

PRAKTIKUM 14

ANALISIS REGRESI SEDERHANA LANJUTAN

Uji Analisis Regresi Linier Sederhana

Analisis regresi linier sederhana adalah representasi hubungan linear antara variabel independen (X) dan variabel dependen (Y). Fungsinya adalah untuk mengungkapkan arah hubungan antara variabel bebas dan variabel terikat, baik positif maupun negatif. Selain itu, analisis ini dapat digunakan untuk memperkirakan nilai variabel terikat ketika variabel bebas mengalami perubahan. Regresi sederhana umumnya diterapkan pada data dengan skala interval atau rasio. Rumus regresi linier sederhana digunakan sebagai alat untuk mengukur dan menggambarkan hubungan matematis di antara variabel tersebut.

$$Y = a + bX$$

dimana:

Y = Variabel tak bebas (nilai yang akan diprediksi)

X = Variabel bebas (variabel independent)

a = Konstanta (nilai dari Y apabila X=0)

b = koefisien regresi (pengaruh positif atau negatif)

Analisis Korelasi

Pada penelitian ini, dilakukan analisis korelasi menggunakan metode koefisien korelasi Pearson. Metode ini digunakan untuk mengevaluasi hubungan antara disiplin kerja (X) dan kinerja (Y). Sugiyono (2012) memberikan rumus yang digunakan dalam uji korelasi ini.

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{(n \sum x^2 - (\sum x)^2)(n \sum y^2 - (\sum y)^2)}}$$

dimana:

r = Koefisien korelasi

n = Jumlah data

X = Variabel Bebas (Independen)

Y = Variabel terikat (Dependen)

Uji Asumsi Klasik

Pengecekan asumsi klasik ini merupakan langkah prasyarat yang dilakukan sebelum melakukan analisis lebih lanjut terhadap data yang telah terkumpul. Tujuan dari pengujian asumsi klasik adalah untuk memastikan bahwa model regresi yang dikembangkan memenuhi kriteria BLUE (Best Linear Unbiased Estimator). Agar dapat menentukan apakah model regresi yang akan digunakan memenuhi kriteria BLUE, diperlukan serangkaian pengujian, termasuk Uji Normalitas, Uji Multikolinieritas, Uji Heteroskedastisitas, dan Uji Autokorelasi.

SOURCE CODE

Library yang di butuhkan

```
> library(rmarkdown) #membuat markdown dengan Bahasa R menggunakan
Rstudio
> library(prettydoc) #membuat tampilan dokumen menjadi lebih cantik
> library(tinytex) #membantu fungsi untuk menginstall dan memperbaiki
TeXLive dan meng-compile LaTeX dokumen
> library(car) #pendamping analisis regresi yang diterapkan
> library(tseries) #analisis time series dan komputasi keuangan
> library(lmtest) #pengujian model regresi linier
```

Data

Dataset yang digunakan adalah gaji karyawan berdasarkan lama pengalaman bekerja. Kumpulan data diambil, diurutkan dan dikonversikan ke angka untuk memudahkan dalam melakukan analisis. Data penelitian ini bersumber dari:

<https://www.kaggle.com/datasets/karthickveerakumar/salary-data-simple-linear-regression?resource=download> .

```
> datagaji<- read.csv("Salary_Data.csv", header=TRUE)
> head(datagaji)
  YearsExperience Salary
1             1.1  39343
2             1.3  46205
3             1.5  37731
4             2.0  43525
5             2.2  39891
6             2.9  56642
> str(datagaji)
'data.frame':   30 obs. of  2 variables:
 $ YearsExperience: num  1.1 1.3 1.5 2 2.2 2.9 3 3.2 3.2 3.7 ...
 $ Salary         : num  39343 46205 37731 43525 39891 ...
> paged_table(as.data.frame(datagaji))
```

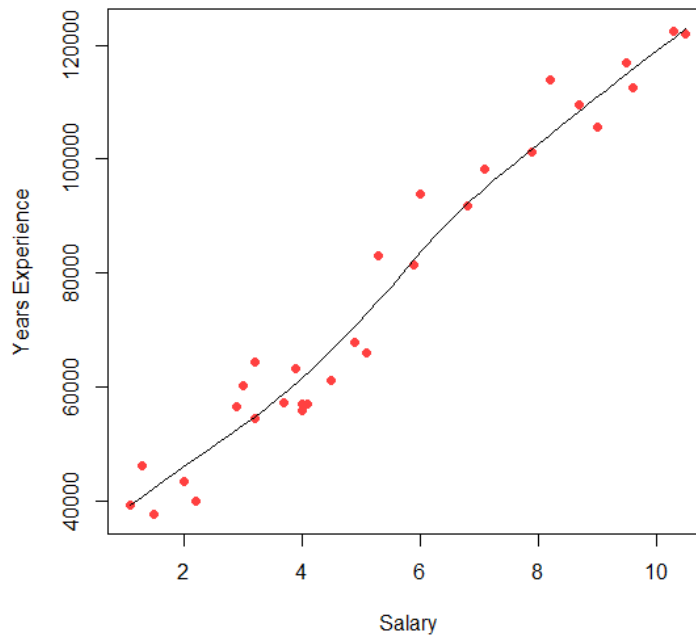
Statistika Deskriptif

```
> summary(datagaji)
```

Scatter Plot

```
> scatter.smooth(x=datagaji$YearsExperience, y=datagaji$Salary,
main="Gambar 1. Smooth Scatter Plot Salary~YearsExperience",
xlab="Salary", ylab="Years Experience", pch=16, col="brown1") #
scatterplot
```

Gambar 1. Smooth Scatter Plot Salary~YearsExperience



Plot pencar bersama dengan garis smoothing di atas menunjukkan hubungan yang meningkat secara linier antara variabel 'YearsExperience' dan 'Salary'. Ini adalah hal yang baik, karena salah satu asumsi yang mendasari regresi linier adalah bahwa hubungan antara variabel respon dan prediktor adalah linier dan aditif.

Box Plot

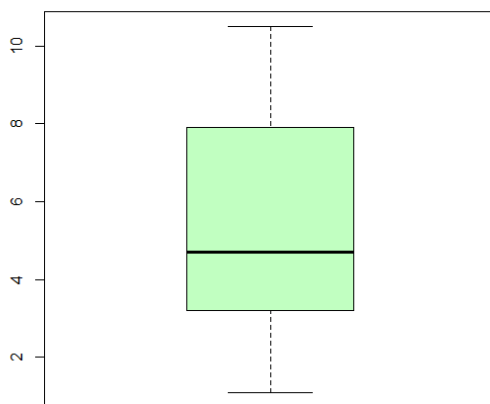
Box plot digunakan untuk menemukan pengamatan outlier dalam variabel. Data yang memiliki outlier di prediktor dapat secara drastis memengaruhi prediksi karena mereka dapat dengan mudah memengaruhi arah/kemiringan garis yang paling sesuai.

```

> par(mfrow=c(1,2)) #Membagi area grafik menjadi 2 kolom
> boxplot(datagaji$YearsExperience, main="Gambar 2.1. Boxplot Years
Experience",sub=paste("Outlier rows: ",
boxplot.stats(datagaji$YearsExperience)$out), col="darkseagreen1")
#box plot untuk YearsExperience
> |
> boxplot(datagaji$Salary, main="Gambar 2.2. Boxplot
Salary",sub=paste("Outlier rows: ",
boxplot.stats(datagaji$Salary)$out), col="darksalmon") #box plot
untuk Salary

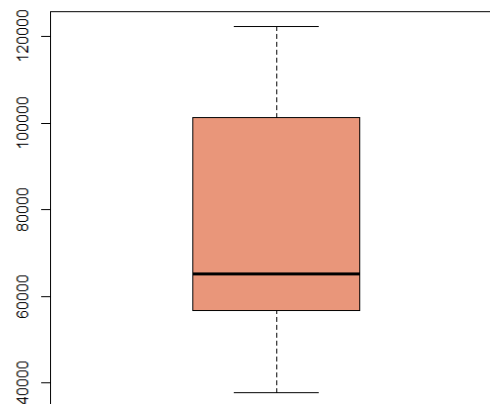
```

Gambar 2.1. Boxplot Years Experience



Outlier rows:

Gambar 2.2. Boxplot Salary



Outlier rows:

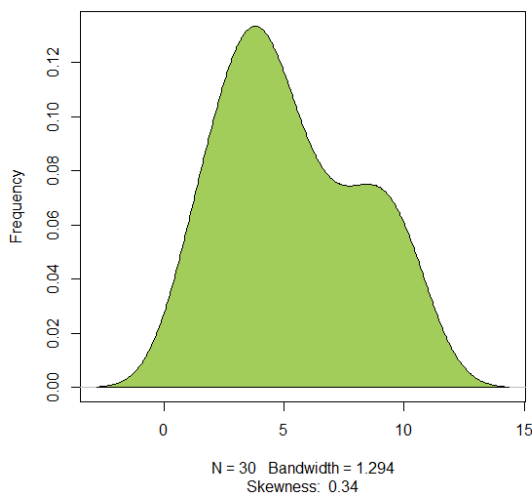
Dari diagram boxplot di atas, diketahui bahwa secara visual dalam data variabel prediktor “Years Experience” dan variabel respons “Salary” tidak setangkup, karena panjang garis whisker tidak sama dan median tidak berada tepat di tengah-tengah kotak. Selain itu, juga dapat diketahui bahwa dari kedua variabel data tidak terdeteksi adanya pencilan atau outlier.

Density Plot

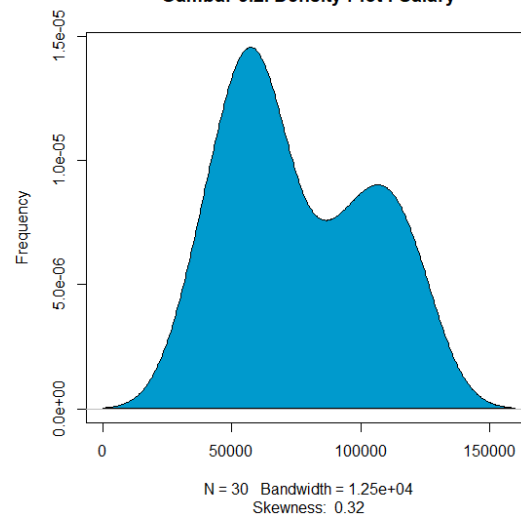
Density plot digunakan untuk melihat distribusi variabel prediktor (memeriksa apakah variabel respons mendekati normalitas). Idealnya, distribusi yang mendekati normal (kurva berbentuk lonceng), tanpa miring ke kiri atau kanan.

```
> par(mfrow=c(1,2)) #Membagi area grafik menjadi 2 kolom
> plot(density(datagaji$YearsExperience), main="Gambar 3.1. Density
Plot : Year Experience", ylab="Frequency", sub=paste("Skewness: ",
round(e1071::skewness(datagaji$YearsExperience),2))) #desity plot
untuk YearsExperience
> polygon(density(datagaji$YearsExperience), col="darkolivegreen3")
>
> plot(density(datagaji$Salary), main="Gambar 3.2. Density Plot :
Salary", ylab="Frequency", sub=paste("Skewness: ",
round(e1071::skewness(datagaji$Salary),2))) #desity plot untuk
Salary
> polygon(density(datagaji$Salary), col="deepskyblue3")
```

Gambar 3.1. Density Plot : Year Experience



Gambar 3.2. Density Plot : Salary



Analisis Korelasi

Analisis korelasi pada kasus ini untuk mengukur kekuatan hubungan antara dua variabel yaitu variabel respons “Salary” dan variabel prediktor “Years Experience” dengan menggunakan metode pearson.

```
> cor(datagaji[1:2], method="pearson") #menghitung korelasi antara
YearsExperience dengan banyaknya gaji
```

	YearsExperience	Salary
YearsExperience	1.0000000	0.9782416
Salary	0.9782416	1.0000000

kuat lemah hubungan diukur diantara jarak (range) dari 0 sampai dengan 1. Dari hasil analisis di atas diketahui bahwa nilai korelasinya 0.9782416, hal ini berarti variabel “YearsExperience” dan variabel “Salary” memiliki hubungan yang sangat kuat. Selain, itu karena koefisien korelasi ditemukan positif (+) maka hubungan tersebut disebut sebagai korelasi sempurna atau hubungan linier sempurna dengan kemiringan (slope) positif.

Analisis Regresi

Analisis untuk meramalkan pengaruh variabel prediktor terhadap variable kriterium untuk membuktikan ada atau tidaknya hubungan fungsional antara variabel bebas (X) dengan sebuah variabel terikat (Y). Setelah melihat hubungan linier secara visual dan dengan menghitung korelasinya, maka langkah selanjutnya yaitu membentuk model analisis regresi sebagai berikut:

Analisis Koefisien Model

```
> X<-datagaji$YearsExperience
> Y<-datagaji$Salary
> regresi<-lm(Y~X, data=datagaji) #membentuk analisis regresi
sederhana
> print(regresi)
Call:
lm(formula = Y ~ X, data = datagaji)
Coefficients:
(Intercept)          X
      25792         9450
```

```
> extract_eq(regresi, use_coefs = TRUE) #membentuk model analisis
regresi sederhana beserta nilai koefisiennya
> summary(regresi) #memanggil koefisien model menggunakan syntax
sebagai berikut.
```

```
Call:
lm(formula = Y ~ X, data = datagaji)

Residuals:
    Min       1Q   Median       3Q      Max
-7958.0 -4088.5 -459.9  3372.6 11448.0

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  25792.2     2273.1   11.35 5.51e-12 ***
X            9450.0       378.8   24.95 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5788 on 28 degrees of freedom
Multiple R-squared:  0.957, Adjusted R-squared:  0.9554
F-statistic: 622.5 on 1 and 28 DF,  p-value: < 2.2e-16
```

- nilai koefisien a sebesar 25792 yang berarti apabila lama pengalaman kerja bernilai 0 maka jumlah gaji akan naik sebesar 25792
- nilai koefisien b sebesar 9450 yang berarti apabila lama pengalaman kerja naik sebesar 1 satuan maka jumlah gaji akan naik sebesar 9450
- Bentuk persamaan regresinya adalah $Y = 25792 + 9450X$