Modul 10b: Analisis Regresi

**Tujuan Pembelajaran**

Setelah memperlajari modul ini, mahasiswa diharapkan:

1. Mampu memahami dan menjelaskan konsep dasar analisis korelasi dan regresi
2. Mampu memodelkan data dengan menggunakan analisis regresi menggunakan software R.
3. Mampu menginterpretasikan hasil pengolahan data yang diperoleh.

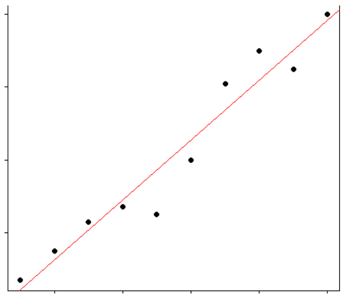
## A. Pendahuluan

Regresi linear merupakan metode statistika untuk melihat pengaruh antara variabel bebas dengan variabel terikat. Dan metode ini sering digunakan pada jenis penelitian expost facto, misalkan survei kepuasaan pelanggan untuk sebuah produk perusahaan atau pelayanan di instansi pemerintah.

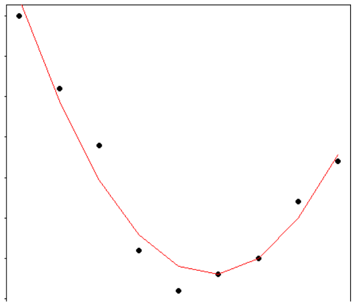
## B. Materi dan Studi Kasus

### 1. Sekilas Mengenai Regresi Linear Berganda

Regresi linear berganda (*multiple linear regression*) merupakan suatu teknik statistika yang menghasilkan suatu persamaan linear. Persamaan linear tersebut menerangkan atau menjelaskan hubungan antara variabel-variabel bebas terhadap variabel tak bebas. Dari persamaan linear tersebut juga dapat diketahui variabel bebas manakah yang memiliki kontribusi terbesar dalam pengaruhnya terhadap variabel tak bebas. Di samping itu, persamaan linear tersebut dapat digunakan untuk keperluan prediksi suatu nilai dari variabel tak bebas berdasarkan masukkan dari nilai-nilai variabel tak bebas. Gambar 1 dan Gambar 2 menyajikan kurva dari persamaan regresi linear dan persamaan persamaan regresi nonlinear.



**Gambar 1. Kurva Regresi Linear**



**Gambar 2. Kurva Regresi nonlinear**

Variabel bebas dan tak bebas yang digunakan untuk membuat persamaan regresi linear bersifat metrik (interval atau rasio). Hair dkk. (2010:151) menyatakan sebagai berikut.

*Multiple regression analysis is a general statistical technique used to analyze the relationship between a single dependent variable and independent variables. Its basic formulation is*

$$\underset{(metric)}{Y} = \underset{(metric)}{X_1 + \cdots + X_n}$$

Sebagai contoh dari variabel yang bersifat metrik adalah pendapatan per bulan, penghasilan per bulan, produksi beras per tahun, tinggi badan, berat badan, dan sebagainya. **Jika variabel tak bebas bersifat non-metrik atau kategori** (nominal atau ordinal), maka alternatif teknik statistika yang dapat digunakan adalah regresi logistik, analisis diskriminan, atau pohon klasifikasi (*classification tree*). Namun **jika variabel tak bebas dan variabel bebas bersifat kategori**, maka alternatif teknik statistika yang dapat digunakan adalah regresi logistik atau pohon klasifikasi.

## 2. Beberapa contoh Aplikasi dari Regresi Linear Berganda

⇨ Membuat suatu persamaan linear untuk memprediksi indeks harga saham gabungan (IHSG) berdasarkan informasi dari tingkat inflasi, harga emas dunia, dan harga minyak mentah dunia. Kemudian dari ketiga faktor tersebut, dapat ditentukan, faktor mana yang memberikan kontribusi terbesar dalam pengaruhnya terhadap indeks harga saham gabungan.

⇨ Membuat suatu persamaan linear untuk memprediksi atau mengestimasi laba perusahaan berdasarkan umur perusahaan, tingkat penjualan, dan besarnya ukuran perusahaan. Di samping itu, dapat diketahui seberapa besar kontribusi yang diberikan dari faktor umur perusahaan terhadap naik/turunnya laba perusahaan, dengan mengontrol pengaruh tingkat penjualan dan besarnya perusahaan. Dapat juga diketahui seberapa besar kontribusi yang diberikan dari faktor tingkat penjualan perusahaan terhadap naik/turunnya laba perusahaan, dengan mengontrol pengaruh umur perusahaan dan besarnya perusahaan.

⇨ Membuat suatu persamaan linear untuk memprediksi pengeluaran per bulan dari suatu rumah tangga berdasarkan informasi penghasilan per bulan, jumlah anggota keluarga, dan jumlah kendaraan yang dimiliki. Kemudian dari ketiga faktor tersebut, dapat ditentukan, faktor mana yang memberikan kontribusi terbesar dalam pengaruhnya terhadap pengeluaran per bulan.

**3. Koefisien Korelasi Linear Pearson (Mengukur Keeratan Hubungan Linear antar Variabel)**

Misalkan seorang peneliti ingin membuat model regresi linear berganda dengan menggunakan variabel indeks prestasi $(Y)$ sebagai variabel tak bebas, variabel jumlah jam belajar dalam sehari $(X_1)$ dan uang jajan dalam sehari $(X_2)$ sebagai variabel bebas. Data yang telah dikumpulkan oleh peneliti disajikan dalam Tabel 1.

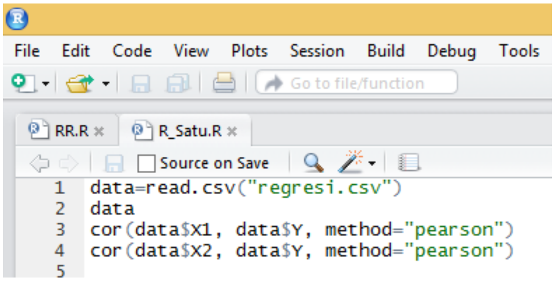
**Tabel 1 (Data Fiktif)**

| Nama | $X_1$ | $X_2$ | $Y$ | Nama | $X_1$ | $X_2$ | $Y$ |
|---|---|---|---|---|---|---|---|
| Ugi | 10 | 7 | 3,01 | Iqbal | 10 | 7 | 3,02 |
| Niar | 10 | 7 | 3,15 | Edi | 12 | 7,2 | 3,16 |
| Alvi | 9 | 11 | 2,9 | Budi | 9 | 6 | 2,95 |
| Fitri | 10 | 8 | 3,1 | Indah | 10 | 8 | 3,12 |
| Ridho | 8 | 7,5 | 2,7 | Tari | 8 | 12 | 2,8 |
| Mifdhal | 11 | 8 | 3,25 | Maura | 11 | 11 | 3,3 |
| Romi | 13 | 7 | 3,6 | Nina | 15 | 10 | 3,57 |
| Wilya | 13 | 12 | 3,7 | Suci | 17 | 8 | 3,64 |
| Windi | 15 | 9,5 | 3,65 | Febri | 16 | 9,5 | 3,6 |
| Evelin | 10 | 10 | 3,15 | Iman | 10 | 10 | 3,15 |

Berdasarkan data pada Tabel 1, jumlah responden yang diteliti sebanyak $n = 20$ responden. Misalkan responden yang diteliti adalah mahasiswa Matematika. Diketahui responden ke-1 bernama Ugi menghabiskan waktu untuk belajar dalam sehari selama 10 jam, uang jajan satu hari Rp. 7000, dan meraih IP 3,01. Responden ke-3 bernama Alvi menghabiskan waktu untuk belajar dalam sehari selama 9 jam, uang jajan dalam satu hari Rp. 11000, dan meraih IP 2,9, dan seterusnya. Misalkan akan ditentukan:

⇨ Nilai koefisien korelasi linear Pearson antara $X_1$ dan $Y$
⇨ Nilai koefisien korelasi linear Pearson antara $X_2$ dan $Y$

Berikut hasil perhitungan nilai koefisien korelasi linear Pearson berdasarkan R.




Gambar 3

```
data=read.csv("regresi.csv")
data
```

```
##    X1   X2    Y
## 1  10  7.0 3.01
## 2  10  7.0 3.15
## 3   9 11.0 2.90
## 4  10  8.0 3.10
## 5   8  7.5 2.70
## 6  11  8.0 3.25
## 7  13  7.0 3.60
## 8  13 12.0 3.70
## 9  15  9.5 3.65
## 10 10 10.0 3.15
## 11 10  7.0 3.02
## 12 12  7.2 3.16
## 13  9  6.0 2.95
## 14 10  8.0 3.12
## 15  8 12.0 2.80
## 16 11 11.0 3.30
## 17 15 10.0 3.57
## 18 17  8.0 3.64
## 19 16  9.5 3.60
## 20 10 10.0 3.15
```

```
cor(data$X1, data$Y, method="pearson")
```

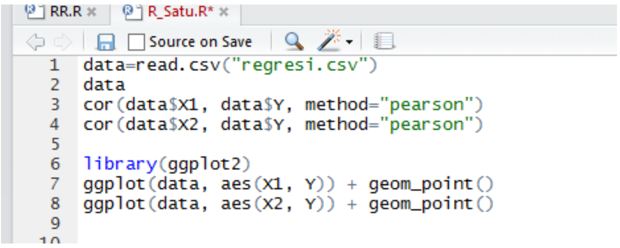
```
## [1] 0.9168692
```

```
cor(data$X2, data$Y, method="pearson")
```

```
## [1] 0.1907533
```

Gambar 4

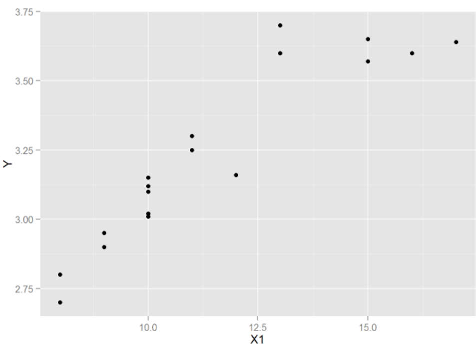
Berdasarkan Gambar 4, diketahui nilai koefisien korelasi linear Pearson antara $X_1$ dan $Y$ sebesar 0,92 (dibulatkan 2 angka di belakang koma), dan nilai koefisien korelasi linear Pearson antara $X_2$ dan $Y$ sebesar 0,19. Berdasarkan hasil tersebut, **diketahui variabel $X_1$ memiliki keeratan linear lebih erat terhadap $Y$, dibandingkan $X_2$.** R code untuk membuat grafik sebaran data dapat dilihat pada Gambar 5. Adapaun Grafik sebaran data antara $X_1$ dan $Y$, serta $X_2$ dan $Y$, disajikan pada Gambar 6 dan Gambar 7.



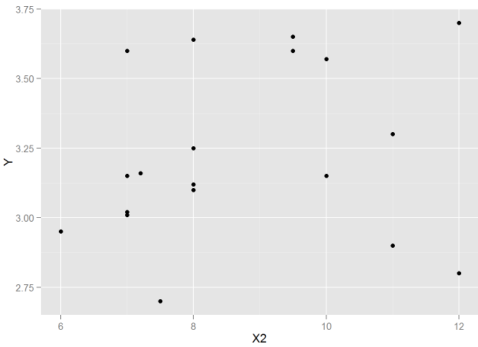
```r
data=read.csv("regresi.csv")
data
cor(data$X1, data$Y, method="pearson")
cor(data$X2, data$Y, method="pearson")

library(ggplot2)
ggplot(data, aes(X1, Y)) + geom_point()
ggplot(data, aes(X2, Y)) + geom_point()
```

**Gambar 5**




**Gambar 6**




**Gambar 7**

Perhatikan bahwa Gambar 6 merupakan grafik sebaran data antara $X_1$ dan $Y$, sementara Gambar 7 merupakan grafik sebaran data antara $X_2$ dan $Y$. **Dapat dilihat bahwa sebaran data pada Gambar 6 lebih linear dibandingkan sebaran data pada Gambar 7. Hal dapat diartikan bahwa variabel jam ($X_1$) memiliki keeratan linear lebih tinggi terhadap variabel IP ($Y$), dibandingkan variabel uang jajan ($X_2$).**

### 4. Mengestimasi Persamaan Regresi Linear Berganda

Pada pembahasan sebelumnya, diketahui bahwa variabel jam ($X_1$) memiliki keeratan linear lebih tinggi terhadap variabel IP ($Y$), dibandingkan variabel uang jajan ($X_2$). Selanjutnya akan diestimasi persamaan regresi linear berganda. Persamaan regresi linear berganda untuk kasus ini memiliki bentuk sebagai berikut.

$$\hat{Y} = \hat{\alpha} + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$$

Berikut akan dihitung $\hat{\alpha}$, $\hat{\beta}_1$, dan $\hat{\beta}_2$.

$$p = n\sum X_1Y - \sum X_1 \sum Y = (20)(746{,}3) - (227)(64{,}52) = 279{,}96$$

$$q = n\sum X_2^2 - \left(\sum X_2\right)^2 = (20)(1606{,}59) - (175{,}7)^2 = 1261{,}31$$

$$r = n\sum X_1X_2 - \sum X_1 \sum X_2 = (20)(2001{,}9) - (227)(175{,}7) = 154{,}1$$

$$s = n\sum X_2Y - \sum X_2 \sum Y = (20)(568{,}817) - (175{,}7)(64{,}52) = 40{,}176$$

$$t = n\sum X_1^2 - \left(\sum X_1\right)^2 = (20)(2709) - (227)^2 = 2651$$

$$u = tq - r^2 = (2651)(1261{,}31) - (154{,}1)^2 = 3319986$$

Nilai $X_1$, $X_2$, $X_1Y$, $X_2Y$, $X_1^2$, $X_2^2$, $Y^2$ dan $X_1X_2$ diberikan padaTabel 2 berikut:

**Tabel 2. Nilai- Nilai beberapa variabel yang digunakan dalam menghitung parameter pada regresi linear**

| $X_1$ | $X_2$ | $Y$ | $X_1Y$ | $X_2Y$ | $X_1^2$ | $X_2^2$ | $Y^2$ | $X_1X_2$ |
|---|---|---|---|---|---|---|---|---|
| 10 | 7 | 3.01 | 30.1 | 21.07 | 100 | 49 | 9.0601 | 70 |
| 10 | 7 | 3.15 | 31.5 | 22.05 | 100 | 49 | 9.9225 | 70 |
| 9 | 11 | 2.9 | 26.1 | 31.9 | 81 | 121 | 8.41 | 99 |
| 10 | 8 | 3.1 | 31 | 24.8 | 100 | 64 | 9.61 | 80 |
| 8 | 7.5 | 2.7 | 21.6 | 20.25 | 64 | 56.25 | 7.29 | 60 |
| 11 | 8 | 3.25 | 35.75 | 26 | 121 | 64 | 10.5625 | 88 |
| 13 | 7 | 3.6 | 46.8 | 25.2 | 169 | 49 | 12.96 | 91 |
| 13 | 12 | 3.7 | 48.1 | 44.4 | 169 | 144 | 13.69 | 156 |
| 15 | 9.5 | 3.65 | 54.75 | 34.675 | 225 | 90.25 | 13.3225 | 142.5 |
| 10 | 10 | 3.15 | 31.5 | 31.5 | 100 | 100 | 9.9225 | 100 |
| 10 | 7 | 3.02 | 30.2 | 21.14 | 100 | 49 | 9.1204 | 70 |
| 12 | 7.2 | 3.16 | 37.92 | 22.752 | 144 | 51.84 | 9.9856 | 86.4 |
| 9 | 6 | 2.95 | 26.55 | 17.7 | 81 | 36 | 8.7025 | 54 |
| 10 | 8 | 3.12 | 31.2 | 24.96 | 100 | 64 | 9.7344 | 80 |
| 8 | 12 | 2.8 | 22.4 | 33.6 | 64 | 144 | 7.84 | 96 |
| 11 | 11 | 3.3 | 36.3 | 36.3 | 121 | 121 | 10.89 | 121 |
| 15 | 10 | 3.57 | 53.55 | 35.7 | 225 | 100 | 12.7449 | 150 |
| 17 | 8 | 3.64 | 61.88 | 29.12 | 289 | 64 | 13.2496 | 136 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 16 | 9.5 | 3.6 | 57.6 | 34.2 | 256 | 90.25 | 12.96 | 152 |
| | 10 | 10 | 3.15 | 31.5 | 31.5 | 100 | 100 | 9.9225 | 100 |
| Jumlah | 227 | 175.7 | 64.52 | 746.3 | 568.817 | 2709 | 1606.59 | 209.9 | 2001.9 |

$$\hat{\beta}_1 = \frac{pq - rs}{u}$$

$$\hat{\beta}_1 = \frac{(279{,}96)(1261{,}31) - (154{,}1)(40{,}176)}{3319986} = 0{,}104496$$

$$\hat{\beta}_2 = \frac{st - pr}{u}$$

$$\hat{\beta}_2 = \frac{(40{,}176)(2651\,) - (279{,}96)(154{,}1)}{3319986} = 0{,}019086$$

$$\hat{\alpha} = \frac{\sum Y - \hat{\beta}_1 \sum X_1 - \hat{\beta}_2 \sum X_2}{n}$$

$$\hat{\alpha} = \frac{64{,}52 - (0{,}104496)(227) - (0{,}019086)(175{,}7)}{20} = 1{,}872301$$

Maka diperoleh persamaan regresi linear berganda

$$\hat{Y} = 1{,}872301 + 0{,}104496X_1 + 0{,}019086X_2.$$

Berikut disajikan hasil perhitungan berdasarkan R.

```
simpan_data=read.csv("regresi.csv")
simpan_data

regresi=lm(formula = Y ~ X1 + X2, data = simpan_data)
regresi
```

**Gambar 8**

```
## Call:
## lm(formula = Y ~ X1 + X2, data = simpan_data)
##
## Coefficients:
## (Intercept)           X1           X2
##     1.87230      0.10450      0.01909
```

**Gambar 9**

### 5. Memprediksi Nilai Variabel Tak Bebas

Persamaan regresi linear berganda yang telah dihasilkan sebelumnya, dapat digunakan untuk memprediksi atau mengestimasi nilai dari variabel tak bebas, berdasarkan masukan nilai-nilai dari variabel bebas. Diketahui persamaan regresi linear berganda berdasarkan perhitungan sebelumnya sebagai berikut.

$$Y = 1{,}872301 + 0{,}104496X_1 + 0{,}019086X_2$$

Misalkan akan diprediksi nilai IP, ketika jumlah jam belajar dalam sehari $X_1 = 10$ dan uang jajan dalam sehari $X_2 = 7$ (dalam ribuan).

$$\hat{Y} = 1{,}872301 + 0{,}104496(10) + 0{,}019086(7) = 3{,}050862$$

Misalkan akan diprediksi nilai IP, ketika jumlah jam belajar dalam sehari $X_1$= 6 dan uang jajan dalam sehari $X_2 = 12$ (dalam ribuan).

$$\hat{Y} = 1{,}872301 + 0{,}104496(6) + 0{,}019086(12) = 2{,}728307$$

Misalkan akan diprediksi nilai IP, ketika jumlah jam belajar dalam sehari $X_1 = 12$ dan uang jajan dalam sehari $X_2 = 10$ (dalam ribuan).

$$\hat{Y} = 1{,}872301 + 0{,}104496(12) + 0{,}019086(10) = 3{,}317112$$

Berikut disajikan ilustrasi dalam R.

```
RR.R ×   R_Satu.R* ×
Source on Save
simpan_data=read.csv("regresi.csv")
simpan_data

regresi=lm(formula = Y ~ X1 + X2, data = simpan_data)

intersep=regresi$coefficient[1]
B1=regresi$coefficient[2]
B2=regresi$coefficient[3]

intersep + 10*B1 + 7*B2
intersep + 6*B1 + 12*B2
intersep + 12*B1 + 10*B2
```

**Gambar 10**

```
regresi=lm(formula = Y ~ X1 + X2, data = simpan_data)

intersep=regresi$coefficient[1]
B1=regresi$coefficient[2]
B2=regresi$coefficient[3]

intersep + 10*B1 + 7*B2
```

```
## (Intercept)
##    3.050862
```

```
intersep + 6*B1 + 12*B2
```

```
## (Intercept)
##    2.728307
```

```
intersep + 12*B1 + 10*B2
```

```
## (Intercept)
##    3.317112
```

**Gambar 11**

### 6. Menghitung Nilai Residual Untuk Setiap Pengamatan

Residual (dilambangkan dengan $\hat{e}$) merupakan selisih antara nilai variabel tak bebas ($Y$) dan nilai estimasi dari variabel tak bebas ($\hat{Y}$). Niali nilai variable yang terkait dalam menghitung nilai residual diberikan pada Tabel 3.

**Tabel 3. Tabel nilai-nilai variabel yang terkait dengan perhitungan nilai Residual**

| No | $X_1$ | $X_2$ | $Y$ | $\hat{Y}$ | $\hat{e} = Y - \hat{Y}$ |
|---|---|---|---|---|---|
| 1 | 10 | 7 | 3.01 | 3.050862 | -0.04086 |
| 2 | 10 | 7 | 3.15 | 3.050862 | 0.099138 |
| 3 | 9 | 11 | 2.9 | 3.02271 | -0.12271 |
| 4 | 10 | 8 | 3.1 | 3.069948 | 0.030052 |
| 5 | 8 | 7.5 | 2.7 | 2.851413 | -0.15141 |
| 6 | 11 | 8 | 3.25 | 3.174444 | 0.075556 |
| 7 | 13 | 7 | 3.6 | 3.36435 | 0.23565 |
| 8 | 13 | 12 | 3.7 | 3.459779 | 0.240221 |
| 9 | 15 | 9.5 | 3.65 | 3.621057 | 0.028943 |
| 10 | 10 | 10 | 3.15 | 3.10812 | 0.04188 |
| 11 | 10 | 7 | 3.02 | 3.050862 | -0.03086 |
| 12 | 12 | 7.2 | 3.16 | 3.263671 | -0.10367 |
| 13 | 9 | 6 | 2.95 | 2.92728 | 0.02272 |
| 14 | 10 | 8 | 3.12 | 3.069948 | 0.050052 |
| 15 | 8 | 12 | 2.8 | 2.937299 | -0.1373 |
| 16 | 11 | 11 | 3.3 | 3.231702 | 0.068298 |

| | | | | | |
|---|---|---|---|---|---|
| 17 | 15 | 10 | 3.57 | 3.6306 | -0.0606 |
| 18 | 17 | 8 | 3.64 | 3.80142 | -0.16142 |
| 19 | 16 | 9.5 | 3.6 | 3.725553 | -0.12555 |
| 20 | 10 | 10 | 3.15 | 3.10812 | 0.04188 |

Berdasarkan Tabel 3, nilai estimasi $Y$ untuk responden ke-1 adalah 3,050862. Nilai tersebut diperoleh berdasarkan hasil perhitungan berikut.

$$\hat{Y} = 1{,}872301 + 0{,}104496(10) + 0{,}019086(7) = 3{,}050862$$

Nilai residual untuk responden ke-1 dihitung sebagai berikut.

$$e = \hat{Y} - Y = 3{,}01 - 3{,}050862 = -0{,}04086$$

Nilai estimasi $Y$ untuk responden ke-20 adalah 3,15. Nilai tersebut diperoleh berdasarkan hasil perhitungan berikut.

$$\hat{Y} = 1{,}872301 + 0{,}104496(10) + 0{,}019086(10) = 3{,}10812$$

Nilai residual untuk responden ke-20 dihitung sebagai berikut.

$$e = \hat{Y} - Y = 3{,}15 - 3{,}1082 = -0{,}04086$$

Hasil perhitungan dengan R untuk memperoleh nilai estimasi IP dari tiap-tiap responden, beserta residualnya dapt dilihat pada Gambar 12 dan Gambar 12.

```
simpan_data=read.csv("regresi.csv")
simpan_data

X1=simpan_data$X1
X2=simpan_data$X2
Y=simpan_data$Y

regresi=lm(formula = Y ~ X1 + X2, data = simpan_data)

intersep=regresi$coefficient[1]
B1=regresi$coefficient[2]
B2=regresi$coefficient[3]

Y_estimasi = intersep + B1*X1 + B2*X2

Y_estimasi

Residual = Y-Y_estimasi

Residual
```

**Gambar 12**

**Gambar 13**

```
X2=simpan_data$X2
Y=simpan_data$Y

regresi=lm(formula = Y ~ X1 + X2, data = simpan_data)

intersep=regresi$coefficient[1]
B1=regresi$coefficient[2]
B2=regresi$coefficient[3]

Y_estimasi = intersep + B1*X1 + B2*X2

Y_estimasi
```

```
##  [1] 3.050862 3.050862 3.022710 3.069948 2.851413 3.174444 3.364350
##  [8] 3.459779 3.621057 3.108120 3.050862 3.263671 2.927280 3.069948
## [15] 2.937299 3.231702 3.630600 3.801420 3.725553 3.108120
```

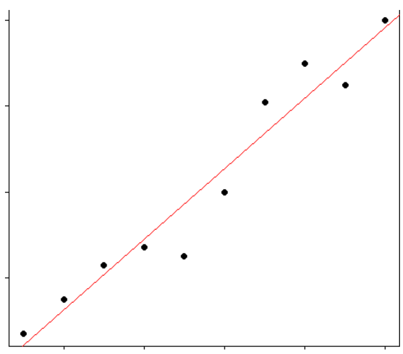
```
Residual = Y-Y_estimasi

Residual
```

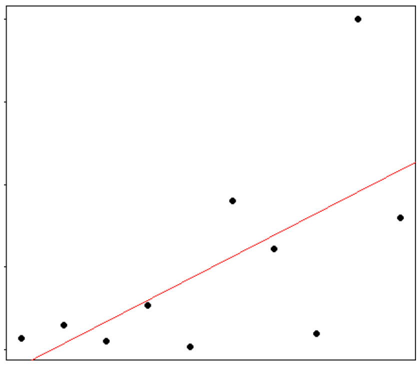
```
##  [1] -0.04086218  0.09913782 -0.12270957  0.03005198 -0.15141312
##  [6]  0.07555599  0.23564985  0.24022062  0.02894325  0.04188029
## [11] -0.03086218 -0.10367133  0.02271966  0.05005198 -0.13729942
## [16]  0.06829845 -0.06059967 -0.16141996 -0.12555274  0.04188029
```

## 7. Mengukur Kecocokan Model Regresi Linear Berganda terhadap Data dengan Koefisien Determinasi ($r^2$)

Dalam regresi linear, baik sederhana (melibatkan satu variabel bebas) maupun berganda (melibatkan lebih dari satu variabel bebas), nilai dari koefisien determinasi ($r^2$) digunakan untuk mengukur kemampuan persamaan regresi linear dalam mencocokkan atau menyesuiakan (*fits*) data. Sebagai ilustrasi perhatikan Gambar 14 dan Gambar 15. Gambar 14 dan Gambar 15 menyajikan garis persamaan regresi linear. Pada Gambar 14, garis persamaan regresi linear lebih baik dalam hal mencocokkan data dibandingkan garis persamaan regresi linear pada Gambar 15. Pada Gambar 14, titik-titik cenderung menyebar lebih dekat pada garis persamaan regresi linear, dibandingkan pada Gambar 15.

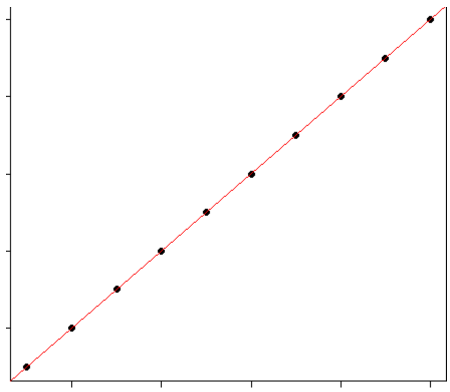


**Gambar 14**



**Gambar 15**

Nilai koefisien determinasi berkisar di antara 0 dan 1. Nilai koefisien determinasi yang bernilai 1 berarti persamaan regresi linear secara sempurna dalam mencocokkan data (Gambar 16). Nilai koefisien determinasi yang semakin mendekati 0, berarti kemampuan persamaan regresi linear semakin tidak baik dalam mencocokkan data. Dengan kata lain, kemampuan variabel-variabel bebas yang digunakan dalam persamaan regresi linear secara bersamaan atau simultan kurang mampu dalam hal menjelaskan *variation* variabel tak bebas (Gambar 15).



**Gambar 16**

Semakin tinggi nilai koefisien determinasi (mendekati 1), maka akan semakin baik suatu persamaan regresi linear dalam mencocokkan data. Dengan kata lain, kemampuan variabel-variabel bebas yang digunakan dalam persamaan regresi linear secara bersamaan atau simultan semakin baik dalam hal menjelaskan *variation* variabel tak bebas (Gujarati, 2993:87).

Pada pembahasan sebelumnya, telah diperoleh persamaan regresi linear berganda sebagai berikut.

$$\hat{Y} = 1{,}872301 + 0{,}104496X_1 + 0{,}019086X_2$$

Berikut akan dihitung nilai koefisien determinasi dari persamaan regresi linear berganda tersebut.

$$r^2 = \frac{\sum(\hat{Y} - \bar{Y})^2}{\sum(Y - \bar{Y})^2}$$

$$r^2 = \frac{1,501074522}{1,75848}$$

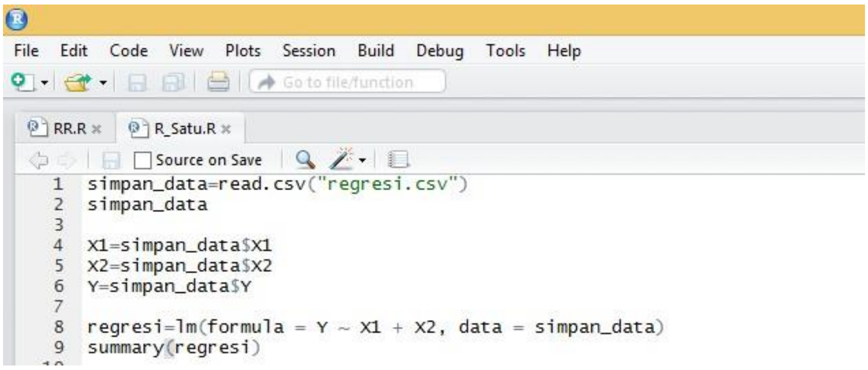
$$r^2 = 0,85362$$

Nilai-nilai yang terkait dalam perhitungan $r^2$ diberikan pada Tabel 4.

Tabel 4. Nilai-nilai yang terkait dalam perhitungan $r^2$

| | $Y$ | $\hat{Y}$ | $(\hat{Y} - \bar{Y})^2$ | $(Y - \bar{Y})^2$ |
|---|---|---|---|---|
| | 3.01 | 3.050862 | 0.030673257 | 0.046656 |
| | 3.15 | 3.050862 | 0.030673257 | 0.005776 |
| | 2.9 | 3.02271 | 0.041326999 | 0.106276 |
| | 3.1 | 3.069948 | 0.024352219 | 0.015876 |
| | 2.7 | 2.851413 | 0.140315331 | 0.276676 |
| | 3.25 | 3.174444 | 0.00265802 | 0.000576 |
| | 3.6 | 3.36435 | 0.019140764 | 0.139876 |
| | 3.7 | 3.459779 | 0.054652798 | 0.224676 |
| | 3.65 | 3.621057 | 0.156069832 | 0.179776 |
| | 3.15 | 3.10812 | 0.013895762 | 0.005776 |
| | 3.02 | 3.050862 | 0.030673257 | 0.042436 |
| | 3.16 | 3.263671 | 0.001419129 | 0.004356 |
| | 2.95 | 2.92728 | 0.089233434 | 0.076176 |
| | 3.12 | 3.069948 | 0.024352219 | 0.011236 |
| | 2.8 | 2.937299 | 0.083348022 | 0.181476 |
| | 3.3 | 3.231702 | 3.25077E-05 | 0.005476 |
| | 3.57 | 3.6306 | 0.163700891 | 0.118336 |
| | 3.64 | 3.80142 | 0.331108128 | 0.171396 |
| | 3.6 | 3.725553 | 0.249552936 | 0.139876 |
| | 3.15 | 3.10812 | 0.013895762 | 0.005776 |
| Jumlah | 64.52 | 64.52 | 1.501074522 | 1.75848 |
| Rata-Rata | 3.226 | 3.226 | 0.075053726 | 0.087924 |

Gambar 17 dan Gambar 18 merupakan hasil perhitungan dengan R. Pada Gambar 18, nilai koefisien determinasi (***R-squared***) bernilai 0,8536. Nilai tersebut dapat diinterpretasikan variabel jumlah jam belajar dan uang jajan mampu menjelaskan atau menerangkan *variation* dari variabel IP sebesar 85,36%, sisanya sebesar 14,64% dijelaskan oleh variabel atau faktor lain.




**Gambar 17**

```
X1=simpan_data$X1
X2=simpan_data$X2
Y=simpan_data$Y

regresi=lm(formula = Y ~ X1 + X2, data = simpan_data)
summary(regresi)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2, data = simpan_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.16142 -0.10843  0.02583  0.05461  0.24022
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.87230    0.17730  10.560 6.93e-09 ***
## X1           0.10450    0.01073   9.742 2.27e-08 ***
## X2           0.01909    0.01555   1.227    0.236
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1231 on 17 degrees of freedom
## Multiple R-squared:  0.8536, Adjusted R-squared:  0.8364
## F-statistic: 49.57 on 2 and 17 DF,  p-value: 8.065e-08
```

**Gambar 18**

### 8. Menguji Kecocokan Persamaan Regresi Linear Terhadap Data dengan Uji F

Uji $F$ digunakan untuk menguji apakah persamaan regresi linear yang telah diperoleh benar- benar bermakna atau signifikan secara statistika (*statistically significant*) mampu, dalam hal mencocokkan data. Hipotesis nol menyatakan bahwa kemampuan persamaan regresi linear dalam mencocokkan data tidak signifikan. Dengan kata lain, kemampuan variabel-variabel bebas secara simultan atau bersamaan dalam menjelaskan *variation* variabel tak bebas tidak signifikan. Secara matematis, untuk hipotesis nol dapat dinyatakan dalam persamaan sebagai berikut.

$$H_0: \beta_1 = \beta_2 = \beta_3 = \cdots = \beta_m = 0.$$

Pada persamaan di atas berarti seluruh koefisien regresi populasi dari variabel bebas bernilai

0. Perhatikan bahwa$m$ menyatakan jumlah variabel bebas yang digunakan dalam persamaan regresi linear. Hipotesis alternatif menyatakan bahwa kemampuan persamaan regresi linear dalam mencocokkan data signifikan secara statiska mampu menjelaskan *variation* dari variabel bebas.

Nilai statistik dari uji $F$ ($F_{hitung}$) dibandingkan dengan nilai kritis $F$ ($F_{kritis}$) untuk pengambilan keputusan terhadap hipotesis. Untuk menentukan nilai kritis $F$, terlebih dahulu menghitung nilai derajat bebas pembilang (*numerator*) dan derajat bebas penyebut (*denominator*). Derajat bebas pembilang dan derajat bebas penyebut dihitung dengan rumus sebagai berikut.

$Derajat\ bebas\ pembilang = k - 1.$

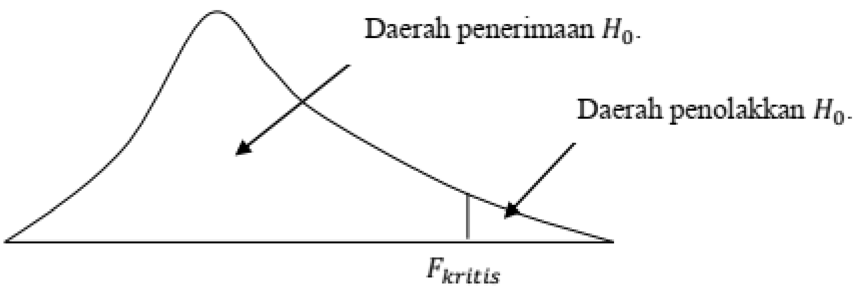
$Derajat\ bebas\ penyebut = n - k.$

Perhatikan bahwa $k$ menyatakan jumlah variabel, sedangkan $n$ menyatakan jumlah pengamatan atau elemen dalam sampel. Berikut aturan pengambilan keputusan berdasarkan uji $F$.

$Jika\ F_{hitung} \leq F_{kritis}, maka\ \ H_0\ diterima\ dan\ H_1 ditolak.$

$Jika\ F_{hitung} > F_{kritis}, maka\ \ H_0\ ditolak\ dan\ H_1 diterima.$

Gambar 19 menyajikan daerah keputusan untuk uji $F$.



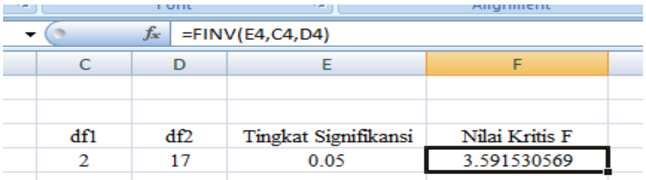

**Gambar 19**

Pengambilan keputusan terhadap hipotesis juga dapat dilakukan dengan menggunakan pendekatan nilai probabilitas dari uji F. Nilai probabilitas dari uji F dibandingkan dengan tingkat signifikansi yang digunakan. Berikut aturan pengambilan keputusan terhadap hipotesis berdasarkan pendekatan nilai probabilitas.

$Jika\ nilai\ probabilitas \geq tingkat\ signifikansi,\ maka\ H_0\ diterima\ dan\ H_1\ ditolak.$

$Jika\ nilai\ probabilitas < tingkat\ signifikansi,\ maka\ H_0\ ditolak\ dan\ H_1\ diterima.$

Diketahui nilai statistik dari uji $F$ (*F-statistic*) adalah 49,57 (perhatikan Gambar 3.28). Diketahui jumlah pengamatan atau elemen dalam sampel adalah $n = 20$ dan jumlah variabel adalah $k = 3$. Maka nilai derajat bebas pembilang adalah $k - 1 = 3 - 1 = 2$ dan nilai derajat bebas penyebut adalah $n - k = 20 - 3 = 17$. Nilai kritis $F$ dengan derajat bebas pembilang 2, derajat bebas penyebut 17, dan tingkat signifikansi 5% adalah 3,59.

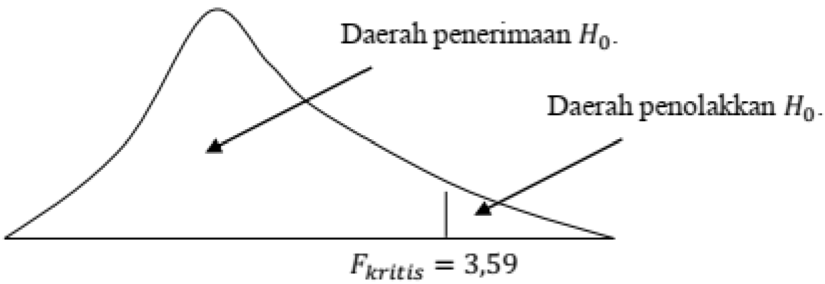

=FINV(E4,C4,D4)

| C | D | E | F |
|---|---|---|---|
| df1 | df2 | Tingkat Signifikansi | Nilai Kritis F |
| 2 | 17 | 0.05 | 3.591530569 |

**Gambar 20 Menentukan Nilai Kritis F dengan *Microsoft Excel***

Berikut aturan pengambilan keputusan terhadap hipotesis berdasarkan uji $F$.

$$Jika\ F_{hitung} \leq F_{kritis},\ maka\ H_0\ diterima\ dan\ H_1\ \text{ditolak}.$$
$$jika\ F_{hitung} > F_{kritis},\ maka\ H_0\ ditolak\ dan\ H_1\ diterima.$$




**Gambar 21**

Perhatikan bahwa karena nilai statistik dari uji $F$, yakni 49,57 lebih besar dibandingkan nilai kritis $F$, maka hipotesis nol ditolak dan hipotesis alternatif diterima. Hal ini berarti persamaan regresi linear yang dihasilkan signifikan secara statistika mampu dalam hal mencocokkan data.

Pengambilan keputusan terhadap hipotesis juga dapat dilakukan dengan menggunakan pendekatan nilai probabilitas dari uji $F$. Nilai probabilitas dari uji $F$ dibandingkan dengan tingkat signifikansi yang digunakan. Berikut aturan pengambilan keputusan terhadap hipotesis berdasarkan nilai probabilitas.

$$Jika\ nilai\ probabilitas \geq tingkat\ signifikansi, maka\ H_0\ diterima\ dan\ H_1\ \ ditolak.$$
$$Jika\ nilai\ probabilitas < tingkat\ signifikansi, maka\ H_0\ ditolak\ dan\ H_1\ \ diterima.$$

Berdasarkan Gambar 21, diketahui nilai probabilitas (*p-value* atau *probability-value*) adalah

$$8.065e-08 = \frac{8.065}{10^8} = 0{,}0000000865.$$

Karena nilai probabilitas tersebut lebih kecil dibandingkan $\alpha = 0{,}05$, maka hipotesis nol ditolak dan hipotesis alternatif diterima. Berikut rumus untuk menghitung nilai statistik dari uji $F$.

$$F_{hitung} = \frac{\frac{r^2}{k-1}}{\frac{1-r^2}{n-k}}$$

Sehingga nilai statistik dari uji $F$ diperoleh sebagai berikut.

$$F = \frac{\frac{0{,}85362}{3-1}}{\frac{1-0{,}85362}{20-3}}$$

$$F = 49{,}56804208$$

**9. Uji signifikansi Koefisien Regresi Secara Individu dengan Uji t.**

Dalam regresi linear berganda, uji $t$ digunakan untuk menguji signifikansi dari masing-masing koefisien regresi populasi. Signifikansi koefisien regresi populasi diuji berdasarkan koefisien regresi sampel. Berikut perumusan hipotesis untuk uji signifikansi koefisien regresi secara individu.

$H_0: \beta_i = 0$

$H_1: \beta_i \neq 0$

Perhatikan bahwa hipotesis nol menyatakan koefisien regresi populasi ke-$i$ ($\beta_i$) bernilai nol. Dengan kata lain, variabel bebas ke-$i$ memiliki pengaruh yang tidak signifikan secara statistika terhadap variabel tak bebas, dengan mengontrol pengaruh dari variabel bebas lain. Hipotesis alternatif menyatakan koefisien regresi populasi ke-$i$ ($\beta_i$) tidak bernilai nol. Dengan kata lain, variabel bebas ke-$i$ memiliki pengaruh yang signifikan secara statistika terhadap variabel tak bebas, dengan mengontrol pengaruh dari variabel bebas lain.

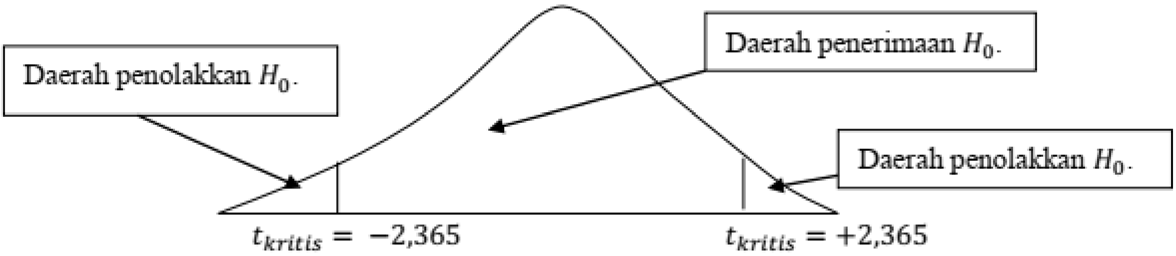
Untuk pengambilan keputusan terhadap hipotesis, dapat dilakukan dengan membandingkan nilai statistik dari uji $t$ ($t_{hitung}$) terhadap nilai kritis $t$ ($t_{kritis}$). Sebelum menghitung nilai kritis $t$, terlebih dahulu menghitung nilai derajat. Berikut rumus untuk menghitung nilai derajat bebas.

$$Derajat\ bebas\ = n - k.$$

Perhatikan bahwa $n$ menyatakan jumlah pengamatan atau elemen dalam sampel, sedangkan $k$ merupakan jumlah variabel. Andaikan jumlah pengamatan atau elemen dalam sampel sebanyak 10 dan jumlah variabel adalah 3 (jumlah variabel bebas adalah 2 dan variabel tak bebas adalah 1), sehingga derajat bebas adalah $10 - 3 = 7$. Misalkan tingkat signifikansi yang digunakan adalah 5%, sehingga nilai kritis $t$ dengan derajat bebas 7 dan tingkat signifikansi 5% adalah $\pm 2{,}365$. Berikut aturan pengambilan keputusan terhadap hipotesis berdasarkan uji $t$.

$$Jika\ |t_{hitung}| \leq |t_{kritis}|, maka\ H_0\ diterima\ dan\ H_1 ditolak.$$
$$Jika\ |t_{hitung}| > |t_{kritis}|, maka\ H_0\ ditolak\ dan\ H_1 diterima.$$

Pengambilan keputusan terhadap hipotesis juga dapat dilakukan dengan menggunakan pendekatan nilai probabilitas dari uji $t$. Nilai probabilitas dari uji $t$ dibandingkan dengan tingkat signifikansi yang digunakan. Berikut aturan pengambilan keputusan terhadap hipotesis berdasarkan pendekatan nilai probabilitas.




**Gambar 22**

$Jika\ nilai\ probabilitas \geq tingkat\ signifikansi, maka\ H_0\ diterima\ dan\ H_1\ ditolak.$

$Jika\ nilai\ probabilitas < tingkat\ signifikansi,\ maka\ H_0\ ditolak\ dan\ H_1\ diterima.$

Berikut akan ditentukan apakah faktor jumlah jam belajar dalam sehari mempengaruhi IP secara signifikan (signifikan secara statistika), dengan mengontrol pengaruh uang jajan dalam sehari. *Output* R pada Gambar 18 disajikan kembali pada Gambar 23.

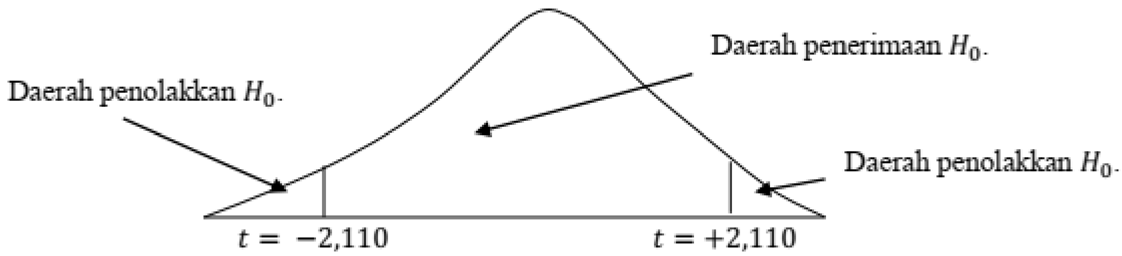
```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.87230    0.17730  10.560 6.93e-09 ***
## X1           0.10450    0.01073   9.742 2.27e-08 ***
## X2           0.01909    0.01555   1.227    0.236
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1231 on 17 degrees of freedom
## Multiple R-squared:  0.8536, Adjusted R-squared:  0.8364
## F-statistic: 49.57 on 2 and 17 DF,  p-value: 8.065e-08
```

**Gambar 23**

Diketahui nilai statistik dari uji $t$ untuk variabel jumlah jam belajar dalam sehari $(X_2)$ adalah 9,742. Nilai kritis $t$ dengan derajat bebas $n - k = 20 - 3 = 17$ dan tingkat signifikansi 5% adalah $\pm 2{,}110$. Berikut aturan pengambilan keputusan terhadap hipotesis berdasarkan uji $t$.

$Jika\ |t_{hitung}| \leq\ |t_{kritis}|, maka\ H_0\ diterima\ dan\ H_1 ditolak.$

$Jika\ |t_{hitung}| >\ |t_{kritis}|, maka\ H_0\ ditolak\ dan\ H_1 diterima.$




**Gambar 24**

Perhatikan bahwa karena $|t_{hitung}| > |t_{kritis}|$, yakni $9{,}742 > 2{,}110$, maka disimpulkan bahwa faktor jumlah jam belajar memiliki pengaruh yang signifikan (signifikan secara statistika) terhadap IP, dengan mengontrol pengaruh uang jajan dalam sehari.

Pengambilan keputusan terhadap hipotesis juga dapat dilakukan dengan menggunakan pendekatan nilai probabilitas dari uji $t$. Nilai probabilitas dari uji $t$ dibandingkan dengan tingkat signifikansi yang digunakan. Berikut aturan pengambilan keputusan terhadap hipotesis berdasarkan nilai probabilitas.

$Jika\ nilai\ probabilitas \geq tingkat\ signifikansi, maka\ H_0\ diterima\ dan\ H_1\ ditolak.$

$Jika\ nilai\ probabilitas < tingkat\ signifikansi, maka\ H_0\ ditolak\ dan\ H_1\ diterima.$

Nilai probabilitas dari uji $t$ berdasarkan variabel jumlah jam belajar dalam sehari (lihat kolom Pr(> |t|)) adalah

$$2.27e - 08 = \frac{2.27}{10^8} = 0.0000000227.$$

Karena nilai probabilitas tersebut lebih kecil dibandingkan tingkat signifikansi $\alpha = 5\%$, maka disimpulkan bahwa faktor jumlah jam belajar dalam sehari mempengaruhi IP secara signifikan (signifikan secara statistika), dengan mengontrol pengaruh uang jajan dalam sehari. Diketahui nilai koefisien regresi untuk variabel jumlah jam belajar dalam sehari adalah 0,10450. Nilai tersebut dapat diinterpretasikan ketika jumlah jam belajar dalam sehari ditambah satu jam, maka diharapkan (*expected*) nilai indeks prestasi meningkat sebesar 0,1045, ketika pengaruh dari uang jajan dalam sehari dipertahankan konstan.

Selanjutnya akan ditentukan apakah faktor uang jajan dalam sehari $(X_2)$ mempengaruhi IP secara signifikan (signifikan secara statistika), dengan mengontrol pengaruh jumlah jam belajar dalam sehari. Perhatikan bahwa karena $|t_{hitung}| \leq |t_{kritis}|$, yakni $1{,}227 < 2{,}110$, maka disimpulkan bahwa faktor uang jajan dalam sehari tidak mempengaruhi IP secara signifikan (signifikan secara statistika), dengan mengontrol pengaruh uang jajan dalam sehari. Dengan kata lain, pengaruh yang diberikan oleh faktor uang jajan dalam sehari terhadap IP sangat lemah, dengan mengontrol pengaruh jumlah jam belajar dalam sehari. Nilai probabilitas dari uji $t$ berdasarkan variabel uang jajan dalam sehari adalah 0,236. Karena nilai probabilitas tersebut lebih besar dibandingkan tingkat signifikansi $\alpha = 5\%$, maka disimpulkan bahwa faktor uang jajan dalam sehari tidak mempengaruhi IP secara signifikan (signifikan secara statistika), dengan mengontrol pengaruh jumlah jam belajar dalam sehari.

## C. Studi Kasus

## D. Latihan Mandiri

Misalnya, kita akan meneliti variasi berat badan (Y) dikaitkan dengan tinggi badan($X_1$) dan umur ($X_2$) untuk anak-anak yang memiliki masalah atau kekurangan gizi. Andaikan bahwa sebuah sampel acak terdiri 12 anak yang mengunjungi klinik. Data berat (kg), tinggi (cm), dan umur (tahun) diperoleh untuk setiap anak dan diberikan pada tabel berikur ini (taraf kesignifikanan 5% atau 0,05)

| Anak | Y (kg) | $X_1$ (cm) | $X_2$ (thn) |
|---|---|---|---|
| 1 | 32 | 57 | 8 |
| 2 | 36 | 59 | 10 |
| 3 | 27 | 49 | 6 |
| 4 | 34 | 62 | 11 |
| 5 | 28 | 51 | 8 |
| 6 | 29 | 50 | 7 |
| 7 | 39 | 55 | 10 |

| | | | |
|---|---|---|---|
| 8 | 29 | 48 | 9 |
| 9 | 28 | 42 | 10 |
| 10 | 26 | 42 | 6 |
| 11 | 38 | 61 | 12 |
| 12 | 39 | 57 | 9 |

Lakukan tahap tahap analisis regresi seperti yang telah dijelaskan pada modul dengan menggunakan program R!

**E. Rangkuman**

Modul ini telah memperkenalkan tahapan dalam analisis regresi baik secara teori maupun data real. Modul ini juga telah dilengkapi dengan beberapa studi kasus beserta tahapan pengolahan datanya dengan menggunakan software R.

**F. Daftar Pustaka**

1. Agresti, A. dan B. Finlay. 2009. *Statistical Methods for the Social Sciences, 4th Edition*. United States of America: Prentice Hall.

2. Field, A. 2009. *Discovering Statistics Using SPSS, 3rd Edition*. London: Sage.

3. Gio, P.U. dan E. Rosmaini, 2015. Belajar Olah Data dengan SPSS, Minitab, R, Microsoft Excel, EViews, LISREL, AMOS, dan SmartPLS. USUpress.

4. Gujarati, D.N. 2003. *Basic Econometrics, 4th Edition*. New York: McGraw-Hill.

5. Hair, J.F Jr., R.E. Anderson, B.J. Babin, dan W.C. Black. 2010. *Multivariate Data Analysis, 7th Edition*. Pearson Prentice Hall.

6. Johnson, R.A. dan D.W. Wichern. 2007. *Applied Multivariate Statistical Analysis, 6th Edition*. United States of America: Prentice Hall.

7. Malhotra, N.K. dan D.F. Birks. 2006. *Marketing Research, An Applied Approach, 2nd European Edition*. London: Prentice Hall.

8. Montgomery, D.C. dan G.C. Runger. 2011. *Applied Statistics and Probability for Engineers, 5th Edition*. United States of America: John Wiley & Sons, Inc.

9. Stevens, J.P. 2009. *Applied Multivariate Statistics For The Social Science, 5th Edition*. New York: Routledge.

10. Supranto, J. 2004. Ekonometri, Buku Kedua. Jakarta: Ghalia Indonesia.

11. Supranto, J. 2005. Ekonometri, Buku Kesatu. Jakarta: Ghalia Indonesia.