 x 1

Evaluasi Model

Kriteria Ekonomi

Kriteria ini adalah tanda (*sign*) dan besaran (*magnitude*) dari parameter dugaan baik konstanta maupun koefisien dugaan. Untuk fungsi konsumsi, hasil regresi seharusnya menghasilkan:

1. Konstanta yang positif: *minimum need* (biasa disebut autonomous C)
2. Koefisien dugaan bernilai antara 0 dan 1: MPC

Kriteria Statistika

Kriteria ini mencakup:

1. *Goodness of fit* atau **kebaiksaan**: koefisien determinasi atau R^2
2. Uji hipotesis: uji *t-student*; uji F

Evaluasi Model

Koefisien Determinasi atau R^2

Ukuran sejauhmana model mampu mem-*fit*-kan data. Nilai R^2 berada antara 0 sampai 1 (100%). Untuk regresi dengan data deret waktu, tuntutan R^2 lebih tinggi daripada data kerat lintang. **R^2 akan semakin besar bila variabel bebas ditambah**

$$R^2 = \frac{ESS}{TSS} = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS} \quad \text{dimana} \quad \sum_t (c_t - \bar{c}_i)^2 = \sum_t (c_t - \bar{c})^2 + \sum_t e_i^2$$

Uji t

Adalah uji hipotesis terhadap koefisien dugaan secara masing-masing. Sebagai contoh untuk fungsi konsumsi, **pengujian terhadap mpc** adalah:

$$H_0: \beta_1 = 1$$

$$H_1: \beta_1 < 1$$

Kondisi Gauss-Markov

1. $E\{\varepsilon_i\} = 0, i = 1, \dots, N$ \Rightarrow **zero mean**
2. $V\{\varepsilon_i\} = \sigma^2, i = 1, \dots, N$ \Rightarrow **homoskedastisitas**
3. $Cov\{\varepsilon_i, \varepsilon_j\} = 0, i = 1, \dots, N$ untuk $i \neq j$: \Rightarrow **tidak ada autokorelasi**
4. $\{\varepsilon_1, \dots, \varepsilon_N\}$ dan $\{X_1, \dots, X_N\}$ tidak saling berhubungan: **data panel**

Beberapa referensi menyebutkan:

- Linearitas (dalam parameter: $Y = a + \mathbf{b}_1 \cdot X_1 + \mathbf{b}_2 \cdot X_2 + \mathbf{b}_3 \cdot X_1^2$)
- Data ditarik secara acak dari populasi
- Tidak ada korelasi **sempurna** antar variabel independen

Kondisi Gauss Markov

- Jika kondisi Gauss-Markov terpenuhi, maka OLS dapat dikatakan *valid* dalam pendugaan koefisien
- Dalam observasi lapang, seringkali kondisi tersebut tidak secara penuh terpenuhi. Ini menjadi tolak ukur dalam mendapatkan model yang *robust*
- Dengan kondisi Gauss-Markov dipenuhi, maka penduga OLS bersifat **BLUE**: **Best**, **Linear**, **Unbiased** estimator. Artinya penduga varians yang minimum (efisien), serta untuk sampel yang berulang penduga (b_2) secara rata-rata sama dengan β_2 .

Estimasi Regresi Linier dengan Satu Variabel Bebas

```
regresi1 = lm(Y ~ MODAL, data=data1)
print(regresi1)
summary(regresi1)
anova(regresi1)
```

```
> regresi1 = lm(Y ~ MODAL, data=data1)
> print(regresi1)

Call:
lm(formula = Y ~ MODAL, data = data1)

Coefficients:
(Intercept)      MODAL
      5.8436      0.7307
```

```
> summary(regresi1)

Call:
lm(formula = Y ~ MODAL, data = data1)

Residuals:
    Min       1Q   Median       3Q      Max
-1.4298 -0.5673 -0.1126  0.3961  3.9112

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.8436     1.5757   3.709 0.000663 ***
MODAL         0.7307     0.1057   6.913 3.22e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9815 on 38 degrees of freedom
Multiple R-squared:  0.5571,    Adjusted R-squared:  0.5454
F-statistic: 47.79 on 1 and 38 DF,  p-value: 3.222e-08
```

```
> anova(regresi1)
Analysis of Variance Table

Response: Y
      Df Sum Sq Mean Sq F value    Pr(>F)
MODAL   1  46.038   46.038  47.794 3.222e-08 ***
Residuals 38  36.604    0.963
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Estimasi Regresi Linier dengan Lebih dari Satu Variabel Bebas

```
regresi = lm(Y ~ MODAL + LU + JK + PROMOSI, data=data1)  
summary(regresi)
```

```
> regresi = lm(Y ~ MODAL + LU + JK + PROMOSI, data=data1)  
> summary(regresi)
```

```
Call:  
lm(formula = Y ~ MODAL + LU + JK + PROMOSI, data = data1)
```

```
Residuals:  
    Min       1Q   Median       3Q      Max  
-1.7197 -0.5905  0.1473  0.4702  1.7360
```

```
Coefficients:  
              Estimate Std. Error t value Pr(>|t|)  
(Intercept)  4.42437    1.42316   3.109  0.00372 **  
MODAL         0.58973    0.09145   6.448 1.99e-07 ***  
LU            0.15877    0.05114   3.105  0.00376 **  
JK            0.25862    0.08701   2.972  0.00532 **  
PROMOSI       0.50462    0.28983   1.741  0.09046 .  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.7896 on 35 degrees of freedom  
Multiple R-squared:  0.7359,    Adjusted R-squared:  0.7057  
F-statistic: 24.38 on 4 and 35 DF,  p-value: 1.054e-09
```

```
anova(regresi)
```

```
> anova(regresi)  
Analysis of Variance Table  
  
Response: Y  
      Df Sum Sq Mean Sq F value    Pr(>F)  
MODAL  1 46.038  46.038 73.8330 3.833e-10 ***  
LU      1  7.044   7.044 11.2973 0.001888 **  
JK      1  5.845   5.845  9.3743 0.004211 **  
PROMOSI 1  1.890   1.890  3.0313 0.090458 .  
Residuals 35 21.824   0.624  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Joint Test library(car)
linearHypothesis(regresi, c("MODAL=0", "LU=0", "JK=0", "PROMOSI=0"))

```
> linearHypothesis(regresi, c("MODAL=0", "LU=0", "JK=0", "PROMOSI=0"))  
Linear hypothesis test  
  
Hypothesis:  
MODAL = 0  
LU = 0  
JK = 0  
PROMOSI = 0  
  
Model 1: restricted model  
Model 2: Y ~ MODAL + LU + JK + PROMOSI  
  
  Res.Df  RSS Df Sum of Sq    F    Pr(>F)  
1      39 82.642  
2      35 21.824  4     60.818 24.384 1.054e-09 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Uji Asumsi (Normalitas, Heteroskedastisitas)

```
install.packages("gvlma")  
library(gvlma)  
gvlma::gvlma(regresi)
```

```
> gvlma::gvlma(regresi)  
  
Call:  
lm(formula = Y ~ MODAL + LU + JK + PROMOSI, data = data1)  
  
Coefficients:  
(Intercept)      MODAL          LU          JK      PROMOSI  
      4.4244      0.5897      0.1588      0.2586      0.5046  
  
ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS  
USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:  
Level of significance = 0.05  
  
Call:  
gvlma::gvlma(x = regresi)  
  
              value p-value      Decision  
Global stat    1.91103 0.7521 Assumptions acceptable.  
Skewness       0.17735 0.6737 Assumptions acceptable.  
Kurtosis       0.03161 0.8589 Assumptions acceptable.  
Link Function  0.78185 0.3766 Assumptions acceptable.  
Heteroscedasticity 0.92022 0.3374 Assumptions acceptable.
```

Hatur nuhun



IPB University
— Bogor Indonesia —

Department of Economics - IPB University
IPB Dramaga Campus, Bogor 16680
Telp.: 0251-8622602
E-mail: ilmu_ekonomi@ipb.ac.id
Web : <http://ilmuekonomi.fem.pb.ac.id>