

BTS BUKU TENTANG STATISTIKA

REGRESI PARAMETRIK DAN NON PARAMETRIK
TEORI DAN APLIKASI DENGAN SOFTWARE R

Andrea Tri Rian Dani, S.Stat., M.Stat.
Fachrian Bimantoro Putra, S.Stat.
Ludia Ni'matuzzahroh, S.Stat., M.Stat.
Prof. Dr. Vita Ratnasari, M.Si.
Prof. Dr. Drs. I Nyoman Budiantara, M.Si.

**BTS (Buku Tentang Statistika):
Regresi Parametrik dan Nonparametrik, Teori
dan Aplikasi dengan Software R**

Andrea Tri Rian Dani, Fachrian Bimantoro Putra,
Ludia Ni'matuzzahroh, Vita Ratnasari,
I Nyoman Budiantara



BTS (Buku Tentang Statistika): Regresi Parametrik dan Nonparametrik, Teori dan Aplikasi dengan Software R

Penulis:

Andrea Tri Rian Dani, Fachrian Bimantoro Putra,
Ludia Ni'matuzzahroh, Vita Ratnasari,
I Nyoman Budiantara

Editor:

Erik Santoso

Layouter :

Tim Kreatif PRCI

Cover:

Rusli

Cetakan Pertama : Mei 2024

Hak Cipta 2024, pada Penulis. Diterbitkan pertama kali oleh:

Perkumpulan Rumah Cemerlang Indonesia

ANGGOTA IKAPI JAWA BARAT

Pondok Karisma Residence Jalan Raflesia VI D.151
Panglayungan, Cipedes Tasikmalaya – 085223186009

Website : www.rcipress.rcipublisher.org

E-mail : rumahcemerlangindonesia@gmail.com

Copyright © 2024 by Perkumpulan Rumah Cemerlang Indonesia

All Right Reserved

- Cet. I – : Perkumpulan Rumah Cemerlang Indonesia, 2024

; 14,8 x 21 cm

ISBN : 978-623-448-863-0

Hak cipta dilindungi undang-undang

Dilarang memperbanyak buku ini dalam bentuk dan dengan
cara apapun tanpa izin tertulis dari penulis dan penerbit

Undang-undang No.19 Tahun 2002 Tentang

Hak Cipta Pasal 72

Undang-undang No.19 Tahun 2002 Tentang Hak Cipta
Pasal 72

Barang siapa dengan sengaja melanggar dan tanpa hak melakukan perbuatan sebagaimana dimaksud dalam pasal ayat (1) atau pasal 49 ayat (1) dan ayat (2) dipidana dengan pidana penjara masing-masing paling sedikit 1 (satu) bulan dan/atau denda paling sedikit Rp.1.000.000,00 (satu juta rupiah), atau pidana penjara paling lama 7 (tujuh) tahun dan/atau denda paling banyak Rp.5.000.000.000,00 (lima miliar rupiah).

Barang siapa dengan sengaja menyiarkan, memamerkan, mengedarkan, atau menjual kepada umum suatu ciptaan atau barang hasil pelanggaran hak cipta terkait sebagai dimaksud pada ayat (1) dipidana dengan pidana penjara paling lama 5 (lima) tahun dan/atau denda paling banyak Rp.500.000.000,00 (lima ratus juta rupiah)

KATA PENGANTAR

Buku Tentang Statistika (BTS) dengan materi pokok Regresi Parametrik dan Nonparametrik dipersiapkan oleh penulis agar dapat menjadi bahan bacaan pengantar kepada mahasiswa dan siapa saja yang ingin mempelajari statistika, khususnya pada bidang Analisis Regresi. BTS dengan materi pokok Regresi Parametrik dan Nonparametrik diharapkan dapat menjadi referensi dan membuat para pembaca untuk bisa mempelajari, memahami materi analisis regresi dengan mudah dan langsung praktek dengan kasus-kasus yang diberikan menggunakan software R. Pada buku ini, diperkenalkan estimator-estimator dalam regresi nonparametrik yang bisa menjadi alternatif pendekatan pemodelan menggunakan analisis regresi.

Dengan rasa syukur yang mendalam, karya ini di persembahkan untuk seseorang yang selalu menemani penulis untuk melewati berbagai rintangan dan mencapai banyak pencapaian sampai dengan saat ini. Pertemuan kita bukanlah suatu kebetulan, namun selalu punya makna tersirat. Suka Duka Cita Tangis Tawa tersimpan dalam memori yang tersimpan diantara Mentari dan Rembulan. Semoga harapan Bahagia bersama, sukses bersama terpancar melalui janji-janji yang sudah diikrarkan.

Akhir kata, penulis mengucapkan terima kasih sebesar-besarnya kepada semua pihak yang telah membantu dan memberikan dukungan. Penulis menyadari sepenuhnya, bahwa buku ini masih jauh dari kata sempurna, sedemikian sehingga saran, kritik, dan masukan yang membangun dari para pembaca sangat diharapkan agar pada edisi selanjutnya bisa diperbaiki dan menjadi lebih baik. Saran dapat disampaikan melalui email: andreatririandani@fmipa.unmul.ac.id. Semoga buku ini dapat bermanfaat untuk kita semua, Aamiin.

Penulis

DAFTAR ISI

KATA PENGANTAR	I
DAFTAR ISI	II
BAB 1 KORELASI	1
A. Konsep Korelasi	1
B. Korelasi Pearson <i>Product Moment</i>	4
C. Korelasi Rank-Spearman	7
D. Korelasi Kendall-Tau	9
E. Korelasi Phi	11
F. Korelasi Koefisien Kontingensi	13
BAB 2 ANALISIS REGRESI	16
A. Pengertian Analisis Regresi	16
B. Regresi Parametrik dan Nonparametrik	18
C. Regresi Semiparametrik	20
BAB 3 REGRESI LINEAR	21
A. Regresi Linear Sederhana	21
B. Regresi Linear Berganda	25
C. Asumsi pada Regresi Linear	32
D. Contoh Kasus Regresi Linear Sederhana	38
E. Contoh Kasus Regresi Linear Berganda	45
BAB 4 REGRESI POLINOMIAL	55
A. Regresi Polinomial Univariabel	55
B. Regresi Polinomial Multivariabel	60
C. Contoh Kasus Regresi Polinomial Univariabel	61
BAB 5 REGRESI NONPARAMETRIK <i>SPLINE TRUNCATED</i>	69
A. Konsep Dasar Regresi <i>Spline Truncated</i>	69
B. Pemilihan Titik Knot Optimal	70
C. Regresi Nonparametrik <i>Spline Truncated</i> Univariabel	73
D. Regresi Nonparametrik <i>Spline Truncated</i> Multivariabel	79
E. Contoh Kasus Regresi Nonparametrik <i>Spline Truncated</i> Univariabel	84
F. Contoh Kasus Regresi Nonparametrik <i>Spline Truncated</i> Multivariabel	95

BAB 6 REGRESI NONPARAMETRIK DERET FOURIER	115
A. Konsep Dasar Regresi Deret Fourier	116
B. Pemilihan Osilasi yang Optimal	117
C. Regresi Nonparametrik Deret Fourier Univariabel	119
D. Contoh Kasus Regresi Nonparametrik Deret Fourier Multivariabel	123
 BAB 7 REGRESI NONPARAMETRIK KERNEL	 132
A. Pendahuluan	132
B. Konsep Dasar Regresi Kernel	133
C. Pemilihan <i>Bandwidth</i> yang Optimal	136
D. Contoh Kasus Regresi Nonparametrik Kernel	138
 DAFTAR PUSTAKA	 145

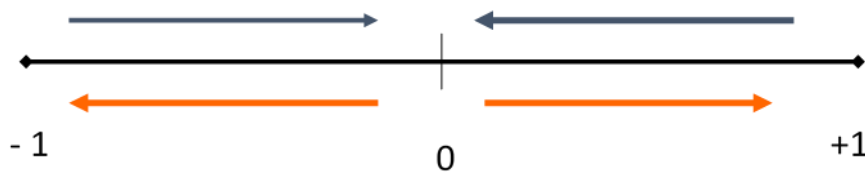
BAB 1

KORELASI

A. Konsep Korelasi

Analisis Korelasi merupakan metode yang dapat digunakan untuk mengukur ataupun menganalisis derajat keeratan hubungan antar dua variabel. Pada korelasi, tidak memperhatikan variabel mana yang mempengaruhi ataupun variabel mana yang dipengaruhi.

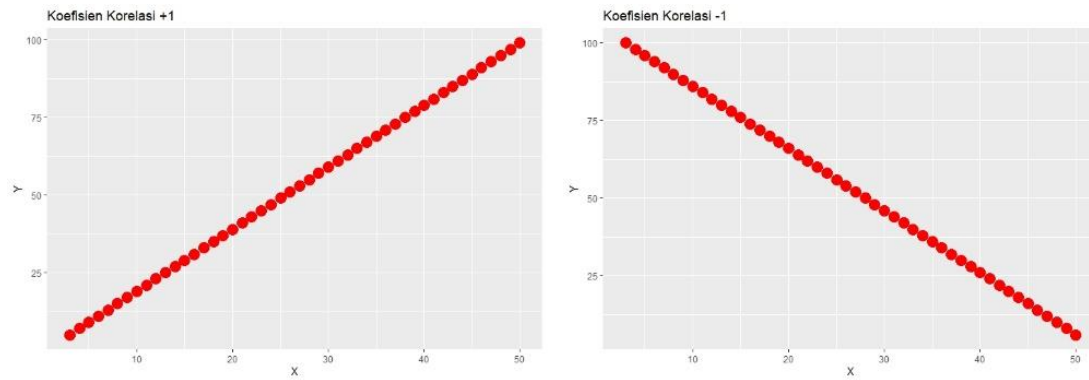
Pada analisis korelasi, besar kecilnya hubungan antar dua variabel dapat dinyatakan dalam koefisien korelasi. Nilai dari koefisien korelasi berada pada interval -1 sampai dengan 1, dimana tanda dari koefisien korelasi dapat bernilai negatif (-) ataupun positif (+). Tanda koefisien menyatakan arah hubungan antar kedua variabel.



Gambar 1.1 Ilustrasi Rentang Nilai Korelasi

Berdasarkan Gambar 1.1 dapat kita jabarkan jika semakin mendekati 0, maka derajat keeratan hubungan antara dua variabel semakin lemah, Namun semakin mendekati 1, maka derajat keeratan hubungannya akan semakin kuat bahkan sempurna

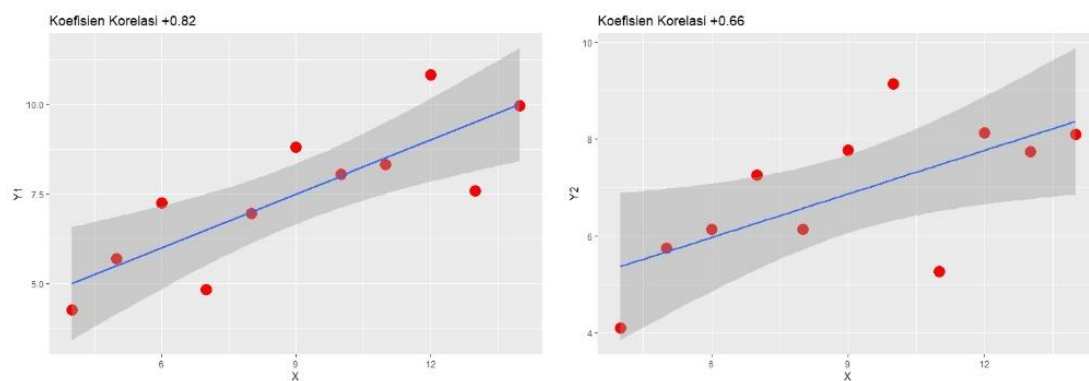
Untuk dapat mengetahui derajat hubungan antara kedua variabel (X dan Y) dengan menggunakan koefisien korelasi, dapat divisualisasikan seperti pada diagram pencar:



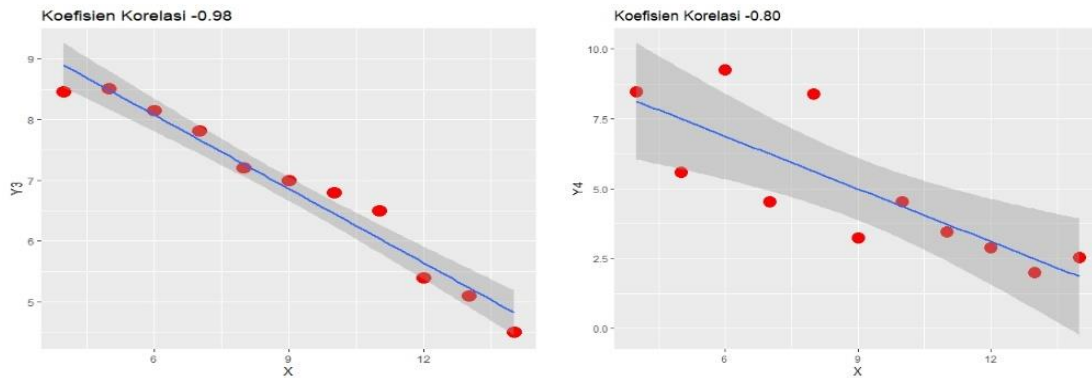
Gambar 1.2 Korelasi Sempurna (+1) dan (-1)

Apabila kedua variabel (X dan Y) mempunyai koefisien korelasi sebesar +1 atau -1, maka berarti hubungan antara kedua variabel tersebut sangat sempurna. Tanda (+) menyatakan arah hubungan, dimana apabila salah satu variabel (sebagai contoh X) mengalami peningkatan, maka variabel Y juga akan mengalami peningkatan. Dan sebaliknya, apabila tandanya (-), maka apabila variabel X mengalami peningkatan, maka variabel Y akan mengalami penurunan, dan atau sebaliknya.

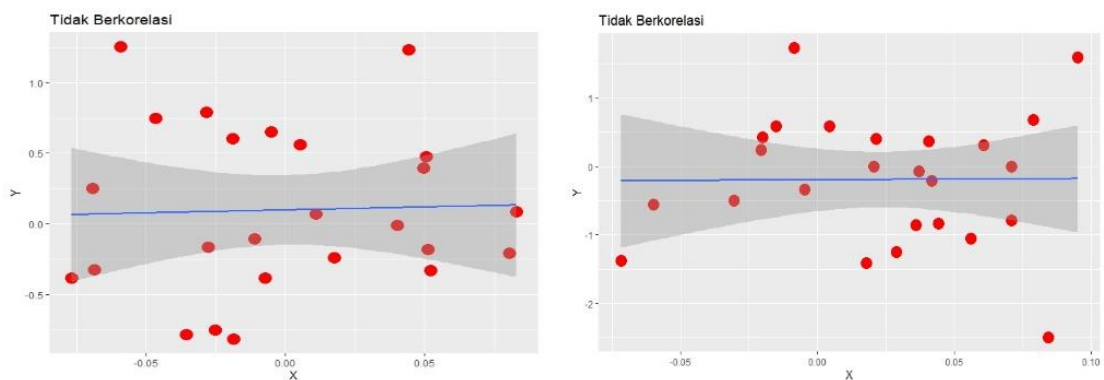
Ditampilkan beberapa contoh pola hubungan antara kedua variabel sebagai berikut:



Gambar 1.3 Korelasi Positif



Gambar 1.4 Korelasi Negatif



Gambar 1.5 Tidak ada Korelasi

Seperti yang kita ketahui, bahwa nilai dari koefisien korelasi berada pada interval 0 sampai dengan 1, sehingga untuk memudahkan dalam tahap interpretasi, dapat dibuat kategorisasi derajat (keamatan) hubungan sebagai berikut:

Tabel 1.1 Kategorisasi Nilai Koefisien Korelasi

0	: Tidak ada korelasi
0,00-0,25	: Korelasi sangat lemah
0,26-0,50	: Korelasi cukup
0,51-0,75	: Korelasi kuat
0,76-0,99	: Korelasi sangat kuat
1	: Korelasi sempurna

Banyak sekali metode-metode dalam analisis korelasi yang dapat digunakan, sesuai dengan jenis dan skala data penelitian. Pada buku ini, akan dibahas beberapa metode yang sering digunakan dalam penelitian, diantaranya: *Korelasi Pearson Product Moment*; *Korelasi Rank-Spearman*; *Korelasi Kendall-Tau*; *Korelasi Phi*; *Korelasi Koefisien Kontingensi*.

B. Korelasi Pearson *Product Moment*

Korelasi *Pearson Product Moment* merupakan korelasi yang paling sering digunakan oleh para peneliti. Jenis korelasi ini digunakan untuk mengukur derajat keeratan hubungan yang “linear” antara dua variabel. Dan juga, korelasi ini digunakan apabila kedua variabel, sebagai contoh X dan Y masing-masing mempunyai skala pengukuran data yaitu “interval atau rasio”. Adapun formula dalam menghitung koefisien korelasi menggunakan *Pearson Product Moment* adalah:

$$r_{x,y} = \frac{n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{\sqrt{\left(n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right) \left(n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 \right)}} \quad 1.1$$

Dimana:

$r_{x,y}$: Koefisien korelasi *Pearson Product Moment*

n : Banyaknya pasangan pengamatan (X dan Y)

=== CONTOH KASUS KORELASI PEARSON ===

Diberikan data berpasangan (x_i, y_i) , dimana $i = 1, 2, \dots, 25$. Variabel x merupakan volume pemesanan dan y merupakan waktu pelayanan, dengan data ditampilkan pada Tabel 1.2. Hitunglah derajat keeratan antara kedua variabel tersebut menggunakan korelasi *Pearson Product Moment*.

Tabel 1.2 Data Korelasi Pearson

Y	X	Y	X
16,68	7,00	19,75	6,00
11,50	3,00	24,00	9,00
12,03	3,00	29,00	10,00
14,88	4,00	15,35	6,00
13,75	6,00	19,00	7,00
18,11	7,00	9,50	3,00
8,00	2,00	35,10	17,00
17,83	7,00	17,90	10,00
79,24	30,00	52,32	26,00
21,50	5,00	18,75	9,00
40,33	16,00	19,83	8,00
21,00	10,00	10,75	4,00
13,50	4,00		

Sumber: Montgomery, Peck, and Vining (Data Example 2.9)

Aplikasi dengan R Software:

Langkah awal yang perlu dipersiapkan adalah mendownload *packages* (**stats**) dan (**ggplot2**). Kemudian untuk menghitung nilai korelasi Pearson *Product Moment* di R dapat dilakukan dengan *syntax* sebagai berikut:

```
library(stats)
data<-read.csv(file.choose(),header=TRUE)
data
X<-data$X
Y<-data$Y

cor.test(X,Y,method="pearson") #Calculate Pearson Correlation
```

```
df<-data.frame(X,Y)
library(ggplot2)
ggplot(df,aes(X,Y)) +
geom_point(colour="red",lwd=5)+geom_smooth(method='lm',formula
=y~x)+labs(title = "Korelasi Pearson")
```

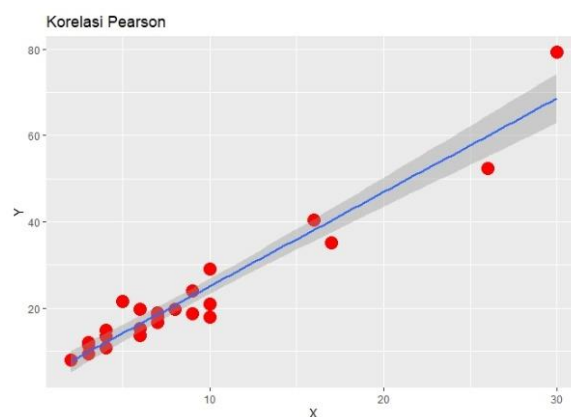
Sehingga didapatkan *output*:

```
> cor.test(X,Y)

Pearson's product-moment correlation

data:  X and Y
t = 17.546, df = 23, p-value = 8.22e-15
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9202275 0.9845031
sample estimates:
cor
0.9646146
```

Berdasarkan *output* korelasi Pearson *Product Moment* diketahui bahwa hubungan antara variabel *x* (volume pemesanan) dengan variabel *y* (waktu pelayanan) adalah sebesar 0,96 yang menunjukkan bahwa korelasi berada dalam kategori “*sangat kuat*”. Sementara tanda hubungan bernilai positif mengindikasikan bahwa pola hubungan antara volume pemesanan dengan waktu pelayanan adalah searah (berbanding lurus). Ini berarti bahwa semakin tinggi volume pemesanan maka semakin tinggi pula waktu pelayanan. Ditampilkan diagram pencar pola hubungan antara variabel *x* (volume pemesanan) dengan variabel *y* (waktu pelayanan) pada Gambar 1.6.



Gambar 1.6 Diagram Pencar

C. Korelasi Rank-Spearman

Korelasi Rank-Spearman digunakan apabila kedua variabel, sebagai contoh X dan Y masing-masing mempunyai skala pengukuran data yaitu ordinal. Adapun formula dalam menghitung koefisien korelasi menggunakan Rank-Spearman adalah:

$$\rho_{x,y} = 1 - \frac{6 \sum_{i=1}^n [R(x_i) - R(y_i)]^2}{n(n^2 - 1)} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad 1.2$$

Dimana:

$\rho_{x,y}$: Koefisien korelasi Rank-Spearman

$R(x_i)$: Rank skor data variabel X

$R(y_i)$: Rank skor data variabel Y

d_i : Selisih setiap pasangan rank

n : Banyaknya pasangan pengamatan (X dan Y)

==== CONTOH KASUS KORELASI RANK-SPEARMAN ====

Untuk mengetahui tingkat kepuasan dan loyalitas dari pelanggan di salah satu tempat perbelanjaan kebutuhan sehari-hari, akhirnya pihak manager melakukan suatu penelitian apakah terdapat hubungan antara kepuasan dan loyalitas. Instrumen yang digunakan adalah kuesioner. Data ditampilkan sebagai berikut:

Tabel 1.3 Data Korelasi Rank-Spearman

Kepuasan	Loyalitas	Kepuasan	Loyalitas
5	5	4	4
4	4	5	5
5	5	4	4
5	4	5	5
4	4	2	3
5	5	5	5
5	5	4	5

Kepuasan	Loyalitas	Kepuasan	Loyalitas
5	5	5	5
5	4	5	5
5	5	5	4
4	4	4	4
5	5	5	4
5	5	5	5
4	5	5	4
5	5	4	4

Aplikasi dengan R Software:

Langkah awal yang perlu dipersiapkan adalah mendownload *packages* (**stats**). Kemudian untuk menghitung nilai korelasi Rank-Spearman di R dapat dilakukan dengan *syntax* sebagai berikut:

```
library(stats)
data<-read.csv(file,choose(),header=TRUE)

X<-data$kepuasan
Y<-data$loyalitas

cor.test(X,Y,method="spearman") #Spearman Correlation
```

Sehingga didapatkan *output*:

```
> cor.test(X,Y,method="spearman")

Spearman's rank correlation rho

data:  X and Y
S = 1960.3, p-value = 0.001174
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.5638875
```

Berdasarkan *output* korelasi Rank-Spearman diketahui bahwa hubungan antara kepuasan dengan loyalitas adalah sebesar 0,56 yang menunjukkan bahwa korelasi berada dalam kategori “kuat”. Sementara tanda hubungan bernilai positif

mengindikasikan bahwa pola hubungan antara kepuasan dengan loyalitas adalah searah (berbanding lurus).

D. Korelasi Kendall-Tau

Korelasi Kendall-Tau hampir sama dengan korelasi Rank-Spearman, dimana dapat digunakan apabila kedua variabel, yaitu X dan Y masing-masing mempunyai skala pengukuran data yaitu ordinal. Formula dalam menghitung koefisien korelasi menggunakan Kendall-Tau adalah:

$$\tau_{x,y} = \frac{2S}{n(n-1)} = \frac{2(C-D)}{n(n-1)} \quad 1.3$$

Jika terdapat rank yang sama, maka formula nya dilengkapi dengan faktor koreksi rank yang sama, yaitu:

$$\tau_{x,y} = \frac{2(C-D)}{\sqrt{n(n-1)-T_x} \sqrt{n(n-1)-T_y}} \quad 1.4$$

$$T_x = \sum_{i=1}^s (t_{i(x)}^2 - t_{i(x)}) \text{ dan } T_y = \sum_{i=1}^s (t_{i(y)}^2 - t_{i(y)})$$

Dimana:

- $\tau_{x,y}$: Koefisien korelasi Kendall-Tau
- C : Banyaknya pasangan Konkordansi
- D : Banyaknya pasangan Diskordansi
- T_x : Koreksi Rank untuk X yang sama
- T_y : Koreksi Rank untuk Y yang sama
- s : Jumlah Konkordansi dan Diskordansi
- n : Banyaknya pasangan pengamatan (X dan Y)

==== CONTOH KASUS KORELASI KENDALL-TAU ====

Seorang peneliti ingin melihat dan mengetahui apakah terdapat korelasi antara variabel kepuasan dengan pelayanan di salah satu rumah sakit di kota X. Untuk itu peneliti tersebut membagikan kuesioner secara acak kepada 15 responden yang

terpilih untuk dijadikan observasi penelitian. Adapun tabulasi data jawaban dari pada responden ditampilkan pada Tabel:

Tabel 1.4 Data Korelasi Kendall-Tau

Responden	Kepuasan (X)	Skor X	Pelayanan Y)	Skor Y
1	Tidak Puas	2	Cukup Baik	3
2	Puas	4	Sangat Baik	5
3	Puas	4	Sangat Baik	5
4	Sangat Puas	5	Sangat Baik	5
5	Puas	4	Baik	4
6	Sangat Puas	5	Baik	4
7	Cukup Puas	3	Baik	4
8	Puas	4	Baik	4
9	Sangat Puas	5	Sangat Baik	5
10	Puas	4	Baik	4
11	Sangat Puas	5	Sangat Baik	5
12	Tidak Puas	2	Cukup Baik	3
13	Cukup Puas	3	Baik	4
14	Sangat Puas	5	Sangat Baik	5
15	Tidak Puas	2	Baik	4

Berdasarkan Tabel 1.4, skala pengukuran pada kuesioner yang digunakan adalah Likert sebagai berikut:

Kepuasan (X) ->

- (1) : Sangat Tidak Puas
- (2) : Tidak Puas
- (3) : Cukup Puas
- (4) : Puas
- (5) : Sangat Puas

Pelayanan (Y) ->

- (1) : Sangat Tidak Baik
- (2) : Tidak Baik
- (3) : Cukup Baik
- (4) : Baik
- (5) : Sangat Baik

Aplikasi dengan R Software:

Langkah awal yang perlu dipersiapkan adalah mendownload *packages* (**stats**). Kemudian untuk menghitung nilai korelasi Kendall-Tau di R dapat dilakukan dengan *syntax* sebagai berikut:

```
library(stats)
data<-read.csv(file,choose(),header=TRUE)

X<-data$kepuasan
Y<-data$pelayanan

cor.test(X,Y,method="kendall") #Kendal Tau Correlation
```

Sehingga didapatkan *output*:

```
> cor.test(X,Y,method="kendall")

      Kendall's rank correlation tau

data:  X and Y
z = 2.9687, p-value = 0.002991
alternative hypothesis: true tau is not equal to 0
sample estimates:
      tau 
0.7006585
```

Berdasarkan *output* korelasi Kendall-Tau diketahui bahwa hubungan antara kepuasan dengan loyalitas adalah sebesar 0,70 yang menunjukkan bahwa korelasi berada dalam kategori “kuat”. Sementara tanda hubungan bernilai positif mengindikasikan bahwa pola hubungan antara kepuasan dengan pelayanan adalah searah (berbanding lurus).

E. Korelasi Phi

Korelasi Phi merupakan salah satu uji asosiatif yang digunakan untuk mengukur derajat keeratan hubungan antara dua variabel dengan skala data nominal, yang kemudian dapat di sajikan ke dalam tabel kontingensi yang berukuran 2 x 2.

Tabel 1.5 Tabel Kontingensi 2x2

	Kolom		Jumlah
Baris	A	B	A+B
	C	D	C+D
Jumlah	A+C	B+D	N

Adapun formula dalam menghitung korelasi Phi adalah sebagai berikut:

$$r_{x,y} = \frac{AD - BC}{\sqrt{(A+B)(C+D)(A+C)(B+D)}} \quad 1.5$$

Dimana:

A, B, C, dan D merupakan frekuensi yang ada pada tabel kontingensi.

==== CONTOH KASUS KORELASI PHI ====

Berikut disajikan data pengendara mobil yang mengalami kecelakaan di jalan raya. Terdapat 2 variabel, yaitu Cidera dan Penggunaan Sabuk Pengaman. Variabel cidera dikategorikan sebagai Cidera Fatal dan Tidak Fatal, sedangkan penggunaan sabuk pengaman dikategorikan menggunakan sabuk pengaman dan tidak menggunakan sabuk pengaman. Data ditampilkan dalam bentuk tabel kontingensi pada Tabel 1.6.

Tabel 1.6 Data Korelasi Phi

Penggunaan Sabuk Pengaman	Cidera	
	Fatal	Tidak Fatal
Tidak Menggunakan	1601	163
Menggunakan Sabuk Pengaman	510	412

Sumber: Alan Agresti (Data for Problem 2.3)

Aplikasi dengan R Software:

Langkah awal yang perlu dipersiapkan adalah mendownload *packages* (**psych**), Kemudian untuk menghitung nilai korelasi Phi di R dapat dilakukan dengan *syntax* sebagai berikut:

```
library(psych)
data=matrix(c(1601,510,163,412), nrow=2)

phi(data) #Calculate Phi Correlation
```

Sehingga didapatkan *output*:

```
> data=matrix(c(1601,510,163,412),nrow=2)
> data
      [,1] [,2]
[1,] 1601  163
[2,]  510  412
> phi(data)
[1] 0.41
```

Berdasarkan *output* korelasi Phi, diketahui bahwa hubungan antara penggunaan sabuk pengaman dan cedera adalah sebesar 0,41 yang menunjukkan bahwa korelasi berada dalam kategori “cukup”. Sementara tanda hubungan bernilai positif mengindikasikan bahwa pola hubungan antara kedua variabel adalah searah (berbanding lurus).

F. Korelasi Koefisien Kontingensi

Korelasi Koefisien Kontingensi merupakan salah satu uji asosiatif yang serupa dengan korelasi Phi, namun digunakan untuk mengukur derajat keeratan hubungan antara dua variabel dengan skala data nominal politomi. Sebagai contoh di tampilkan tabel kontingensi berukuran $b \times k$.

Tabel 1.7 Tabel Kontingensi 2x3

	Kolom			Jumlah
Baris	A	B	C	A+B+C
	D	E	F	D+E+F
Jumlah	A+D	B+E	C+F	N

Adapun formula dalam menghitung korelasi Koefisien Kontingensi adalah sebagai berikut:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} \quad 1.6$$

$$\chi^2 = \sum_{j=1}^k \sum_{i=1}^b \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad 1.7$$

Dimana:

O_{ij} : Frekuensi observasi

E_{ij} : Frekuensi harapan

χ^2 : Nilai *Chi-Squared*

n : Frekuensi pengamatan

=== CONTOH KASUS KORELASI KOEFISIEN KONTINGENSI ===

Berikut disajikan data kecenderungan pemilihan jenis partai berdasarkan ras warna kulit disalah satu Negara. Data ditampilkan dalam bentuk tabel kontingensi pada Tabel 1.8.

Tabel 1.8 Data Korelasi Koefisien Kontingensi

Ras Warna Kulit	Jenis Partai		
	Demokrat	Independent	Republik
Hitam	103	15	11
Putih	341	105	405

Sumber: Alan Agresti (Data for Problem 3.4)

Aplikasi dengan R Software:

Langkah awal yang perlu dipersiapkan adalah mendownload *packages* (**stats**). Kemudian untuk menghitung nilai korelasi koefisien kontingensi di R, kita perlu menghitung terlebih dahulu nilai χ^2 dengan *syntax* sebagai berikut:

```
library(stats)
data=matrix(c(103,341,15,105,11,405),nrow=2)
data

Chisq=chisq.test(data)
```

Sehingga didapatkan *output*:

```
> data=matrix(c(103,341,15,105,11,405),nrow=2)
> data
      [,1] [,2] [,3]
[1,]  103   15   11
[2,]  341  105  405
>
> Chisq=chisq.test(data)
> Chisq

      Pearson's Chi-squared test

data:  data
X-squared = 79.431, df = 2, p-value < 2.2e-16
```

Langkah selanjutnya menghitung korelasi koefisien kontingensi menggunakan Persamaan (1.4) sebagai berikut:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$
$$C = \sqrt{\frac{79,431}{79,431 + 980}} = 0,27$$

Berdasarkan hasil perhitungan korelasi koefisien kontingensi diketahui bahwa hubungan antara pemilihan jenis partai dan ras warna kulit penduduk adalah sebesar 0,27 yang menunjukkan bahwa korelasi berada dalam kategori “*cukup*”. Sementara tanda hubungan bernilai positif mengindikasikan bahwa pola hubungan antara kedua variabel adalah searah (berbanding lurus).

BAB 2

ANALISIS REGRESI

A. Pengertian Analisis Regresi

Analisis regresi adalah metode statistik yang digunakan untuk memodelkan hubungan antara variabel respon dengan satu atau lebih variabel prediktor. Variabel respon adalah variabel yang dipengaruhi, sedangkan variabel prediktor adalah variabel yang diduga mempengaruhi.

Beberapa pengertian analisis regresi menurut para ahli:

- Draper dan Smith (1981) menjelaskan bahwa analisis regresi adalah metode statistik yang digunakan untuk menjelaskan hubungan antara variabel respon dan satu atau lebih variabel prediktor
- Montgomery dan Runger (2003) menjelaskan bahwa regresi adalah metode untuk menganalisis hubungan antara variabel prediktor terhadap variabel respon
- Budiantara (2010) mendefinisikan regresi sebagai metode statistik yang digunakan untuk mempelajari hubungan fungsional antara variabel respon dan variabel prediktor

Budiantara (2010) menjelaskan bahwa secara umum terdapat tiga tujuan utama pada pemodelan regresi, yaitu:

1. Mengetahui pengaruh variabel prediktor (X) terhadap variabel respon (Y)
2. Memprediksi nilai variabel respon (Y) berdasarkan nilai pada variabel prediktor (X)
3. Mendapatkan model terbaik yang mampu menjelaskan hubungan fungsional antara variabel prediktor dengan variabel respon,

Secara umum, ketiga tujuan pemodelan regresi dapat divisualisasikan pada Gambar:

Contoh: Regresi kernel, regresi spline, regresi deret Fourier, dll.

Kelebihan:

Lebih fleksibel dalam menangkap hubungan yang kompleks.

Tidak sensitif terhadap pelanggaran asumsi.

Kekurangan:

Lebih sulit dipahami dan diinterpretasikan.

Memiliki lebih sedikit teori dan metode yang tersedia.

Kurang efisien dalam hal estimasi parameter.

3. Regresi Semiparametrik

Asumsi: Bentuk hubungan antara variabel dependen dan variabel independent diasumsikan sebagian diketahui.

Contoh: Regresi semiparametrik spline truncated, Regresi semiparametrik deret Fourier

Kelebihan:

Menggabungkan keunggulan regresi parametrik dan nonparametrik.

Lebih fleksibel daripada regresi parametrik

Kekurangan:

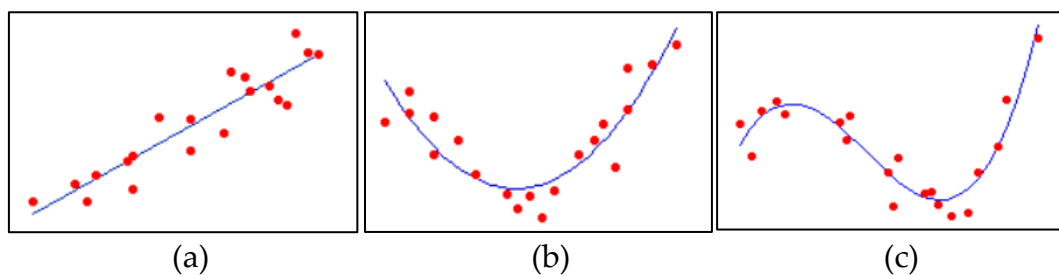
Lebih kompleks daripada regresi parametrik dan nonparametrik.

B. Regresi Parametrik dan Nonparametrik

Salah satu tahapan awal dalam analisis regresi adalah melakukan pendugaan bentuk kurva regresi. Pendugaan bentuk kurva regresi digunakan untuk menjelaskan hubungan antara variabel respon dengan variabel prediktor. Model regresi umumnya dibagi menjadi 3 bagian, yaitu model regresi parametrik, model regresi nonparametrik dan model regresi semiparametrik. Bentuk dari model regresi sangat tergantung pada kurva $f(x_i)$. Apabila bentuk kurva $f(x_i)$ diketahui, maka pendekatan model regresi yang digunakan adalah pendekatan model regresi parametrik. Namun apabila bentuk kurva $f(x_i)$ tidak diketahui atau

tidak terdapat informasi masa lalu yang lengkap tentang bentuk pola datanya, maka pendekatan model regresi yang digunakan adalah pendekatan model regresi nonparametrik. Dalam beberapa kasus, sebagian bentuk pola data diketahui, sedangkan sebagian lainnya tidak diketahui bentuk pola datanya. Pada kasus ini, pendekatan model regresi yang disarankan adalah pendekatan model regresi semiparametrik.

Pendekatan pendugaan yang sering dijumpai adalah pendekatan parametrik. Asumsi yang mendasari pendekatan regresi parametrik adalah bahwa bentuk dari kurva regresi dapat digambarkan dalam pola tertentu. Jika kurva regresi mengikuti bentuk yang telah diketahui sebelumnya seperti linier, kuadratik, kubik, eksponensial dan lain sebagainya, maka dapat digunakan pendekatan regresi parametrik. Beberapa grafik ditampilkan sebagai berikut:

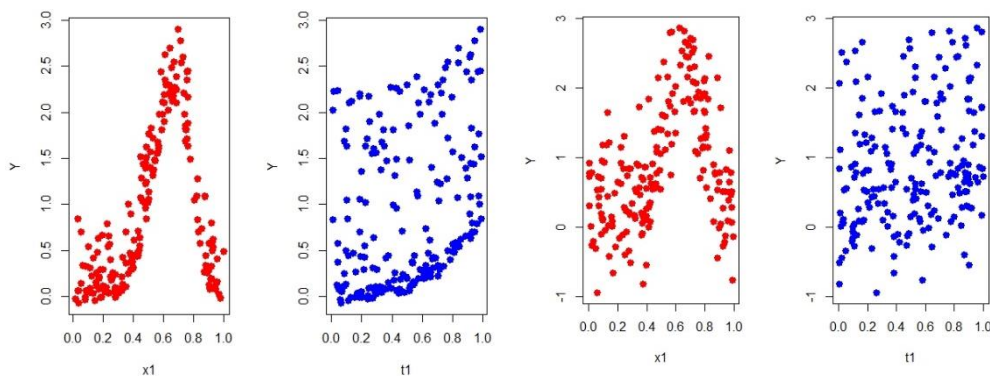


Gambar 2.1 Grafik dari beberapa model regresi parametrik (a) linier;
(b) kuadratik; (c) kubik

Pendekatan model regresi parametrik memiliki sifat yang sangat baik dari sisi statistika inferensi yaitu: sederhana, memiliki estimator yang tidak bias, efisien dan konsisten serta mudah untuk diinterpretasikan, yang sangat jarang dimiliki oleh pendekatan regresi lainnya seperti regresi nonparametrik dan regresi semiparametrik. Estimasi parameter dari model diperoleh berdasarkan berbagai metode penaksir yang telah diketahui dalam ilmu statistika, seperti *least square*, *maximum likelihood estimation*, dan *Bayesian*.

Hardle (1994), menyatakan bahwa pendekatan regresi nonparametrik mulai dikenal dikisaran abad-ke 19 tepatnya pada Tahun 1857. Regresi nonparametrik

merupakan salah satu pendekatan yang digunakan untuk mengetahui pola hubungan antara variabel prediktor dan variabel respon yang tidak diketahui bentuk kurva regresinya dan tidak terdapat informasi masa lalu yang lengkap tentang bentuk pola datanya. Berdasarkan hal itu, maka secara visual pola data yang diberikan oleh variabel prediktor dan variabel respon tidak mempunyai pola yang mengikuti pola tertentu. Dalam pandangan regresi nonparametrik, data akan mencari sendiri bentuk estimasi dari kurva regresinya tanpa harus dipengaruhi oleh faktor subyektifitas peneliti. Ini berarti pendekatan model regresi nonparametrik memiliki fleksibilitas yang tinggi (sangat fleksibel) dan obyektif.



Gambar 2.2 Grafik dari pola hubungan prediktor dan respon yang tidak diketahui polanya

C. Regresi Semiparametrik

Pendekatan regresi semiparametrik merupakan kombinasi antara pendekatan regresi parametrik dan nonparametrik. Pada prakteknya, sering kali ditemukan kasus dimana sebagian bentuk pola kurva regresinya diketahui dan sebagian lagi tidak diketahui, sehingga pendekatan yang diusulkan adalah pendekatan regresi semiparametrik. Variabel prediktor yang memiliki pola data yang diketahui atau terdapat informasi sebelumnya mengenai pola datanya akan dimodelkan dengan pendekatan regresi parametrik, sedangkan variabel yang tidak diketahui pola datanya akan dimodelkan dengan pendekatan regresi nonparametrik.

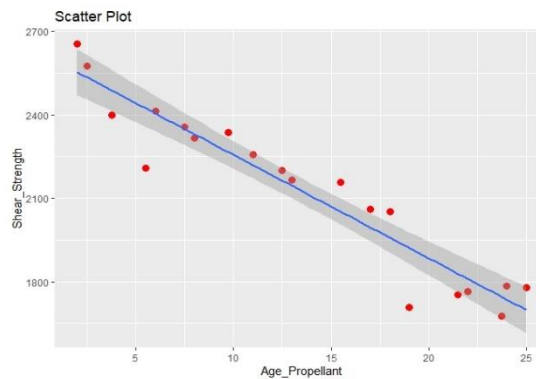
BAB 3

REGRESI LINEAR

Regresi linear (*linear regression*) merupakan salah satu metode yang digunakan untuk mengetahui pola hubungan serta pengaruh antara variabel prediktor terhadap variabel respon. Seperti dengan sebutannya, yaitu “regresi linear”, maka tentunya pola hubungan yang dapat dimodelkan adalah pola hubungan yang linear. Berdasarkan definisi tersebut, maka regresi linear masuk kedalam pendekatan regresi parametrik.

A. Regresi Linear Sederhana

Regresi linear sederhana adalah salah satu bentuk dari regresi linear, yang dapat digunakan untuk menggambarkan hubungan serta mengetahui pengaruh antara satu variabel prediktor (X) terhadap satu variabel respon (Y).



Gambar 3.1 Diagram pencar

Hubungan dari variabel prediktor dan variabel respon dapat digambarkan sebagai suatu garis lurus (linear). Bentuk persamaan dari regresi linear sederhana dapat dituliskan:

$$y_i = f(x_i) + \varepsilon_i \quad 3.1$$

Dimana $f(x_i)$ adalah persamaan regresi atau biasa disebut fungsi regresi. Pada regresi linear sederhana, maka $f(x_i)$ dapat dituliskan sebagai berikut:

$$f(x_i) = \beta_0 + \beta_1 x_{1i} \quad 3.2$$

Sehingga berdasarkan Persamaan (3.1) dan (3.2), dapat dituliskan model regresi linear sederhana:

$$y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i \quad 3.3$$

dengan:

y_i : Variabel respon

x_{1i} : Variabel prediktor

β_0 : Intersep / Konstanta

β_1 : Koefisien regresi

ε_i : *Error* yang diasumsikan berdistribusi Normal Independen dan Identik,

dengan mean 0 dan variansi σ^2 , $\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$

Selanjutnya β_0 dan β_1 merupakan parameter-parameter dari populasi yang bernilai real dan tidak diketahui, yang kemudian akan diestimasi nilainya. Untuk mengestimasi parameter β_0 dan β_1 , terdapat beberapa metode yang dapat digunakan, salah dua metode estimasi yang cukup terkenal diantaranya *Ordinary Least Squares* (OLS) dan *Maximum Likelihood Estimation* (MLE).

1. Estimasi Koefisien Regresi

Pada model regresi linear sederhana, parameter β_0 dan β_1 dapat diestimasi dengan cara sebagai berikut:

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \quad 3.4$$

dan

$$\hat{\beta}_1 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \quad 3.5$$

Berdasarkan Persamaan (3.4) dan (3.5), maka dapat dituliskan model regresi linear sederhana setelah diestimasi:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} \quad 3.6$$

2. Pengujian Parameter Model Regresi Linear Sederhana

Langkah selanjutnya adalah melakukan pengujian parameter model regresi linear sederhana. Tujuan dari pengujian ini adalah menguji apakah variabel prediktor (X) memberikan pengaruh secara signifikan terhadap variabel respon (Y). Langkah pengujian hipotesisnya sebagai berikut:

- a. Rumusan hipotesis

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

- b. Menentukan tangka signifikansi sebesar α

- c. Daerah kritis

$$\text{Tolak } H_0 \text{ jika } F_{hit} > F_{\alpha(k, n-k-1)}$$

- d. Statistik uji

Statistik uji yang digunakan ditampilkan pada tabel *Analysis of Variance* (ANOVA) dalam Tabel 3.1.

Tabel 3.1 ANOVA

Sumber Variasi	Derajat Bebas (df)	Jumlah Kuadrat	Rata-Rata Jumlah Kuadrat
Regresi	k	$JK_{regresi} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$RJK_{regresi} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{k}$

Sumber Variasi	Derajat Bebas (df)	Jumlah Kuadrat	Rata-Rata Jumlah Kuadrat
Error	$n - k - 1$	$JK_{error} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$	$RJK_{error} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - k - 1}$
Total	$n - 1$	$JK_{total} = \sum_{i=1}^n (y_i - \bar{y})^2$	

Selanjutnya statistik uji F_{hit} dapat dihitung dengan:

$$F_{hit} = \frac{RJK_{regresi}}{RJK_{error}} \quad 3.7$$

Dibawah hipotesis H_0 , statistik uji F_{hit} mengikuti distribusi F , dengan derajat bebas (df) yaitu $(k, n - k - 1)$. Untuk kasus regresi linear sederhana, dikarenakan variabel prediktor hanya 1, maka nilai k adalah 1.

e. Kesimpulan

Berdasarkan nilai statistik F_{hit} yang kemudian dibandingkan dengan tabel distribusi F , dapat ditarik kesimpulan apakah hasilnya adalah menerima atau menolak H_0 . Jika H_0 ditolak, maka dapat diartikan jika variabel prediktor berpengaruh signifikan terhadap variabel respon. Hal sebaliknya jika H_0 gagal ditolak, maka artinya adalah variabel prediktor tidak berpengaruh terhadap variabel respon.

3. Koefisien Determinasi

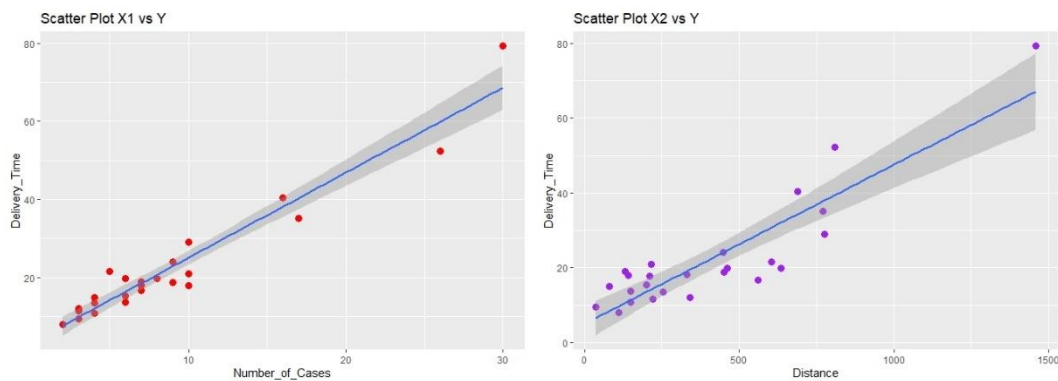
Koefisien determinasi dapat digunakan untuk melihat kebaikan dari suatu model regresi, disimbolkan dengan R^2 . Nilai R^2 menyatakan proporsi keberagaman dalam variabel respon yang dapat dijelaskan oleh hubungan linear dari variabel prediktor. Semakin tinggi nilai R^2 , maka semakin baik model yang dibentuk, karena semakin besar keberagaman variabel respon yang dapat dijelaskan atau diterangkan oleh variabel prediktor. Nilai R^2 biasanya dinyatakan dalam

persentase, untuk memudahkan dalam interpretasi. Formula untuk menghitung nilai R^2 dituliskan sebagai berikut:

$$R^2 = 1 - \frac{JK_{error}}{JK_{total}} \quad 3.8$$

B. Regresi Linear Berganda

Pada analisis regresi, sangat mungkin jika mempunyai lebih dari satu variabel prediktor. Regresi linear berganda merupakan salah satu bentuk dari regresi linear, yang dapat digunakan untuk menggambarkan hubungan serta mengetahui pengaruh antara lebih satu variabel prediktor terhadap satu variabel respon.



Gambar 3.2 Ilustrasi diagram pencar antara X1 dan X2 terhadap Y

Struktur data dari regresi linear berganda ditampilkan pada Tabel 3.2.

Tabel 3.2 Struktur Data

Observasi	Variabel Respon (y)	Variabel Prediktor			
		x_1	x_2	...	x_p
1	y_1	x_{11}	x_{21}	...	x_{p1}
2	y_2	x_{12}	x_{22}	...	x_{p2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
n	y_n	x_{1n}	x_{2n}	...	x_{pn}

Jika diberikan variabel prediktor sebanyak p dan diberikan data berpasangan $(x_1, x_2, \dots, x_p, y)$ yang selanjutnya hubungan antara variabel prediktor $(x_1, x_2, \dots, x_p)$ dengan variabel respon (y) mengikuti model regresi linear:

$$y_i = f(x_{1i}, x_{2i}, \dots, x_{pi}) + \varepsilon_i \quad 3.9$$

dimana $i = 1, 2, \dots, n$.

Kurva regresi $f(x_{1i}, x_{2i}, \dots, x_{pi})$ diasumsikan aditif, sehingga dapat dituliskan:

$$f(x_{1i}, x_{2i}, \dots, x_{pi}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} \quad 3.10$$

Sehingga berdasarkan Persamaan (3.9) dan (3.10), dapat dituliskan model regresi linear berganda:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \varepsilon_i \quad 3.11$$

dengan:

y_i : Variabel respon

$x_{1i}, x_{2i}, \dots, x_{pi}$: Variabel prediktor sebanyak p

β_0 : Intersep / Konstanta

$\beta_1, \beta_2, \dots, \beta_p$: Parameter model regresi

ε_i : Error yang diasumsikan berdistribusi Normal Independen dan

Identik, dengan mean 0 dan variansi σ^2 , $\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$

$\beta_0, \beta_1, \beta_2, \dots, \beta_p$ merupakan parameter-parameter dari populasi yang bernilai real dan tidak diketahui, yang kemudian akan diestimasi nilainya.

1. Estimasi Model Regresi Linear Berganda

Menjadi permasalahan tersendiri dalam mengestimasi parameter-parameter dalam model regresi linear berganda. Untuk melakukan estimasi parameter pada model regresi linear berganda, dapat menggunakan beberapa metode estimasi yang cukup terkenal dan *familiar*, diantaranya *Ordinary Least Squares* (OLS) ataupun *Maximum Likelihood Estimation* (MLE). Perlu diingat kembali, jika model regresi linear merupakan model-model yang linear dalam parameter.

Berdasarkan Persamaan (3.11), untuk setiap $i=1,2,...,n$ maka dapat dituliskan:

untuk $i=1$, maka

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{21} + \dots + \beta_p x_{p1} + \varepsilon_1$$

untuk $i=2$, maka

$$y_2 = \beta_0 + \beta_1 x_{12} + \beta_2 x_{22} + \dots + \beta_p x_{p2} + \varepsilon_2$$

Hingga untuk $i=n$, maka

$$y_n = \beta_0 + \beta_1 x_{1n} + \beta_2 x_{2n} + \dots + \beta_p x_{pn} + \varepsilon_n$$

Model regresi diatas, dapat kita tuliskan ke dalam bentuk matriks:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p1} \\ 1 & x_{12} & x_{22} & \cdots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{pn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad 3.12$$

atau dapat ditulis:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad 3.13$$

dimana $\mathbf{y}$ merupakan vektor dari variabel respon berukuran $(n \times 1)$, $\mathbf{X}$ merupakan matriks berukuran $n \times (p+1)$, $\boldsymbol{\beta}$ adalah vektor parameter koefisien regresi yang akan diestimasi dan berukuran $(1+p) \times 1$ serta $\boldsymbol{\varepsilon}$ adalah vektor *error* acak berukuran $(n \times 1)$.

Vektor parameter $\boldsymbol{\beta}$ dapat diestimasi dengan menggunakan metode *Ordinary Least Squares* (OLS) yaitu metode untuk mendapatkan estimator dengan cara meminimumkan jumlah kuadrat *error*. *Error* dari model regresi pada Persamaan (3.13) dapat dituliskan sebagai berikut:

$$\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \quad 3.14$$

Jumlah kuadrat *error* diberikan oleh:

$$\sum_{i=1}^n \varepsilon_i^2 = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad 3.15$$

Metode OLS mendapatkan estimator $\boldsymbol{\beta}$ melalui penyelesaian optimasi:

$$\min\{Q(\boldsymbol{\beta})\} = \min\{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\} \quad 3.16$$

Untuk menyelesaikan optimasi pada Persamaan (3.16) dapat menggunakan derivatif parsial, sebagai berikut:

$$\begin{aligned} \frac{\partial(Q(\boldsymbol{\beta}))}{\partial\boldsymbol{\beta}} &= \frac{\partial\{\mathbf{y}^T\mathbf{y} - 2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}\}}{\partial\boldsymbol{\beta}} \\ &= -2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \end{aligned} \quad 3.17$$

Kemudian Persamaan (3.17), disamadengankan dengan 0:

$$\frac{\partial(Q(\boldsymbol{\beta}))}{\partial\boldsymbol{\beta}} = 0$$

$$-2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = 0$$

Kemudian dilakukan pemindahan ruas sehingga menjadi $\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{y}$. Matriks $\mathbf{X}$ adalah matriks *full rank*, sehingga $\mathbf{X}^T\mathbf{X}$ adalah matriks yang non-singular, yang artinya matriks tersebut memiliki suatu invers. Dengan mengalikan kedua ruas dengan $(\mathbf{X}^T\mathbf{X})^{-1}$, maka akan kita dapatkan estimasi parameter dari $\hat{\boldsymbol{\beta}}$ sebagai berikut:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \quad 3.18$$

Estimasi parameter dari model regresi linear, umumnya sangat digemari dalam dalam proses pemodelan, dikarenakan sifat dari estimatornya adalah *Best Linear Unbiased Estimator* (BLUE).

(a) Estimator $\hat{\boldsymbol{\beta}}$ mempunyai sifat *unbiased* (tidak bias)

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= E[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}] \\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^TE[\mathbf{y}] \\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \\ E(\hat{\boldsymbol{\beta}}) &= \boldsymbol{\beta} \end{aligned}$$

(b) Estimator $\hat{\boldsymbol{\beta}}$ mempunyai variansi yang minimum