

PRAKTIKUM 13

ANALISIS REGRESI LINIER SEDERHANA

DESKRIPSI SINGKAT

Regresi linier sederhana adalah teknik statistik yang digunakan untuk mempelajari hubungan antara dua variabel. Dalam regresi linier sederhana, kita mencari hubungan antara variabel independen (x) dan variabel dependen (y) dengan menggunakan garis lurus terbaik yang meminimalkan jarak antara titik data dan garis lurus. Dalam artikel ini, kita akan membahas tata cara melakukan regresi linier sederhana menggunakan software R/R Studio.

LANGKAH-LANGKAH REGRESI LINIER SEDERHANA

Berikut adalah langkah-langkah untuk melakukan regresi linier sederhana di R/R Studio:

| |
|--|
| Import Data |
| Pertama-tama, impor data yang ingin dianalisis. Data dapat diimpor dalam berbagai format seperti <code>.csv</code> , <code>.txt</code> , <code>.xlsx</code> , atau <code>.rds</code> . Jika data sudah berada dalam format yang sesuai, kita dapat menggunakan fungsi <code>read.csv()</code> , <code>read.table()</code> , <code>read.xlsx()</code> , atau <code>readRDS()</code> untuk memuat data |
| Visualisasi Data |
| Setelah data diimpor, periksa dan visualisasikan data tersebut menggunakan scatter plot atau histogram. Hal ini penting untuk memeriksa apakah data kita berdistribusi normal atau tidak. |
| Regresi Linier Sederhana |
| Lakukan regresi linier sederhana dengan menggunakan fungsi <code>lm()</code> . Fungsi ini akan membentuk garis lurus terbaik yang meminimalkan jarak antara titik data dan garis lurus. |
| Analisis Hasil Regresi |
| Setelah melakukan regresi linier sederhana, kita dapat menganalisis hasil regresi dengan menggunakan fungsi <code>summary()</code> . Fungsi ini akan memberikan informasi tentang koefisien regresi, intercept, nilai p, dan lain-lain. |
| Plot Regresi |
| Terakhir, plot garis regresi menggunakan fungsi <code>plot()</code> dan <code>abline()</code> . |

OUTPUT ANALISIS REGRESI LINIER SEDERHANA

Output analisis regresi linier sederhana dapat memberikan informasi tentang kualitas model dan signifikansi hubungan antara variabel independen dan dependen. Beberapa output yang umum ditemukan dalam analisis regresi linier sederhana adalah:

| | |
|----------------------------------|---|
| Koefisien Regresi (slope) | Menunjukkan seberapa besar perubahan nilai variabel dependen ketika nilai variabel independen berubah satu satuan. |
| Intercept | Menunjukkan nilai dari variabel dependen ketika nilai variabel independen sama dengan nol. |
| R-squared | Menunjukkan seberapa baik model kita cocok dengan data. Nilai r-squared berkisar antara 0 dan 1, dengan nilai yang lebih tinggi menunjukkan model yang lebih baik. |
| Nilai p | Menunjukkan signifikansi hubungan antara variabel independen dan dependen. Nilai p yang rendah menunjukkan bahwa hubungan antara variabel independen dan dependen adalah signifikan secara statistik. |

SOAL

Contoh kasus yang digunakan adalah untuk menganalisis hubungan antara curah hujan dengan penjualan payung. Curah hujan dijadikan variabel bebas (X) dan penjualan payung dijadikan variabel terikat (Y). Tujuan analisis ini adalah untuk mengetahui seberapa besar pengaruh curah hujan terhadap penjualan payung.

| Curah Hujan (X) | Penjualan Payung (Y) |
|-----------------|----------------------|
| 3.2 | 20 |
| 4.5 | 25 |
| 2.1 | 15 |
| 5.7 | 30 |
| 6.3 | 35 |
| 2.8 | 18 |
| 4.1 | 24 |
| 5.5 | 32 |
| 1.9 | 12 |
| 3.8 | 22 |
| 4.6 | 26 |
| 2.5 | 16 |
| 6.1 | 36 |
| 3.6 | 21 |
| 5.2 | 28 |
| 2.3 | 14 |
| 3.9 | 23 |
| 4.8 | 27 |
| 2.9 | 19 |
| 6.5 | 38 |
| 5.4 | 31 |
| 4.4 | 25 |
| 3.5 | 20 |
| 2.2 | 14 |
| 5.8 | 33 |
| 3.3 | 19 |
| 4.9 | 28 |
| 6.2 | 36 |
| 2.7 | 17 |
| 4.3 | 24 |

SOLUSI

Contoh kasus yang digunakan adalah untuk menganalisis hubungan antara curah hujan dengan penjualan payung. Curah hujan dijadikan variabel bebas (X) dan penjualan payung dijadikan variabel terikat (Y). Tujuan analisis ini adalah untuk mengetahui seberapa besar pengaruh curah hujan terhadap penjualan payung.

Langkah-langkah untuk melakukan analisis regresi linier sederhana di R/R Studio adalah sebagai berikut:

Memasukkan data ke dalam R/R Studio

Pertama, kita perlu memasukkan data curah hujan dan penjualan payung ke dalam R/R Studio.

Untuk melakukan hal ini, kita dapat menulis script berikut:

```
> curah_hujan <- c(3.2, 4.5, 2.1, 5.7, 6.3, 2.8, 4.1, 5.5, 1.9,
3.8, 4.6, 2.5, 6.1, 3.6, 5.2, 2.3, 3.9, 4.8, 2.9, 6.5, 5.4, 4.4,
3.5, 2.2, 5.8, 3.3, 4.9, 6.2, 2.7, 4.3)
> penjualan_payung <- c(20, 25, 15, 30, 35, 18, 24, 32, 12, 22,
26, 16, 36, 21, 28, 14, 23, 27, 19, 38, 31, 25, 20, 14, 33, 19,
28, 36, 17, 24)
```

Selain itu, jika data sudah disimpan dalam suatu file, missal excel, maka bisa mengimport dengan script berikut (warna merah melupakan lokasi folder file)

```
> library(readxl)
> Data1 <- read_excel("C:/Users/Data/Folder/Data1.xlsx")
```

Visualisasi Data

Setelah memasukkan data, kita dapat memvisualisasikan data menggunakan plot. Hal ini dapat dilakukan dengan menggunakan script berikut:

```
> plot(curah_hujan, penjualan_payung, main = "Hubungan antara
Curah Hujan dan Penjualan Payung", xlab = "Curah Hujan (mm)", ylab
= "Penjualan Payung")
```

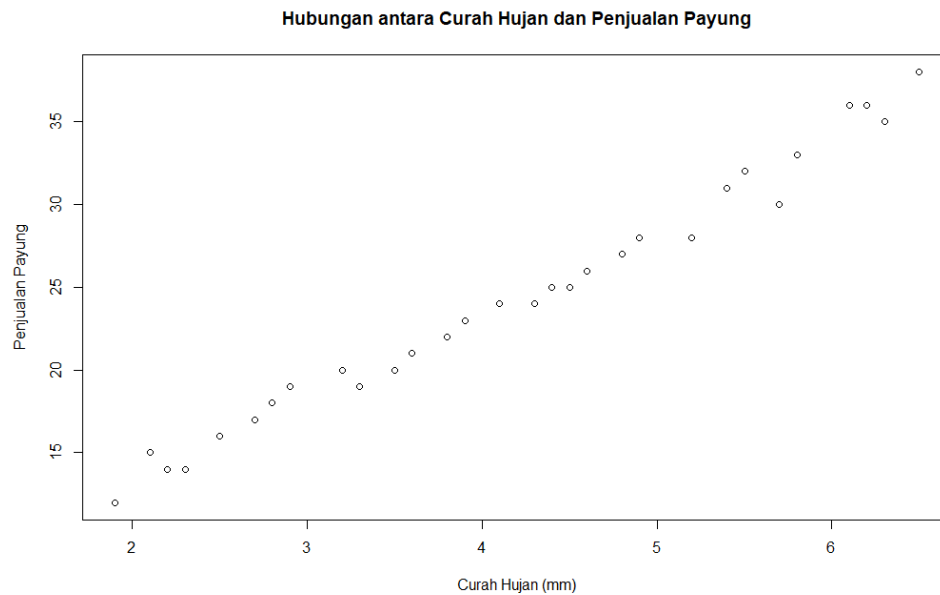
Melakukan Regresi Linier Sederhana

Setelah memvisualisasikan data, langkah selanjutnya adalah melakukan regresi linier sederhana. Kita dapat melakukan regresi linier sederhana dengan menggunakan fungsi `lm()`. Scriptnya adalah sebagai berikut:

```
> model_regresi <- lm(penjualan_payung ~ curah_hujan)
> summary(model_regresi)
```

OUTPUT

Setelah script telah selesai di running / di jalankan, berikut adalah contoh output dari analisis regresi linier sederhana menggunakan Software R/R Studio



```
> model_regresi <- lm(penjualan_payung ~ curah_hujan)
> summary(model_regresi)
```

Call:

```
lm(formula = penjualan_payung ~ curah_hujan)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-2.23880 -0.54954 -0.06048  0.66911  1.68151
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.6033     0.5589   4.658 7.06e-05 ***
curah_hujan    5.1992     0.1274  40.808 < 2e-16 ***
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 0.9567 on 28 degrees of freedom

Multiple R-squared: 0.9835, Adjusted R-squared: 0.9829

F-statistic: 1665 on 1 and 28 DF, p-value: < 2.2e-16

INTERPRETASI OUTPUT

Output tersebut merupakan hasil regresi linier sederhana antara variabel penjualan_payung dan curah_hujan. Berikut adalah interpretasi dari setiap bagian output:

Call

Berisi informasi tentang fungsi yang digunakan untuk melakukan analisis regresi linier sederhana. Dalam hal ini, fungsi `lm()` digunakan untuk memodelkan **penjualan_payung** sebagai fungsi dari **curah_hujan**.

Residuals

Merupakan selisih antara nilai observasi aktual dan nilai prediksi yang dihasilkan oleh model regresi. Output ini memberikan informasi tentang sebaran data residual, yang harus sekitar 0 dengan variansi yang homogen.

Coefficients

Merupakan estimasi koefisien regresi. Pada bagian ini terdapat informasi tentang intercept (konstanta) dan koefisien regresi dari variabel **curah_hujan**. Perhatikan bahwa koefisien **curah_hujan** memiliki nilai 5.1992, yang artinya setiap peningkatan satu unit **curah_hujan** akan meningkatkan **penjualan_payung** sebesar 5.1992.

Signif. codes

Mencantumkan informasi tentang signifikansi statistik dari koefisien regresi. Pada output ini, koefisien intercept memiliki tanda asterisk (*) satu di belakangnya, yang berarti signifikan secara statistik pada level signifikansi 0.05. Sementara koefisien **curah_hujan** memiliki tanda asterisk tiga di belakangnya, yang artinya sangat signifikan secara statistik.

Residual standard error

Mengukur seberapa jauh prediksi model regresi dari nilai observasi aktual. Semakin kecil nilai ini, semakin dekat prediksi model dengan nilai observasi aktual.

Multiple R-squared

Mengukur seberapa besar variasi dari variabel penjualan_payung yang dapat dijelaskan oleh variabel curah_hujan. Pada output ini, nilai R-square sebesar 0.9835, yang artinya 98.35% variasi penjualan_payung dapat dijelaskan oleh variabel curah_hujan.

Adjusted R-squared

Mengukur seberapa besar variasi dari variabel penjualan_payung yang dapat dijelaskan oleh variabel curah_hujan setelah mempertimbangkan jumlah parameter dalam model. Semakin dekat nilai adjusted R-squared dengan nilai R-squared, semakin baik model tersebut.

F-statistic

Mengukur signifikansi secara keseluruhan dari model regresi. Pada output ini, nilai F-statistic sebesar 1665 dan p-value kurang dari $2.2e-16$, yang artinya model regresi tersebut signifikan secara statistik.

Model Regresi Linier Sederhana

Model regresi linier sederhana dari output di atas adalah:

$$y = 2.6033 + 5.1992x$$

Di mana y adalah variabel terikat (penjualan_payung) dan x adalah variabel bebas (curah_hujan).

Interpretasi dari model regresi ini adalah sebagai berikut:

- Intercept (a) adalah nilai y ketika variabel x bernilai nol, dalam hal ini adalah 2.6033. Ini berarti bahwa ketika curah hujan sama dengan nol, penjualan payung diperkirakan sebesar 2.6033.
- Coefficient (b) adalah besarnya perubahan y yang diharapkan ketika x (curah_hujan) mengalami perubahan satu satuan, dalam hal ini adalah 5.1992. Ini berarti bahwa setiap kenaikan satu satuan pada curah hujan, diperkirakan akan meningkatkan penjualan payung sebesar 5.1992 unit.
- Koefisien determinasi (R-squared) menunjukkan seberapa baik model ini dapat menjelaskan variasi dalam data, dalam hal ini adalah 0.9835. Ini berarti sekitar 98.35% variasi dalam data penjualan payung dapat dijelaskan oleh curah hujan.

- P-value (nilai p) menunjukkan signifikansi dari masing-masing koefisien dalam model. Dalam hal ini, kedua koefisien (intercept dan curah_hujan) memiliki nilai p yang sangat rendah, kurang dari 0.05, sehingga keduanya signifikan secara statistik dan dapat dipercaya.

Berdasarkan interpretasi di atas, dapat disimpulkan bahwa model regresi linier sederhana antara variabel penjualan_payung dan curah_hujan tersebut cukup baik, karena memiliki nilai R-squared yang cukup tinggi dan koefisien curah_hujan yang signifikan secara statistik.

Kesimpulan

Artikel ini membahas tentang regresi linier sederhana dan bagaimana melakukan analisis regresi linier sederhana menggunakan software R/R Studio. Langkah-langkah untuk melakukan analisis meliputi impor data, visualisasi data, melakukan regresi linier sederhana, analisis hasil regresi, dan plot regresi. Output analisis regresi linier sederhana yang umum ditemukan adalah koefisien regresi, intercept, R-squared, dan nilai p. Artikel ini juga memberikan contoh kasus tentang analisis hubungan antara curah hujan dan penjualan payung. Diharapkan artikel ini dapat membantu pembaca memperluas pengetahuan dan keterampilan analisis data mereka dengan menggunakan R Studio untuk melakukan analisis regresi linier sederhana.