

## BAB IX

### KORELASI DAN REGRESI LINEAR SEDERHANA

#### Capaian Pembelajaran

- Mahasiswa mengerti dan memahami hubungan linear antara kedua peubah dalam kaitannya dengan regresi linear sederhana adalah peubah bebas dan peubah tak bebas
- Mahasiswa mampu untuk menggambarkan tingkat keeratan hubungan linear antara dua peubah atau lebih
- Mahasiswa mampu membentuk model regresi linear sederhana dan menginterpretasikan model regresi yang dihasilkan
- Mahasiswa dapat menentukan kesesuaian model regresi yang dihasilkan.

#### 9.1 Korelasi

Korelasi dan regresi adalah dua konsep penting dalam statistik yang sering digunakan untuk memahami hubungan antara dua atau lebih variabel. Korelasi mengukur hubungan linier antara dua variabel. Nilai korelasi berkisar antara -1 dan 1, di mana -1 menunjukkan hubungan negatif yang kuat, 1 menunjukkan hubungan positif yang kuat, dan 0 menunjukkan bahwa tidak ada hubungan antar dua variabel. Sedikit berbeda dengan korelasi, regresi juga digunakan untuk mengukur hubungan antar variabel, namun hubungan yang diukur adalah hubungan sebab akibat. Regresi memungkinkan kita untuk memprediksi nilai satu variabel dari nilai variabel lain. Ini biasanya dilakukan dengan membangun model matematis yang menggambarkan hubungan antara variabel tersebut. Dengan memahami konsep korelasi dan regresi dan bagaimana mereka berkaitan satu sama lain, metode regresi dapat digunakan untuk memahami dan memprediksi hubungan antara variabel dalam berbagai bidang, seperti bisnis, ekonomi, ilmu sosial, dan lainnya.

Rumus korelasi antara dua variabel X dan Y adalah:

$$r = \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{[n \sum X^2 - (\sum X)^2][\sum Y^2 - (\sum Y)^2]}}$$

di mana

- $n$  : Jumlah observasi,  
 $X$  dan  $Y$  : Nilai dari masing-masing variabel,  
 $\sum X$  dan  $\sum Y$  : Jumlah dari masing-masing variabel,  
 $\sum XY$  : Jumlah produk dari  $X$  dan  $Y$ ,

$\Sigma X^2$  dan  $\Sigma Y^2$  : Jumlah kuadrat dari masing-masing variabel.

## 9.2 Persamaan Regresi Linier Sederhana

Dalam regresi linier sederhana, model adalah garis yang menghubungkan titik-titik data. Garis sederhana hanya cocok untuk memodelkan hubungan linier antara dua variabel. Dalam kasus di mana hubungan antara variabel tidak linier, regresi non-linier atau regresi polinomial mungkin lebih cocok. Dalam regresi berganda, kita dapat memodelkan hubungan antara lebih dari dua variabel. Ini bisa digunakan untuk memprediksi variabel tergantung dari lebih dari satu variabel independen. Sebelum melakukan regresi, penting untuk melakukan analisis korelasi untuk memastikan bahwa ada hubungan antara variabel yang ingin kita modelkan. Jika tidak ada hubungan, maka regresi akan memberikan hasil yang tidak berguna. Dalam analisis regresi, ada beberapa asumsi yang harus dipenuhi agar hasil analisis dapat diinterpretasikan dengan benar. Salah satu asumsi yang paling penting adalah asumsi linearitas, yaitu bahwa hubungan antara variabel harus linier. Asumsi lain meliputi homoskedastisitas (varian yang sama untuk setiap nilai dari variabel independen), independensi observasi/non-multikolinieritas (tidak ada hubungan antar observasi), dan normalitas (distribusi residual harus berdistribusi normal). Setelah asumsi dipenuhi, model regresi dapat dibuat dan diuji untuk melihat apakah memenuhi syarat.

Persamaan model regresi linier sederhana adalah:

$$Y = b_0 + b_1X$$

di mana:

$b_0$  : Intersep,

$b_1$  : Koefisien regresi,

$X$  : Variabel independen.

Koefisien regresi,  $b_1$ , dapat ditemukan dengan rumus:

$$b_i = r \times \frac{S_Y}{S_X}$$

di mana:

$r$  : Korelasi antara X dan Y,

$S_Y$  : Standar deviasi dari Y, dan

$S_X$  : Standar deviasi dari X.

Intersep,  $b_0$ , dapat ditemukan dengan rumus:

$$b_0 = \bar{Y} - b_1\bar{X}$$

di mana:

$\bar{Y}$  : rata-rata dari Y,

$\bar{X}$  : rata-rata dari  $X$ .

Dengan menggunakan rumus ini, kita dapat membuat model regresi dan memprediksi nilai dari variabel dependen berdasarkan nilai variabel independen.

### Uji Serempak dan Uji Individual

Uji serempak adalah uji hipotesis yang menguji apakah seluruh variabel independen secara bersama-sama memiliki pengaruh signifikan terhadap variabel dependen.

$H_0: \beta_1 = 0$  (Variabel bebas tidak mempengaruhi variabel tak bebas),

$H_1: \beta_1 \neq 0$  (Variabel bebas memiliki pengaruh terhadap variabel tak bebas).

Statistik  $F$  digunakan untuk uji serempak, statistik  $F$  dihitung dengan membagi varians dari *mean square error* (MSE) dengan varians *mean square* dari model (MSM).

Uji parsial adalah uji hipotesis yang menguji apakah setiap variabel independen secara terpisah memiliki pengaruh signifikan terhadap variabel dependen.

$H_0: \beta_1 = 0$  (tidak ada hubungan signifikan antara variabel independen dan dependen)

$H_1: \beta_1 \neq 0$  (ada hubungan signifikan antara variabel independen dan dependen)

Statistik  $t$  digunakan untuk uji parsial. Statistik  $t$  dihitung dengan membagi nilai estimasi dari koefisien regresi oleh standar deviasi dari estimasi. Kedua jenis uji hipotesis ini memiliki tujuan yang berbeda, tetapi hasil dari kedua uji ini seringkali digabungkan untuk memperoleh gambaran yang lebih komprehensif tentang pengaruh variabel independen terhadap variabel dependen. Kedua statistik uji (uji  $F$  dan uji  $t$ ) dibandingkan dengan nilai-nilai dari tabel  $F$  atau  $t$  pada tingkat signifikansi tertentu untuk memutuskan apakah hipotesis nol dapat diterima atau ditolak. Jika nilai statistik lebih besar dari nilai tabel, maka hipotesis nol ditolak dan hipotesis alternatif diterima, yang berarti ada pengaruh signifikan antara variabel independen dan dependen.

### 9.3 Aplikasi Menggunakan R

Terdapat banyak dataset yang disediakan oleh *R* untuk diuji hubungannya menggunakan metode korelasi dan dimodelkan menggunakan regresi. Salah satunya adalah *marketing dataset* dalam *package datarium* yang memuat informasi mengenai dampak dari tiga media periklanan (*youtube*, *facebook*, dan surat kabar) terhadap penjualan. Data terdiri dari data anggaran iklan (variabel independen) dan data penjualan (variabel dependen) dalam ribuan dolar. Berikut adalah `summary` dari data.

```
> # Load the package
> data("marketing", package = "datarium")
> head(marketing, 4)
```

	youtube	facebook	newspaper	sales
1	276.12	45.36	83.04	26.52
2	53.40	47.16	54.12	12.48
3	20.64	55.08	83.16	11.16
4	181.80	49.56	70.20	22.20

Sebelum melakukan analisis korelasi dan regresi, sebaiknya dilihat dulu pola hubungan antar variabel melalui *scatter plot*. Berikut adalah pola hubungan antara variabel penjualan dan youtube.