Modul Praktikum I

Judul : Analisis Korelasi Pearson

Tujuan :
1. Untuk mengetahui hubungan linear antara kedua peubah dalam kaitannya dengan regresi linear sederhana adalah peubah bebas dan peubah tak bebas
2. Untuk menggambarkan tingkat keeratan hubungan linear antara dua peubah atau lebih
3. Untuk mendeteksi adanya multikolinearitas antar peubah bebas.

Dasar Teori

Korelasi Pearson

Analisis ini digunakan untuk peubah yang berskala interval dan rasio. Jika kita memiliki pasangan data $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ dan apabila data ini ditebar pada diagram ternyata membentuk seperti garis lurus.

Korelasi digunakan untuk mengetahui hubungan linear antara kedua peubah, dalam kaitannya dengan model regresi linear sederhana. Kedua peubah tersebut adalah peubah bebas dengan peubah tak bebas. Ukuran korelasi linear antara dua peubah yang paling banyak digunakan adalah *koefisien korelasi pearson*. *Koefisien korelasi pearson* digunakan untuk mengukur tingkat kerataan hubungan linear antara dua peubah. *Koefisien korelasi pearson* antara dua peubah X dan Y adalah:

$$r_{yx} = r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

Pada kasus regresi linear sederhana, *koefisien korelasi pearson* dapat dihitung dengan menggunakan rumus $r_{xy} = \sqrt{R^2}$. Untuk menguji signifikansi hubungan tersebut, maka perlu dilakukan uji signifikansi. Statistik uji yang digunakan pada uji signifikansi *koefisien korelasi pearson* adalah statistik uji T, yaitu:

$$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Nilai koefisien korelasi berkisar antara -1 hingga +1, yaitu $-1 \leq r_{xy} \leq 1$. Nilai korelasi negatif berarti hubungan antara dua peubah adalah negatif. Artinya, apabila salah satu peubah menurun, maka peubah lainnya akan meningkat. Sebaliknya, bila korelasi positif, berarti hubungan antara kedau peubah adalah positif. Artinya, apabila salah satu peubah meningkat, maka peubah lainnya meningkat pula. Suatu hubungan antara dua peubah dikatakan berkorelasi kuat apabila makin mendekati 1 atau $|-1|$. Sebaliknya, hubungan antar dua peubah dikatakan lemah apabila semakin mendekati 0 (nol).

Hipotesis

$H_0 \,:\, \rho = 0$

$H_1 \,:\, \rho \neq 0$, dimana ρ adalah korelasi antara 2 peubah.

Daerah Penolakan

$p_{value} < \alpha$

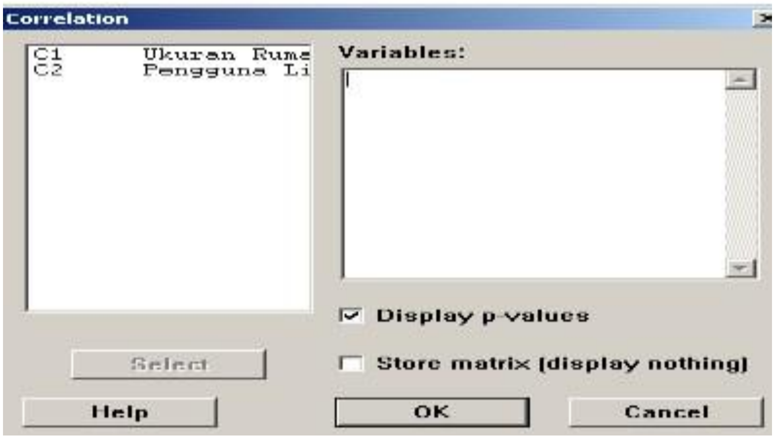
Contoh Kasus

Berikut data penggunaan listrik per bulan:

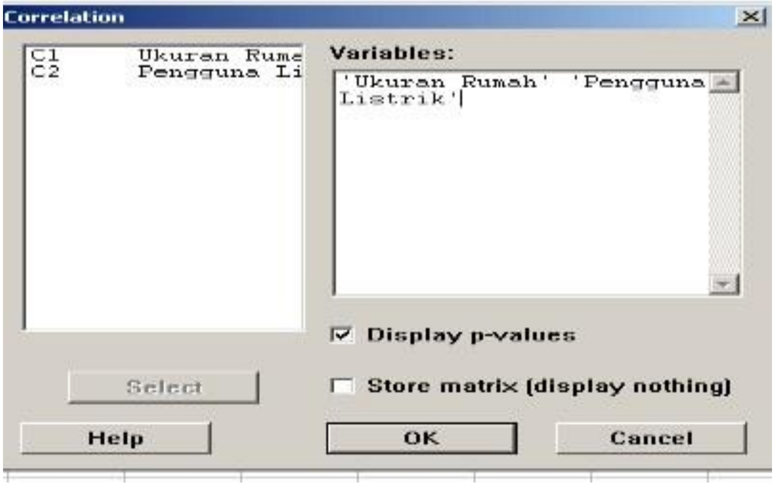
Ukuran Rumah (kaki2)	Penggunaan Listrik
1290	1182
1350	1172
1470	1264
1600	1493
1710	1671
1840	1711
1980	1804
2230	1840
2400	1956
2930	1954

- Langkah-langkah menghitung korelasi antara dua peubah dengan menggunakan minitab:

1. Pilih **Basic Statistic** > **Correlation**



2. Pada kotak dialog berikut, letakkan peubah ukuran rumah dan penggunaan listrik per bulan pada kolom di bawah **Variables**



3. Untuk menampilkan p-value, pilih **Display p-values**. Kemudian klik OK.

   Output :

   Pearson correlation of Ukuran Rumah and Pengguna Listrik = 0,898

   P-Value = 0,000

- Langkah-langkah menghitung korelasi antara dua peubah dengan menggunakan R: Menghitung nilai korelasi beserta pengujiannya dalam *software R*, dapat menggunakan sintaks:

```
>cor.test(data_pertama, data_kedua, methode="pearson")
```

```
> cor.test(prak1$Ukuran.Rumah, prak1$Penggunaan.Listrik, method = "pearson")
```

Output

```
        Pearson's product-moment correlation

data:  prak1$Ukuran.Rumah and prak1$Penggunaan.Listrik
t = 6, df = 8, p-value = 4e-04
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.617 0.976
sample estimates:
  cor
0.898
```

Latihan :

1. Kadar air campuran basah suatu produk (X) diperkirakan berpengaruh pada kepadatan produk akhirnya (Y). Dalam suatu percobaaan, kadar air campuran dikendalikan dan kemudian kepadatan produk akhirnya diukur. Data yang berhasil dikumpulkan adalah sebagai berikut:

X	4.7	5.0	5.2	5.2	5.9	4.7	5.9	5.2	5.3	5.9	5.6	5.0
Y	3	3	4	5	10	2	9	3	7	6	6	4

   a. Hitung koefisien korelasi. Adakah hubungan antara kadar air campuran basah dan kepadatan produk?
   b. Lakukan pengujian terhadap koefisien korelasi

2. Data mengenai umur dan harga jual kendaraan roda empat, umur dinyatakan dalam tahun, sedangkan harga jual dinyatakan dalam juta rupiah. Datanya adalah sebagai berikut:

Umur	12	10	6	4	2	1	3
Harga jual	15	18.5	22.7	32	46	62	43

   a. Hitung koefisien korelasi. Adakah hubungan antara umur kendaraan dan harga jual kendaraan?
   b. Lakukan pengujian terhadap koefisien korelasi.

Modul Praktikum II

Judul : Regresi Linear Sederhana

Tujuan : 1. Membentuk model regresi linear sederhana.

2. Menginterpretasikan model regresi yang dihasilkan.
3. Menentukan kebaikan model yang dihasilkan.

Dasar Teori

Regresi Linear Sederhana

Analisis regresi digunakan untuk *meramalkan nilai peubah respon* berdasarkan nilai satu atau beberapa *peubah prediktor* (peubah bebas). Istilah regresi telah digunakan oleh Sir Francis Galton (1822-1911) yang membandingkan tinggi badan anak laki-laki dengan tinggi badan ayahnya. Galton menunjukkan bahwa tinggi badan anak laki-laki dari ayah yang tinggi setelah beberapa generasi cenderung mundur (*regressed)* mendekati nilai tengah populasi.

Prosedur peramalan peubah respon (Y) berdasarkan peubah bebas (X) adalah:

- Buatlah diagram pencar (*scatter plot*) antara peubah Y dan peubah X.
- Tentukan bentuk hubungan fungsional antara peubah Y dan peubah X.
- Tentukan parameter pada hubungan fungsional tersebut.

Jika hubungan fungsional antara peubah X dan peubah Y adalah fungsi linear, maka modelnya adalah:

$$Y_i = \beta_0 + \beta_1 X_1 + \varepsilon_i \qquad (1)$$

Parameter β_0 dan β_1 diestimasi (diduga) menggunakan rumus:

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

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 X_1 \qquad (3)$$

Asumsi yang harus dipenuhi agar model regresi linear sederhana pada persamaan (1) adalah:

- Nilai galat ke-i tidak berkorelasi dengan galat ke-j, atau *tidak ada autokorelasi* pada galat.
- Variasi (keragaman) ε_i homogen untuk setiap nilai X_i, atau *tidak ada heteroskeastisitas*.
- Galat berdistribusi normal dengan rata-rata 0 dan variansi σ^2.

Nilai-nilai parameter regresi β_0 dan β_1 perlu diuji apakah berbeda signifikan dengan nol, yaitu perlu diuji apakah β_0 = 0 dan β_1 = 0

Uji Serempak dan Uji Individual

Selanjutnya, dapat diestimasi nilai-nilai parameter β_0 dan β_1 pada model regresi linear sederhana $Y_i = \beta_0 + \beta_1 X_1 + \varepsilon_i$. Hipotesis yang diuji adalah:

a. H_0 : Variabel bebas tidak berpengaruh terhadap variabel tak bebas)
   H_1 : $\beta_1 \neq 0$ (Variabel bebas berpengaruh terhadap variabel tak bebas)
   Untuk menguji hipotesis ini dapat digunakan uji t atau uji F melalui tabel ANOVA seperti Tabel 1.

Daerah Penolakan

$p_{value} < \alpha$ atau

$F_{hitng} > F_{\alpha, (k, n-2)}$

Statistik uji F dihitung dengan menggunakan tabel ANOVA berikut:

Tabel 2.1 Analisis variansi regresi linear sederhana

Sumber Keragaman	Jumlah Kuadrat	Derajat Bebas	Rataan Kuadrat	F
Regresi	$JKR = \sum(\hat{y}_i - \bar{y})^2$	1	RKR = JKR/1	RKR/ S^2
Galat	$JKG = \sum(y_i - \hat{y})^2$	n-2	S^2=JKG/(n-2)	
Total	$JKT = \sum(y_i - \bar{y})^2$	n-1		

(Drapper, 1992)

Koefisien determinasi digunakan untuk mengukur variasi nilai respon Y yang dapat dijelaskan ileh peubah bebas X, dan didefinisikan sebagai berikut:

$$R^2 = \frac{JKR}{JKT} = 1 - \frac{JKG}{JKT}$$

Kisaran nilai R^2 adalah $0 \leq R^2 \leq 1$.

b. $H_0: \beta_0 = 0$
   $H_1: \beta_0 \neq 0$
   Untuk menguji hipotesis ini digunakan uji t.

Prosedur Kerja

Data

Contoh

Regresi Linear Sederhana

Dalam tabel berikut, Y menyatakan banyaknya suatu senyawa kimia yang larut dalam 100 gram air pada berbagai suhu (dinyatakan dalam X).

X (^{o}C)	Y (gram)		
0	8	6	8
15	12	10	14
30	25	21	24
45	31	33	28
60	44	39	48
75	48	51	44

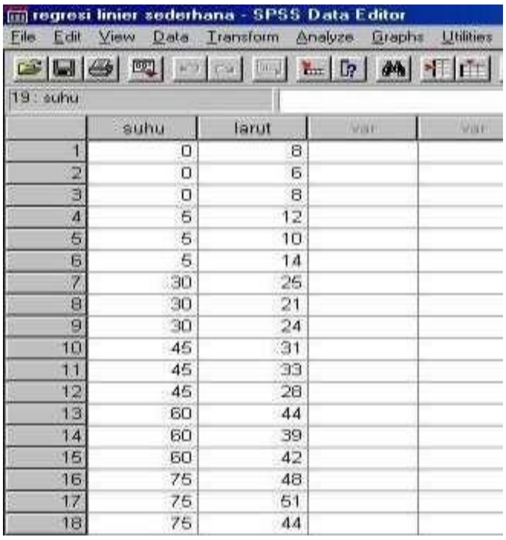
Berdasarkan data tersebut:

a. Buatlah diagram pencar antara Y dan X
b. Tentukan persamaan garis regresinya
c. Dugalah banyaknya senyawa kimia yang akan larut dalam 100 gram pada suhu 50^{o}C

- **Prosedur SPSS**

✓ Prosedur pembuatan diagram pencar menggunakan SPSS:

1. Buatlah dua peubah baru, misalkan diberi nama *suhu* dan *larut*. **Tipe** masing-masing peubah adalah “Numeric” sengan **Decimals** adalah “0”. Pada kedua peubah tersebut, berturut-turut berilah Label “suhu (C)”, “banyak senyawa kimia yang larut dalam 100 gram air (gram)”.
2. Entrikan data pada masing-masing peubah, sehingga diperoleh tampilan seperti gambar berikut:

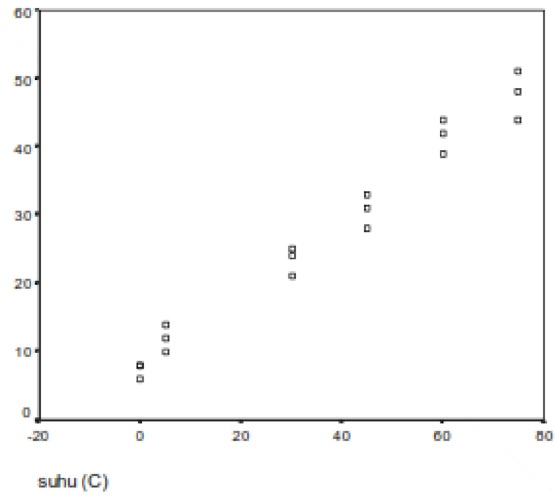


	suhu	larut	var	var
1	0	8		
2	0	6		
3	0	8		
4	5	12		
5	5	10		
6	5	14		
7	30	25		
8	30	21		
9	30	24		
10	45	31		
11	45	33		
12	45	28		
13	60	44		
14	60	39		
15	60	42		
16	75	48		
17	75	51		
18	75	44		

Gambar 2.1 Tampilan dan model regresi linear sederhana

3. Untuk membuat diagram pencar (*scatter plot)*, pilih menu **Graphs>Scatter.** Klik pilihan **Define**. Masukkan peubah "kelarutan senyawa kimia dalam 100 gram air (gram)" ke kotak **Y Axis**. Masukkan peubah "suhu (C)" ke kotak **X Axis**. Selanjutnya klik **OK**.

   Diagram pencar yang diperoleh dapat dilihat pada gambar berikut:



Gambar 2.2 Diagram pencar kelarutan senyawa kimia dalam 100 gram air (gram) vs suhu

✓ Prosedur uji analisis linear sederhana

1. Klik menu **Analyze>Regression>Linear**
2. Masukkan peubah "banyak senyawa kimia yang larut dalam 100 gram air (gram)" ke dalam kotak **Dependant**. Masukkan peubah "suhu (C)" ke kotak **Independent(s).**
3. Klik **OK**.

Output yang diperoleh adalah:

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	X[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Y

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.980[a]	.960	.958	3.19921

a. Predictors: (Constant), X

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3960.686	1	3960.686	386.978	.000[a]
	Residual	163.759	16	10.235		
	Total	4124.444	17			

a. Predictors: (Constant), X

b. Dependent Variable: Y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.730	1.337		4.286	.001
	X	.579	.029	.980	19.672	.000

a. Dependent Variable: Y

Hasil uji F dapat dilihat pada tabel ANOVA. Terlihat bahwa signifikansi pengujian (Sig.) = 0,000. Dengan menggunakan α = 0,05, karena nilai Sig < α, maka hipotesis nol pada poin (a) ditolak. Hal ini berarti bahwa suhu berpengaruh terhadap kelarutan senyawa kimia tersebut dalam air. Selain cara pengujian dalam tabel ANOVA dapat juga digunakan uji t seperti terlihat pada tabel **Coefficients**. Terlihat bahwa untuk **Model: suhu (C)** mempunyai nilai Sig = 0,000 < α yang berarti hipotesis nol pada poin (a) ditolak. Hal ini berarti bahwa suhu berpengaruh terhadap kelarutan senyawa kimia tersebut dalam air.

Selanjutnya, hasil pengujian parameter regresi β_0 = 0 dengan menggunakan uji t dapat dilihat pada tabel **Coefficients** yaitu untuk **Model: (Constant)** mempunyai nilai Sig = 0,000. Dengan menggunakan α = 0,05, karena nilai Sig < α, maka hipotesis nol pada poin (b) ditolak. Hal ini berarti bahwa $\beta_0 \neq 0$. Oleh karena itu model persamaan yang menyatakan hubungan antara kelarutan senyawa kimia tersebut dalam air dan suhu adalah:

$$\hat{Y} = 5{,}730 + 0{,}579X$$

Interpretasi model:

β_0 = Saat suhu bernilai 0, maka kelarutan senyawa kimia dalam air akan bernilai 5,730 gram.

β_1 = Setiap kenaikan 1°C suhu, maka kelarutan senyawa kimia dalam air akan naik sebesar 0,579 gram.

Model tersebut dapat digunakan untuk mengestimasi (menduga) rata-rata banyaknya senyawa kimia yang akan larut dalam 100 gram air pada suhu 50°C yaitu:

$\hat{Y}$ = 5,730 + 0,579 (50) = 34,680 gram

Hal lain yang perlu diperhatikan adalah nilai koefisien determinasi R^2. Nilai R^2 = 0,96 menyatakan bahwa 96% keragaman data kelarutan senyawa kimia dalam air disebabkan oleh faktor suhu, sedangkan sisanya disebabkan oleh faktor lain yang tidak dimasukkan dalam model tersebut.

- Prosedur R

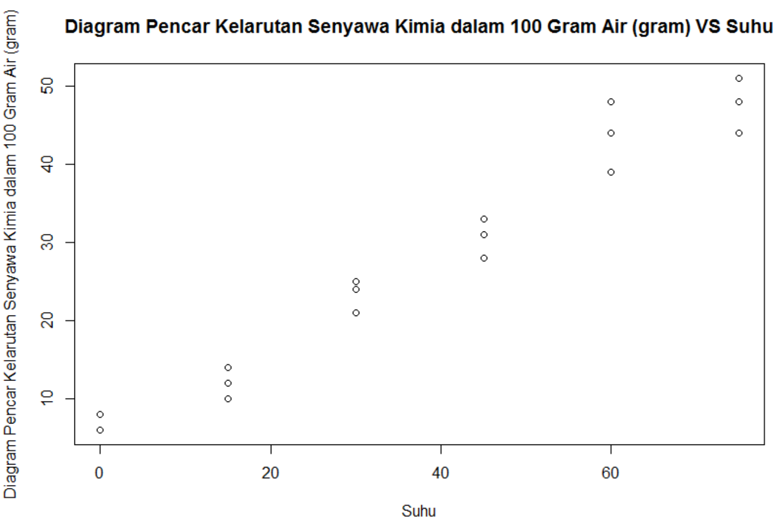
✓ Prosedur pembuatan diagram pencar pada R

1. Langkah pertama adalah dengan memasukkan data pada R, seperti berikut

```
> suhu=c(rep(0, times=3), rep(15, times=3), rep(30, times=3), rep(45, times=3), rep(60, times=3), rep(75, times=3))
> banyaknyasenyawa=c(8, 6, 8, 12, 10, 14, 25, 21, 24, 31, 33, 28, 44, 39, 48, 48, 51, 44)
```

2. Lalu membuat diagram pencar dengan menggunakan sintaks

```
>plot(dataX, dataY, main=”JUDUL”, xlab=”nama sumbu X”, ylab=”nama sumbu Y)
> plot(suhu, banyaknyasenyawa, main="Diagram Pencar Kelarutan Senyawa Kimia dalam 100 Gram Air (gram) VS Suhu", xlab="Suhu", ylab="Diagram Pencar Kelarutan Senyawa Kimia dalam 100 Gram Air (gram)")
```



Gambar 3.3 Diagram pencar kelarutan senyawa kimia dalam 100 gram air (gram) vs suhu

- ✓ Prosedur uji analisis regresi linear sederhana

1. Untuk mencari persamaan regresi, kita dapat menggunakan sintaks sebagai berikut:

```
> lm(dataY~dataX)
> model=(lm(banyaknyasenyawa~suhu))
> model

Call:
lm(formula = banyaknyasenyawa ~ suhu)

Coefficients:
(Intercept)         suhu
      5.730        0.579
```

Persamaan regresinya adalah $\hat{Y} = 5{,}730 + 0{,}579X$

2. Lalu dapat dilakukan uji F dan uji t untuk melihat keberpengaruhan peubah bebas terhadap peubah tak bebas.

```
> summary(model)

Call:
lm(formula = banyaknyasenyawa ~ suhu)

Residuals:
   Min     1Q Median     3Q    Max
-5.159 -1.944 -0.073  1.884  7.527

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.73016    1.33681   4.286 0.000566 ***
suhu         0.57905    0.02944  19.672 1.23e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.199 on 16 degrees of freedom
Multiple R-squared:  0.9603,    Adjusted R-squared:  0.9578
F-statistic:   387 on 1 and 16 DF,  p-value: 1.235e-12
```

Hasil uji F dapat dilihat pada tabel untuk nilai F dan p_{value} dapat dilihat pada bagian **F-statistic** dan **p-value**. Untuk koefisien dari β_0 dan β_1 dapat dilihat pada bagian kolom **Estimate**. Hasil uji individual (uji t) untuk masing-masing koefisien dapat dilihat pada bagian kolom **Pr(>|t|)**. Koefisien determinasi dapat dilihat pada bagian **R-squared**.

Latihan :

Regresi linear sederhana

1. Kadar air campuran basah suatu produk (X) diperkirakan berpengaruh pada kepadatan produk akhirnya (Y). Dalam suatu percobaaan, kadar air campuran dikendalikan dan kemudian kepadatan produk akhirnya diukur. Data yang berhasil dikumpulkan adalah sebagai berikut:

X	4.7	5.0	5.2	5.2	5.9	4.7	5.9	5.2	5.3	5.9	5.6	5.0
Y	3	3	4	5	10	2	9	3	7	6	6	4

a. Buatlah diagram pencar antara Y dan X.
b. Dengan menggunakan model $Y_i = \beta_0 + \beta_1 X_1 + \varepsilon_i$, tentukan nilai dugaan kuadrat terkecil bagi β_0 dan β_1. Tentukan persamaan peramalannya, kemudian interpretasikan persamaan tersebut.
c. Susunlah tabel analisis ragam dan ujilah hipotesis $H_0 : \beta_1 = 0$ dengan taraf nyata α = 0,05.
d. Hitunglah koefisien determinasinya dan jelaskan artinya.

2. Seorang dokter ingin mengetahui apakah ada hubungan antara berat badan seseorang dengan tinggi badan seseorang. Untuk keperluan tersebut selanjutnya dilakukan penelitian terhadap 10 orang dengan data sebagai berikut:

Tinggi (cm)	140	132	136	143	156	158	161	171	166	160
Berat Badan (kg)	42	39	52	47	41	68	46	48	57	62

a. Buatlah diagram pencar antara Y dan X.
b. Dengan menggunakan model $Y_i = \beta_0 + \beta_1 X_1 + \varepsilon_i$, tentukan nilai dugaan kuadrat terkecil bagi β_0 dan β_1. Tentukan persamaan peramalannya, kemudian interpretasikan persamaan tersebut.
c. Susunlah tabel analisis ragam dan ujilah hipotesis $H_0 : \beta_1 = 0$ dengan taraf nyata α = 0,05.
d. Hitunglah koefisien determinasinya dan jelaskan artinya.

Modul Praktikum III

Judul : Diagnostik Model Regresi (pengujian asumsi-asumsi regresi)

Tujuan : 1. Membentuk model regresi linear sederhana.

2. Menginterpretasikan model regresi yang dihasilkan.
3. Menentukan kebaikan model yang dihasilkan.

Dasar Teori

Diagnostik pendugaan model dilakukan melalui analisis sisaan (residuals) untuk memeriksa apakah asumsi-asumsi yang mendasari model regresi terpenuhi. Asumsi-asumsi yang harus dipenuhi sebelum membuat model dalam analisis regresi adalah:

1. Data menyebar normal
2. Antar pengamatan saling bebas (tidak ada autokorelasi antar pengamatan)
3. Ragam galat homogen (konstan/homoskedastisitas)
4. Tidak terdapat multikolinearitas di antara peubah bebasnya

Untuk regresi linear sederhana, model harus memenuhi asumsi 1, asumsi 2, dan asumsi. Sedangkan untuk regresi linear berganda, keempat asumsi harus terpenuhi. Jika asumsi-asumsi tersebut dilanggar atau tidak terpenuhi, maka akan berdampak pada pendugaan parameter, hasil pengujian parameternya serta pengambilan kesimpulan. Beberapa metode yang dapat digunakan untuk mendeteksi asumsi-asumsi tersebut antara lain:

Tabel 4.1 Asumsi dasar regresi linear dan metode-metode yang digunakan untuk mendekati asumsi tersebut.

Asumsi	Metode untuk mendeteksi
Kenormalan data	Uji Kolmogorv-Smirnov, Uji Anderson-Darling, Uji Ryan Joiner.
Autokorelasi antar pengamatan	Durbin-Watson
Ragam antar pengamatan	Breusch-Pagan Test, Goldfeld-Quandt Test, Plot sisaan vs y actual atau y dugaan
Multikolinearitas	*Variance Inflation Factor* (VIF)

Prosedur Kerja

Data

Contoh

Regresi Linear Sederhana

Dalam tabel berikut, Y menyatakan banyaknya suatu senyawa kimia yang larut dalam 100 gram air pada berbagai suhu (dinyatakan dalam X).

X (^{o}C)	Y (gram)		
0	8	6	8
15	12	10	14
30	25	21	24
45	31	33	28
60	44	39	48
75	48	51	44

Berdasarkan data tersebut:

a. Lakukan pengujian asumsi
b. Sebutkan cara mengatasi jika asumsi tidak terpenuhi

✓ **Prosedur Minitab**

Prosedur untuk memeriksa asumsi-asumsi regresi, Model dan ANOVA dengan menggunakan Minitab:

1. Masukkan data untuk peubah X dan Y
2. Pilih menu **Stat>Regression>Regression.** Pada **Graph** klik **normal plot of residual** dan **residual versus fits**. Kemudian pada **Options**, klik **Durbin-Watson** dan **VIF.**

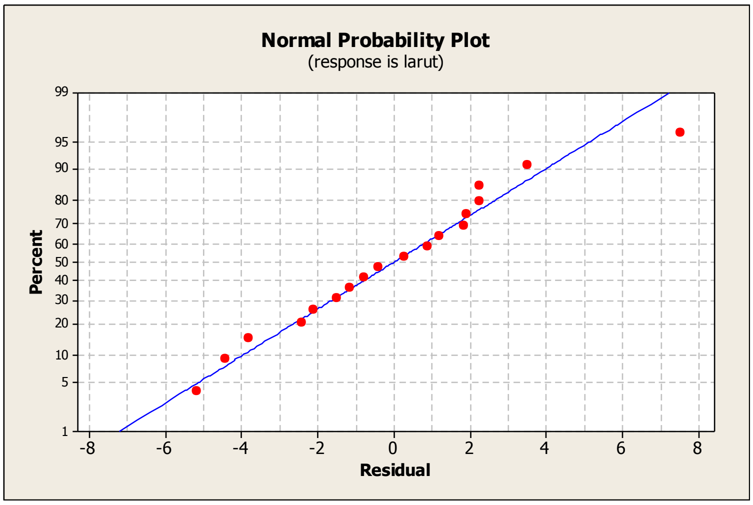
Output

Regression Analysis: larut versus suhu

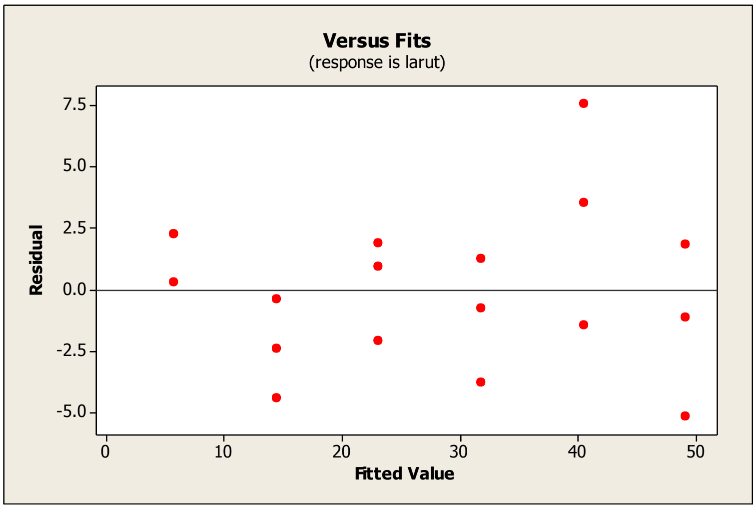
```
The regression equation is
larut = 5.73 + 0.579 suhu


Predictor       Coef  SE Coef      T      P    VIF
Constant       5.730    1.337   4.29  0.001
suhu         0.57905  0.02944  19.67  0.000  1.000


S = 3.19921   R-Sq = 96.0%   R-Sq(adj) = 95.8%
```

```
Analysis of Variance

Source          DF      SS      MS       F      P
Regression       1  3960.7  3960.7  386.98  0.000
Residual Error  16   163.8    10.2
Total           17  4124.4


Unusual Observations

Obs  suhu   larut     Fit  SE Fit  Residual  St Resid
 15  60.0  48.000  40.473   1.004     7.527     2.48R

R denotes an observation with a large standardized residual.


Durbin-Watson statistic = 2.47372
```

Normplot of Residuals for larut

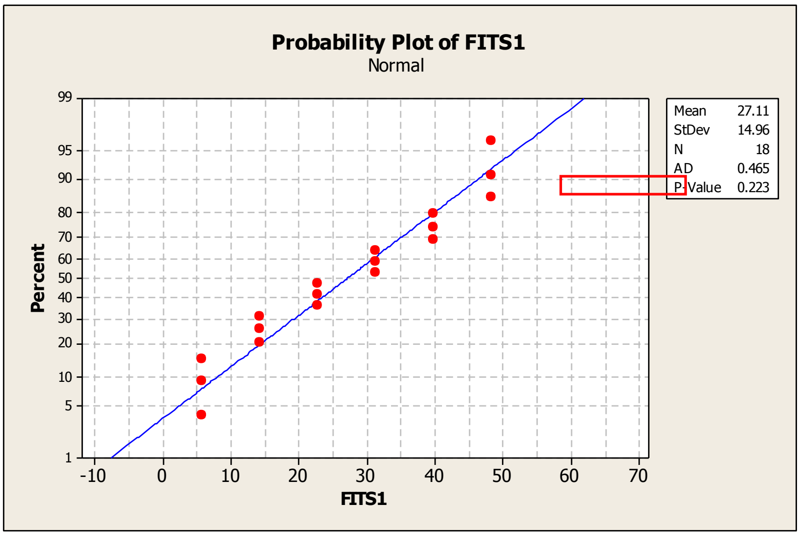


Residuals vs Fits for larut



<u>Cara lain uji kenormalan data:</u>

Pilih menu **Stat>Basic Statistics>Normally test....** Pada kotak **Normally test**, pilih pengujian menggunakan uji **Anderson-Darling**, **Kolmogorov-Smirnov**, atau **Ryan Joiner**.



a. Model dugaan yang terbentuk adalah $\hat{Y} = 5{,}73 + 0{,}579X_1$

b. Pengujian asumsi

1. Asumsi kenormalan data

    - Uji formal (Anderson-Darling)

    Pada pengujian Anderson-Darling, terdapat asumsi seperti berikut:

H_0: Data menyebar normal

H_1: Data tidak menyebar normal

Dengan taraf nyata α = 0,05, karena nilai p_{value} (0,223) > α (0,05), maka tidak cukup bukti untuk tolak H_0 (lihat output **Probability Plot of FITS1**). Hal ini berarti, data menyebar normal dan asumsi kenormalan terpenuhi.

    - Uji non-formal (*normal probability plot*)

    Jika data berada pada sekitar garis, maka data menyebar normal. Pada plot **Normal Probability Plot**, terlihat bahwa data berada pada sekitar garis. Maka, dapat disimpulkan bahwa data menyebar normal dan asumsi kenormalan terpenuhi (lihat output **Normplot of Residuals for larut**).

2. Asumsi autokorelasi antar pengamatan

    Berikut adalah kriteria dalam uji Durbin-Watson dengan hipotesis:

H_0: Tidak ada autokorelasi positif/negatif

H_1: Terdapat autokorelasi positif/negatif

Tabel 4.1 Uji statistik Durbin-Watson

Nilai Statistik Durbin-Watson	Hasil
$0 < d < d_L$	Menolak H_0; ada autokorelasi positif
$d_L \leq d \leq d_U$	Daerah keragu-raguan; tidak ada keputusan
$d_U \leq d \leq 4 - d_U$	Tidak tolak H_0, tidak ada autokorelasi positif/negatif
$4 - d_U \leq d \leq d_L$	Daerah keragu-raguan; tidak ada keputusan
$4 - d_L \leq d \leq 4$	Menolak H_0; ada autokorelasi negatif

Nilai Durbin-Watson (d) adalah 2,50451 (lihat output **Regression Analysis: larut versus suhu)**. Dengan taraf nyata α =0,05, nilai d_U adalah 1,3913 dan $4 - d_U$ = 2,6087 (lihat **tabel Durbin-Watson**). Karena nilai $d_U \leq d \leq 4 - d_U$ maka tidak dapat menolak H_0. Hal ini berarti tidak ada autokorelasi positif/negatif dan asumsi autokorelasi antar pengamatan terpenuhi.

3. Asumsi ragam galat homogen
   - Uji non-formal (*residuals vs fits*)

Ragam galat dikatakan homogen jika **plot residuals vs fits** tidak membentuk pola. Pada tersebut, terlihat bahwa data tidak membentuk pola tertentu, sehingga dapat disimpulkan bahwa ragam galat homogen dan asumsi ragam galat homogen terpenuhi (lihat output **Residuals vs Fits for larut)**.

Ketiga asumsi telah terpenuhi, sehingga dapat disimpulkan bahwa model dugaan $\hat{Y} = 5{,}73 + 0{,}579X_1$ telah layak untuk digunakan.

✓ **Prosedur R**

1. Asumsi kenormalan data

Uji kenormalan data dapat dilakukan dengan menggunakan uji Shapiro-Wilk (mirip dengan uji Ryan Joiner). Uji Shapiro-Wilk dapat menggunakan sintaks berikut:

```
> model=(lm(banyaknyasenyawa~suhu))
> galat=residuals(model)
> library(nortest)
> ad.test(galat)

        Anderson-Darling normality test

data:  galat
A = 0.2006, p-value = 0.8603
```

Hipotesis:

H_0 = Data menyebar normal

H_1 = Data tidak menyebar normal

Dengan taraf nyata α = 0,05, karena nilai p_{value} (0,8603) > α (0,05), maka tidak cukup bukti untuk tolak H_0. Hal ini berarti, data menyebar normal dan asumsi kenormalan terpenuhi.

2. Asumsi autokorelasi antar pengamatan

Pengujian autokorelasi dapat dilakukan dengan uji Durbin-Watson. Uji Durbin-Watson dapat menggunakan sintaks:

```
> library(lmtest)
> dwtest(model)

        Durbin-Watson test

data:  model
DW = 2.4737, p-value = 0.7825
alternative hypothesis: true autocorrelation is greater than 0
```

Hipotesis:

H_0: Tidak ada autokorelasi positif/negatif

H_1: Terdapat autokorelasi positif/negatif

Dengan taraf nyata α = 0,05, karena nilai p_{value} (0,7825) > α (0,05), maka tidak cukup bukti untuk tolak H_0. Hal ini berarti, tidak ada autokorelasi postif/negatif dan asumsi autokorelasi antar pengamatan terpenuhi.

3. Uji ragam galat homogen

Pengujian ragam galat homogen dapat dilakukan dengan uji Breusch-Pagan. Uji Breusch-Pagan menggunakan sintaks:

```
> library(lmtest)
> bptest(model)

	studentized Breusch-Pagan test

data:  model
BP = 1.6148, df = 1, p-value = 0.2038
```

Hipotesis:

H_0: Ragam galat homogen

H_1: Ragam galat tidak homogen

Dengan taraf nyata α = 0,05, karena nilai p_{value} (0,2038) > α (0,05), maka tidak cukup bukti untuk tolak H_0. Hal ini berarti, ragam galat homogen dan asumsi ragam galat homogen terpenuhi.

Ketiga asumsi telah terpenuhi, sehingga dapat disimpulkan bahwa model dugaan $\hat{Y} = 5{,}73 + 0{,}579X_1$ telah layak untuk digunakan. Jika salah satu asumsi tidak terpenuhi, maka dilakukan tranformasi.

Latihan:

1. Data Berikut dikumpulkan untuk menentukan persamaan regresi hubungan antara panjang bayi dengan umur dan berat waktu lahir

Panjang bayi, Y (cm)	Umur, X_1 (hari)	Bobot lahir, X_2 (kg)
57,5	78	2,75
52,8	79	2,15
61,3	77	4,41
67,0	88	5,52
53,5	67	3,21
62,7	80	4,32
56,2	74	2,31

68,5	94	4,30
69,2	102	3,71

a. Lakukan pengujian asumsi
b. Sebutkan cara mengatasi jika asumsi tidak terpenuhi

2. Data populasi daerah per tahun, datanya sebagai berikut:

Tahun	Populasi
1790	3.929
1800	5.308
1810	7.239
1820	9.638
1830	12.866
1840	17069
1850	23191
1860	31.443
1870	39.818
1880	50.155

a. Lakukan pengujian asumsi
b. Sebutkan cara mengatasi jika asumsi tidak terpenuhi

Note:

Untuk pengujian multikolinearitas pada regresi linear berganda (praktikkum 4), dapat dilakukan dengan melihat nilai variance inflation factor (VIF). Jika nilai VIF lebih dari 10 (>10), maka terdapat multikolinearitas.

1. Pada Minitab, dapat dilakukan dengan langkah-langkah sebagai berikut untuk mendapatkan nilai VIF adalah: Pilih menu **Stat>Regression>Regression. P**ada **Options**, klik **VIF.**
2. Sedangkan pada R, dapat menggunakan sintaks sebagai berikut:

```
>VIF(model) #variabel bebas harus lebih dari 1
```

Modul Praktikum IV

Judul : Diagnostik Data Berpengaruh dan Pencilan (Outlier)

Tujuan : Untuk mengetahui data pengamatan yang berpengaruh dan ada tidaknya pencilan

Dasar Teori

Data mempunyai peranan yang penting dalam setiap pendugaan parameter. Dalam perhitungan statistik contoh (*sampel statistics*) seperti rataan, ragam, dan standar deviasi, setiap pengamatan mempunyai bobot yang sama dalam menentukan nilai statistik contoh, walaupun data yang jelek (*bad data*) dapat mengurangi ketepatan nilai statistik sebagai penduga parameter populasi.

Keragaman contoh sangat dipengaruhi oleh variasi data, apalagi jika ada suatu pengamatan yang jelek. Jika pengamatan yang jelek dilibatkan dalam perhitungan, maka rataan contoh akan berbeda banyak dari rataan contoh tanpa pengamatan itu. Dalam regresi linear juga akan terjadi hal serupa. Jika pengamatan yang jelek dilibatkan dalam penentuan model regresi, maka model regresi tersebut akan menggambarkan hubungan yang kurang baik. Pengamatan seperti ini termasuk ke dalam kategori pengamatan yang berpengaruh.

Pengamatan yang berpengaruh dapat dideteksi dengan besaran nilai *R-student*, nilai h_{ii} (*leverage*) yang besar menunjukkan bahwa pengamatan ke-i berada jauh dari pusat data berdasarkan peubah X. Nilai t_i dibandingkan dengan nilai *t-student* dengan derajat bebas n-p-1.

Nilai *R-student* menunjukkan apakah ada pengamatan yang tidak wajar (*unusual observation*) dari sisi peubah Y, sedangkan h_{ii} menunjukkan apakah ada pengamatan yang tidak wajar dari sisi peubah X. Pengaruh pengamatan ke-i terhadap nilai dugaan Y dapat diketahui melalui besaran DFFITS. Pengaruh pengamatan ke-i terhadap nilai koefisien regresi ke-j dapat diketahui melalui besaran DFBETAS.

Pengamatan pencilan (*outlier)* merupakan pengamatan yang berada jauh dari pusat data dan dapat merupakan pengamatan yang berpengaruh. Pencilan dapat juga dideteksi dengan *R-student*, dimana pencilan ini adalah pencila Y. Pencilan belum tentu berpengaruh, tergantung pada nilai h_{ii}. Pengamatan dengan nilai h_{ii} yang besar belum tentu berpengaruh dan pengamatan berpengaruh belum tentu sebagai pencilan.

Prosedur Kerja

Data

Diketahui data tentang umur (X) dan kandungan plasma poliamina (Y) untuk 25 anak sehat sebagai berikut:

X	Y	X	Y
0	13.44	3	7.94
0	12.84	3	6.01
0	11.91	3	5.14
0	20.09	3	6.90
0	15.60	3	6.77
1	10.11	4	4.86
1	11.38	4	5.10
1	10.28	4	5.67
1	8.96	4	5.75
1	8.59	4	6.23
2	9.83		
2	9.00		
2	8.65		
2	7.85		
2	8.88		

Analisa data pengamatan yang berpengaruh dan data pencilan (*outlier)*.

- ✓ Prosedur Minitab

Untuk data berpengaruh

Pilih menu **Stat>Regression>Regression>Storage.** Pada bagian **Storage**, aktifkan **cook's distance, DFFITS, HI**.

Untuk outlier akan otomatis dapat diketahui.

Output

Regression Analysis: Y versus X

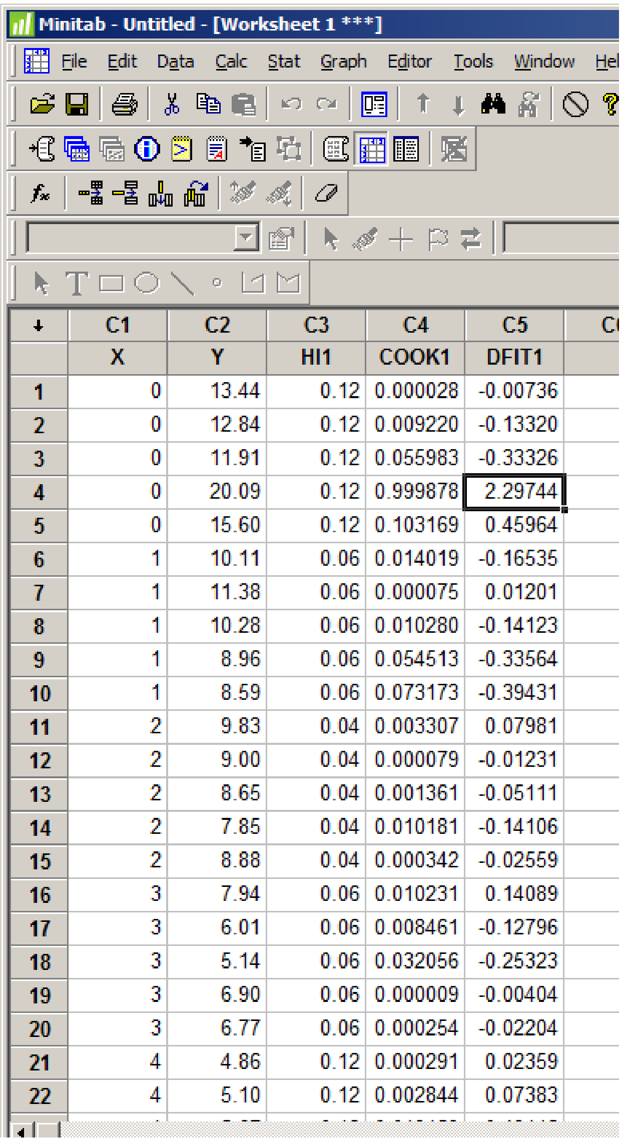
```
The regression equation is
Y = 13.5 - 2.18 X


Predictor      Coef   SE Coef       T       P
```

```
Constant    13.4752   0.6379  21.13  0.000
X           -2.1820   0.2604  -8.38  0.000


S = 1.84135   R-Sq = 75.3%   R-Sq(adj) = 74.3%


Analysis of Variance

Source          DF      SS      MS      F      P
Regression       1  238.06  238.06  70.21  0.000
Residual Error  23   77.98    3.39
Total           24  316.04
```

Unusual Observations

```
Obs     X       Y     Fit  SE Fit  Residual  St Resid
  4  0.00  20.090  13.475   0.638     6.615     3.83R
```

R denotes an observation with a large standardized residual.



↓	C1	C2	C3	C4	C5
	X	Y	HI1	COOK1	DFIT1
1	0	13.44	0.12	0.000028	-0.00736
2	0	12.84	0.12	0.009220	-0.13320
3	0	11.91	0.12	0.055983	-0.33326
4	0	20.09	0.12	0.999878	2.29744
5	0	15.60	0.12	0.103169	0.45964
6	1	10.11	0.06	0.014019	-0.16535
7	1	11.38	0.06	0.000075	0.01201
8	1	10.28	0.06	0.010280	-0.14123
9	1	8.96	0.06	0.054513	-0.33564
10	1	8.59	0.06	0.073173	-0.39431
11	2	9.83	0.04	0.003307	0.07981
12	2	9.00	0.04	0.000079	-0.01231
13	2	8.65	0.04	0.001361	-0.05111
14	2	7.85	0.04	0.010181	-0.14106
15	2	8.88	0.04	0.000342	-0.02559
16	3	7.94	0.06	0.010231	0.14089
17	3	6.01	0.06	0.008461	-0.12796
18	3	5.14	0.06	0.032056	-0.25323
19	3	6.90	0.06	0.000009	-0.00404
20	3	6.77	0.06	0.000254	-0.02204
21	4	4.86	0.12	0.000291	0.02359
22	4	5.10	0.12	0.002844	0.07383

1. Pendeteksian pencilan

Outlier dapat dideteksi dengan melihat pada keluaran bagian **Unusual Observation**. Saat nilai *Standar Residual* lebih besar dari 2 (**Std.Resid** > **2**), maka observasi tersebut merupakan sebuah pencilan. Pada bagian **Unusual Observation** terlihat bahwa observasi ke-4 merupakan sebuah pencilan.

2. Pendeteksian matan berpengaruh

Amatan berpengaruh dapat dideteksi dengan DFFITS. Jika nilai DFFITS dari pencilan lebih besar dari $\frac{2}{\sqrt{\frac{p}{n}}}$. p merupakan banyaknya parameter dalam model, dan n merupakan banyaknya data. $\frac{2}{\sqrt{\frac{p}{n}}}. = \frac{2}{\sqrt{\frac{2}{25}}}. \simeq 7$. Sehingga dapat disimpulkan bahwa observasi ke-4 bukan merupakan amatan yang berpengaruh karena nilai DFFITS observasi ke-4 adalah 2,29744.

✓ Prosedur menggunakan R

1. Langkah pertama adalah dengan mencari galat dari persamaan regresi.

```
> library(DescTools)
> model=lm(Y~X, data=data)
> galat=residuals(model)
```

2. Lalu dicari nilai r-student dan r-standard untuk melihat pencilan

```
> rstandard(model)
          1           2           3           4           5           6 
7           8
-0.02037817 -0.36773327 -0.90613368  3.82947424  1.23010021 -0.66276262 
0.04862052 -0.56753810 
          9          10          11          12          13          14 
15          16
-1.30692845 -1.51418180  0.39841550 -0.06163579 -0.25563332 -0.69905624 
-0.12814922  0.56619376 
         17          18          19          20          21          22 
23          24
-0.51488455 -1.00221000 -0.01635621 -0.08917496  0.06530276  0.20424480 
0.53423215  0.58054616 
         25 
 0.85843025 
> rstudent(model)
          1           2           3           4           5           6 
7           8
-0.01993042 -0.36071219 -0.90247130  6.22148961  1.24470210 -0.65447433 
0.04755425 -0.55899116 
          9          10          11          12          13          14 
15          16
-1.32848164 -1.56073120  0.39100966 -0.06028597 -0.25037025 -0.69107148 
-0.12537718  0.55764832 
         17          18          19          20          21          22 
23          24
-0.50649450 -1.00231081 -0.01599678 -0.08722991  0.06387328  0.19993675 
0.52576159  0.57199170
```

```
          25
0.85334286
```

Jika nilai mutlak r-standard atau mutlak r-student lebih dari 2 (|2|), maka dapat dikatakan bahwa observasi tersebut merupakan pencilan (*outlier*). Observasi ke-4 memiliki nilai r-standard dan r-student lebih dari 2, maka dapat disimpulkan bahwa observasi ke-4 merupakan pencilan.

3. Dapat juga melihat pencilan dengan menggunakan sintaks

```
> Outlier(galat)
     4
6.6148
```

Terlihat bahwa observasi ke-4 merupakan sebuah pencilan.

4. Selanjutnya melihat apakah pencilan tersebut berpengaruh atau tidak dengan melihat nilai DFFITS dengan nilai DFFITS ≃ 7.

```
> influence.measures(model)
Influence measures of
      lm(formula = Y ~ X, data = data) :

      dfb.1_     dfb.X    dffit cov.r   cook.d  hat inf
1   -7.36e-03  6.01e-03 -0.00736 1.242 2.83e-05 0.12
2   -1.33e-01  1.09e-01 -0.13320 1.227 9.22e-03 0.12
3   -3.33e-01  2.72e-01 -0.33326 1.155 5.60e-02 0.12
4    2.30e+00 -1.88e+00  2.29744 0.163 1.00e+00 0.12    *
5    4.60e-01 -3.75e-01  0.45964 1.084 1.03e-01 0.12
6   -1.56e-01  9.55e-02 -0.16535 1.119 1.40e-02 0.06
7    1.13e-02 -6.94e-03  0.01201 1.163 7.54e-05 0.06
8   -1.33e-01  8.15e-02 -0.14123 1.130 1.03e-02 0.06
9   -3.16e-01  1.94e-01 -0.33564 0.996 5.45e-02 0.06
10  -3.72e-01  2.28e-01 -0.39431 0.942 7.32e-02 0.06
11   4.61e-02  5.55e-18  0.07981 1.123 3.31e-03 0.04
12  -7.10e-03 -1.15e-18 -0.01231 1.138 7.91e-05 0.04
13  -2.95e-02  2.58e-17 -0.05111 1.132 1.36e-03 0.04
14  -8.14e-02  3.35e-17 -0.14106 1.091 1.02e-02 0.04
15  -1.48e-02  1.84e-18 -0.02559 1.137 3.42e-04 0.04
16  -8.57e-18  8.13e-02  0.14089 1.131 1.02e-02 0.06
17  -2.57e-17 -7.39e-02 -0.12796 1.136 8.46e-03 0.06
18   6.96e-17 -1.46e-01 -0.25323 1.063 3.21e-02 0.06
19   2.66e-19 -2.33e-03 -0.00404 1.163 8.54e-06 0.06
20  -6.38e-18 -1.27e-02 -0.02204 1.162 2.54e-04 0.06
21  -7.86e-03  1.93e-02  0.02359 1.242 2.91e-04 0.12
22  -2.46e-02  6.03e-02  0.07383 1.238 2.84e-03 0.12
23  -6.47e-02  1.59e-01  0.19415 1.211 1.95e-02 0.12
24  -7.04e-02  1.72e-01  0.21122 1.206 2.30e-02 0.12
25  -1.05e-01  2.57e-01  0.31512 1.164 5.02e-02 0.12
```

Nilai DFFITS observasi ke-4 adalah 2,29744 lebih kecil dari 7, maka dapat disimpulkan bahwa pengamatan tersebut bukan merupakan amatan berpengaruh.

Latihan:

1. Lakukan analisis data pengamatan yang berpengaruh dan data pencilan (*outlier*) pada data di bawah ini

x	y
24.5	138.6
38	198.6
25.5	165.2
36	181.8
99.5	458.7
48.5	195.4
25.5	135.5
37	176.2
32.5	107.2
30	157.6
78	102.7
44	208.8
50	178.4
17.5	298.9
37.5	208.4

2. Dari sebuah survey yang dilakukan di kampung Mulia digunakan untuk mengetahui hubungan antara luas tanah (hektar) dan harganya (Juta). Data tentang luasan dan harga tanah yang diperoleh sebagai berikut :

Luas	Harga
0.75	2.45
0.55	2.20
1.00	2.80
1.25	3.60
2.50	5.80
3.00	7.40
4.50	9.00
3.75	8.50
5.00	10.00
3.25	8.00
3.25	7.50
2.75	6.00
2.75	6.25
2.00	4.00
4.00	8.00

Lakukan analisis data pengamatan yang berpengaruh dan data pencilan (*outlier*).

Modul Praktikum V

Judul : Regresi Linear Berganda

Tujuan : 1. Membentuk model regresi linear berganda .

2. Menginterpretasikan model regresi yang dihasilkan.
3. Menentukan kebaikan model yang dihasilkan.

Dasar Teori

Regresi Linear Berganda

Prosedur regresi linear sederhana dapat dikembangkan jika peubah respon Y dipengaruhi oleh beberapa peubah bebas $X_1, X_2, \dots, X_k$. Bentuk umum model regresi linear berganda adalah:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \quad (1)$$

dengan β_0, β_1, ..., β_k adalah parameter-parameter regresi, yang diduga (diestimasi) menggunakan metode kuadrat terkecil (*Ordinary Least Square*). Menurut Awat N.J. (1995), estimator OLS bersifat *BLUE* (*Best Linear Unbiased Estimator*), yaitu memenuhi syarat sebagai penduga linear terbaik, tak bias dan **berdistribusi normal** serta memenuhi asumsi-asumsi tidak terdapat **Multikolinearitas,** tidak terdapat **Heteroskedastisitas**, dan tidak terdapat **Autokorelasi**.

Asumsi yang harus dipenuhi agar model regresi linear sederhana pada persamaan (1) adalah:

- Nilai galat ke-i tidak berkorelasi dengan galat ke-j, atau *tidak ada autokorelasi* pada galat.
- Variasi (keragaman) ε_i homogen untuk setiap nilai X_i, atau *tidak ada heteroskeastisitas*.
- Galat berdistribusi normal dengan rata-rata 0 dan variansi σ^2.
- Antar peubah bebas tidak terjadi korelasi, atau *tidak terjadinya multikolinearitas*.

Uji Serempak

Untuk menguji model regresi linear berganda pada persamaan (1), digunakan uji F dengan hipotesis sebagai berikut:

- H_0: Peubah bebas secara bersama-sama tidak berpengaruh terhadap variabel respon.
- H_1: Ada setidaknya 1 peubah bebas yang berpengaruh terhadap variabel respon.

Statistik uji F dihitung dengan menggunakan tabel ANOVA berikut:

Tabel 3.1 Analisis variansi regresi linear sederhana

Sumber Keragaman	Jumlah Kuadrat	Derajat Bebas	Rataan Kuadrat	F
Regresi	$JKR = \sum(\hat{y}_i - \bar{y})^2$	k	RKR = JKR/k	RKR/ S^2
Galat	$JKG = \sum(y_i - \hat{y})^2$	n-k-1	S^2=JKG/(n-k-1)	
Total	$JKT = \sum(y_i - \bar{y})^2$	n-1		

Daerah kritisnya adalah $F > F_{\alpha, (k, n-k-1)}$, (Drapper, 1992)

Langkah-langkah pengujiannya adalah:

1. Menentukan tingkat signifikansi pengujian, yaitu α = 0.05.
2. Menghitung statistik uji F melalui tabel ANOVA seperti tabel 2.
3. Menetapkan daerah kritis, yaitu H_0 ditolak jika $F_{hitng} > F_{\alpha, (k, n-k-1)}$. Jika menggunakan bantuan *software*, maka H_0 ditolak jika p_{value} (signifikansi) < α (0.05)

Uji Individual

Penggunaan uji t dimaksudkan untuk melihat pengaruh peubah bebas secara individual terhadap peubah respon. Untuk menguji hipotesis

- H_0: $\beta_i = 0$ (Tidak ada pengaruh secara individu peubah bebas x_i terhadap peubah respon y).
- H_1: $\beta_i \neq 0$ (Ada pengaruh secara individu peubah bebas x_i terhadap peubah respon y).

digunakan statistik uji:

$$t = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)};\ i = 1, 2, ..., k \qquad (2)$$

Langkah-langkah pengujiannya adalah:

1. Menentukan tingkat signifikansi pengujian, yaitu α = 0.05
2. Menentukan daerah kritis yaitu H_0 ditolak jika $t_{hitng} > F_{\alpha/2, df}$. Jika menggunakan bantuan *software*, maka daerah kritisnya adalah p_{value} (signifikansi) < α

Prosedur Kerja

Data

Contoh:

Regresi Linear Berganda

Data Berikut dikumpulkan untuk menentukan persamaan regresi hubungan antara panjang bayi dengan umur dan berat waktu lahir

Panjang bayi, Y (cm)	Umur, X1 (hari)	Bobot lahir,X2 (kg)
57.5	78	2.75
52.8	79	2.15
61.3	77	4.41
67.0	88	5.52
53.5	67	3.21
62.7	80	4.32
56.2	74	2.31
68.5	94	4.30
69.2	102	3.71

Berdasarkan data tersebut:

a. Dugalah model regresi $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$
b. Dugalah panjang rata-rata bayi yang berumur 75 hari dan beratnya waktu lahir 3,15 kg
c. Apakah umur bayi dan bobot bayi secara bersama-sama berpengaruh terhadap panjang bayi?
d. Apakah secara individu umur bayi berpengaruh terhadap panjang bayi dan apakah bobot bayi juga berpengaruh terhadap panjang bayi?
e. Hitung koefisien determinasinya dan jelaskan artinya.

✓ Prosedur penyelesaian menggunakan SPSS

1. Buatlah tiga peubah baru, misalkan diberi nama *umur* (**bertipe** “Numeric", **Decimal** “0”, **Label** “*umur bayi (hari)*”), *bobot* (**bertipe** “Numeric", **Decimal** “2”, **Label** “*bobot bayi lahir (kg)*”), dan *panjang* (**bertipe** “Numeric", **Decimal** “1”, **Label** “*panjang bayi (cm)*”).

2. Entrikan data pada masing-masing peubah, sehingga diperoleh tampilan seperti gambar berikut:

	panjang	umur	bobot
1	57.50	78.00	2.75
2	52.80	79.00	2.15
3	61.30	77.00	4.41
4	67.00	88.00	5.52
5	53.50	67.00	3.21
6	62.70	80.00	4.32
7	56.20	74.00	2.31
8	68.50	94.00	4.30
9	69.20	102.00	3.71

Gambar 3.1 Gambar tampilan data analisis regresi linear berganda

3. Estimasi parameter β_0, β_1, dan β_2 pada model regresi $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$
4. Klik menu **Analyze>Regression>Linear**.
5. Masukkan peubah *panjang* ke kotak **Dependant**. Masukkan peubah *umur* dan *bobot* ke kotak **Independent(s)**.
6. Klik **OK**.

Output yang diperoleh adalah:

Model	Variables Entered	Variables Removed	Method
1	bobot lahir (kg), umur bayi (hari)[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: panjang bayi (cm)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.974[a]	.948	.931	1.66641

a. Predictors: (Constant), bobot lahir (kg), umur bayi (hari)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	304.578	2	152.289	54.841	.000[a]
	Residual	16.662	6	2.777		
	Total	321.240	8			

a. Predictors: (Constant), bobot lahir (kg), umur bayi (hari)

b. Dependent Variable: panjang bayi (cm)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	18.667	4.538		4.113	.006
	umur bayi (hari)	.395	.061	.670	6.532	.001
	bobot lahir (kg)	2.713	.581	.479	4.667	.003

a. Dependent Variable: panjang bayi (cm)

Hipotesis yang diuji adalah

a. Model regresi yang menyatakan hubungan antara panjang bayi, umur bayi, dan bobot lahir bayi adalah $\hat{Y} = 18{,}667 + 0{,}395X_1 + 2{,}713X_2$ (lihat output **Coefficients**).
Interpretasi model:

β_0 = Saat umur bayi dan bobot lahir bayi bernilai 0, maka panjang bayi akan akan sebesar 18,667 cm.

β_1 = Setiap kenaikan 1 hari umur bayi, maka panjang bayi akan naik sebesar 0,395 cm dengan asumsi bobot lahir bayi tetap.

β_2 = Setiap kenaikan 1 kg bobot lahir bayi, maka panjang bayi akan naik sebesar 2,713 cm dengan asumsi umur bayi tetap.

b. Panjang rata-rata bayi yang berumur 75 hari dan beratnya waktu lahir 3,15 kg adalah
$\hat{Y} = 18{,}667 + 0{,}395\ (75) + 2{,}713\ (3{,}15)$ = 56,83795 cm

c. H_0 = Peubah bebas secara bersama-sama tidak berpengaruh terhadap peubah respon.
H_1 = Ada setidaknya 1 peubah bebas berpengaruh terhadap peubah respon.

Untuk menguji hipotesis ini, digunakan uji F. Dengan tingkat signifikan α =0,05, karena nilai Sig = 0,000 < α, maka H_0 ditolak (lihat output **ANOVA**). Hal ini berarti ada setidaknya peubah bebas (umur bayi dan bobot lahir bayi) berpengaruh terhadap panjang bayi.

d. Uji individu
- Peubah umur bayi

H_0: β_1 = 0 (Tidak ada pengaruh yang diberikan umur bayi terhadap panjang bayi).
H_1: $\beta_1 \neq 0$ (Ada pengaruh yang diberikan umur bayi terhadap panjang bayi).

Untuk menguji hipotesis ini digunakan uji t. Dengan tingkat signifikan α =0,05, karena nilai Sig = 0,001 < α, maka H_0 ditolak (lihat output **Coefficients**). Hal ini berarti ada pengaruh yang diberikan umur bayi terhadap panjang bayi.

- Peubah bobot lahir bayi

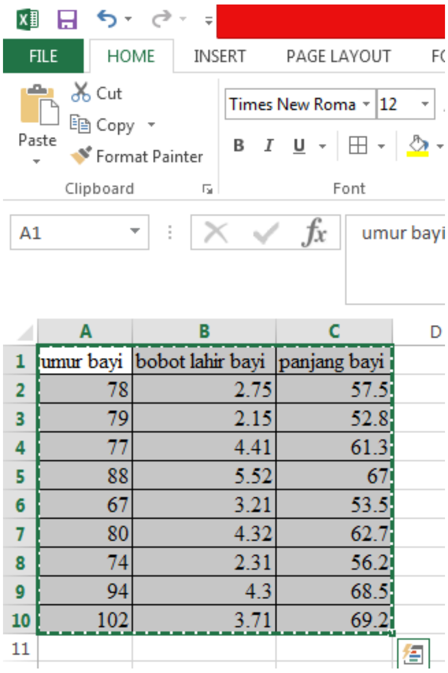
H_0: β_2 = 0 (Tidak ada pengaruh yang diberikan bobot lahir bayi terhadap panjang bayi).
H_1: $\beta_2 \neq 0$ (Ada pengaruh yang diberikan bobot lahir bayi terhadap panjang bayi).

Untuk menguji hipotesis ini digunakan uji t. Dengan tingkat signifikan α =0,05, karena nilai Sig = 0,003 < α, maka H_0 ditolak (lihat output **Coefficients**). Hal ini berarti ada pengaruh yang diberikan bobot lahir bayi terhadap panjang bayi.

e. Koefisien determinasi adalah R^2 = 0,948. Hal ini berarti keragaman data panjang bayi yang dapat dijelaskan oleh peubah bebas umur bayi dan bobot lahir bayi sebesar 94,8%, sedangkan sisanya disebabkan oleh peubah lain yang tidak dimasukkan dalam model regresi.

✓ Prosedur penyelesaian menggunakan R

1. Langkah pertama adalah dengan memanggil data dari format Excel yang berbentuk tabel. Sebelum pemanggilan data, terlebih dahulu untuk memblok semua data yang berbentuk tabel seperti pada gambar di bawah ini:



umur bayi	bobot lahir bayi	panjang bayi
78	2.75	57.5
79	2.15	52.8
77	4.41	61.3
88	5.52	67
67	3.21	53.5
80	4.32	62.7
74	2.31	56.2
94	4.3	68.5
102	3.71	69.2

Lalu panggil tabel tersebut ke dalam R dengan menggunakan sintaks

```
> read.delim("clipboard")
> data=read.delim("clipboard")
```

2. Setelah data dipanggil, maka dicari model dugaannya dengan menggunakan sintaks

```
>lm(dataY~DataX1+DataX2)
>model=lm(panjang.bayi~umur.bayi+bobot.lahir.bayi, data=data)
> model

Call:
lm(formula = panjang.bayi ~ umur.bayi + bobot.lahir.bayi, data = data)

Coefficients:
     (Intercept)          umur.bayi  bobot.lahir.bayi
         18.6669             0.3952            2.7130
```

Persamaan model dugaannya adalah $\hat{Y} = 18{,}6669 + 0{,}3952X_1 + 2{,}7130X_2$.

3. Lalu dicari nilai untuk uji F dan uji individual serta koefisien determinasinya dengan menggunakan sintaks

```
> summary(model)

Call:
lm(formula = panjang.bayi ~ umur.bayi + bobot.lahir.bayi, data = data)

Residuals:
```

```
    Min      1Q  Median      3Q     Max 
-2.9189 -0.3526  0.2400  0.6986  2.0229 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)       18.6669     4.5383   4.113 0.006264 ** 
umur.bayi          0.3952     0.0605   6.532 0.000615 ***
bobot.lahir.bayi   2.7130     0.5814   4.667 0.003443 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.666 on 6 degrees of freedom
Multiple R-squared:  0.9481,	Adjusted R-squared:  0.9308 
F-statistic: 54.84 on 2 and 6 DF,  p-value: 0.0001395
```

Hasil uji F dapat dilihat pada tabel untuk nilai F dan p_{value} dapat dilihat pada bagian **F-statistic** dan **p-value**. Untuk koefisien dari β_0 dan β_1 dapat dilihat pada bagian kolom **Estimate**. Hasil uji individual (uji t) untuk masing-masing koefisien dapat dilihat pada bagian kolom **Pr(>|t|)**. Koefisien determinasi dapat dilihat pada bagian **R-squared**.

Latihan:

Regresi Linear Berganda

1. Suhu plat pembungkus dan jarak plat pembungkus dalam mesin pembungkus sabun mempengaruhi persentase sabun terbungkus yang lolos inspeksi. Beberapa data mengenai tentang peubah itu telah dikumpulkan, yaitu sebagai berikut:

X_1	130	174	143	191	165	194	143	186	139	188	175	156	190	178	132	148
X_2	190	176	205	210	230	192	220	235	240	230	200	218	220	210	208	225
Y	35,0	81,7	42,5	98,3	52,7	82,0	34,5	95,4	56,7	84,4	94,3	44,3	83,3	91,4	43,5	51,7

   a. Dengan menggunakan model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$, tentukanlah nilai dugaan kuadrat terkecil bagi $\beta_0, \beta_1,$ dan β_2. Tentukan persamaan peramalannya, kemudian interpretasikan persamaan tersebut.
   b. Susunlah tabel analisis ragam dan uji parameter dugaan dengan taraf nyata α = 0,05
   c. Hitung koefisien determinasinya dan jelaskan artinya.

2. Delapan runtunan dicobakan pada berbagai kondisi kejenuhan (X_1) dan transisomer (X_2). Responnya, SCI dicantumkan dibawah ini sebagai Y.

X_1	38	41	34	35	31	34	29	32
X_2	47,5	21,3	36,5	18,0	29,5	14,2	21	10
Y	66	43	36	23	22	14	12	7,6

   a. Dengan menggunakan model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$, tentukanlah nilai dugaan kuadrat terkecil bagi $\beta_0, \beta_1,$ dan β_2. Tentukan persamaan peramalannya, kemudian interpretasikan persamaan tersebut.
   b. Susunlah tabel analisis ragam dan uji parameter dugaan dengan taraf nyata α = 0,05
   c. Hitung koefisien determinasinya dan jelaskan artinya.

Modul Praktikum VI

Judul : Regresi dengan Peubah Dummy

Tujuan : Dapat menyelesaikan masalah regresi dengan data peubah bebas.

Dasar Teori

Regresi dengan Peubah Dummy

Misalkan kita ingin memasukkan ke dalam suatu model gagasan bahwa ada dua jenis mesin (misalnya tipe A dan tipe B) yang menhasilkan taraf respon yang berbeda, selain keragaman yang diakibatkan oleh peubah-peubah lain. Salah satu cara yang dapat ditempuh adalah memasukkan satu peubah dummy Z dan satu koefisien regresi α sehingga dalam model muncul tambahan suku αZ. Koefisien α harus diduga bersamaan dengan koefisien-koefisien β. Kepada peubah Z dapat diberikan taraf sebagai berikut:

Z = 0 jika amatan berasal dari mesin A

Z = 1 jika amatan berasal dari mesin B

Sesungguhnya, kepada Z dapat diberikan sembarang dua nilai yang berbeda. Namun, biasanya di atas itulah yang terbaik. Secara umum, untuk r taraf diperlukan (r-1) peubah dummy, pola alokasinya diperoleh dengan menuliskan matriks I yang berukuran (r-1)X(r-1) dan menambahkan satu baris yang terdiri atas (r-1) nol.

Prosedur Kerja

Data

Contoh

Seorang peneliti tertarik untuk memprediksi laba 2 macam perusahaan (swasta asing dan swasta nasional) bila ditinjau dari besarnya biaya iklan yang dikeluarkan oleh perusahaan untuk membuat iklan mengenai produknya. (Untuk perusahaan swasta asing, laba yang diamati adalah laba yang diperoleh dari hasil penjualan produknya di wilayah Indonesia saja). Perusahaan swasta asing dikategorikan 1, dan lainnya 0.

No.	iklan	laba	tipe
1.	10	9.17	1
2.	1	1.32	0
3.	12	8.54	1
4.	12	7.68	1
5.	5	7.15	1
6.	4	2.54	0
7.	8	10.85	1
8.	4	2.39	0
9.	8	1.5	0
10.	8	5.13	0
11.	5	9.08	1
12.	14	8.77	1
13.	2	10.85	1
14.	2	1.49	0
15.	12	7.92	1
16.	9	5.87	0
17.	13	8.97	1
18.	9	7.07	1
19.	3	0.32	0
20.	3	1.84	0

1. Tentukan model persamaan regresi dan interpretasikan modelnya.
2. Lakukan pengujian parameter dan model dengan menggunakan $\alpha = 0{,}05$.

✓ Prosedur Minitab

Setelah penambahan peubah dummy, prosedur kerja akan sama dengan prosedur kerja regresi sederhana dan linear berganda. Untuk prosedur kerja kembali berpedoman kepada modul praktikkum II dan modul praktikkum V.

1. Masukkan data ke dalam Minitab.
2. Lalu klik **Stat>Regression>Regression.**

Output

```
The regression equation is
Laba = 2.16 + 0.070 Iklan + 5.92 Tipe
```

<u>Model</u>

- Model : Y = 2.16 + 0.070X + 5.92 Z

Setiap penambahan satu satuan biaya iklan yang dikeluarkan oleh perusahaan untuk membuat iklan mengenai produknya maka perusahaan akan mendapatkan penambahan laba sebesar 0.070.

- Model regresi untuk tipe perusahaan swasta asing :
  Y = 2.16 + 0.070X + 5.92 (1)
  = 8.08 + 0.070X

Setiap penambahan satu satuan biaya iklan yang dikeluarkan oleh perusahaan swasta asing untuk membuat iklan mengenai produknya maka perusahaan akan mendapatkan penambahan laba sebesar 5.92.

- Model regresi untuk tipe perusahaan swasta nasional :
  Y = 2.16 + 0.070X + 5.92 (0)
  = 2.16 + 0.070X

Penambahan satu satuan biaya iklan yang dikeluarkan oleh perusahaan swasta nasional untuk membuat iklan mengenai produknya tidak menambahkan laba bagi perusahaan tersebut.

Pengujian parameter (β_0, β_1, α)

Pengujian paremeter merupakan suatu uji yang dilakukan untuk mengetahui apakah ada pengaruh atau tidak suatu peubah X *(independen variable*) terhadap terhadap peubah Y (*dependen variable*).

Output :

Predictor	Coef	SE Coef	T	P
Constant	2.1629	0.7239	2.99	0.008
Iklan	0.0699	0.1067	0.65	0.521
Tipe	5.9212	0.8622	6.87	0.000

Interpretasi :

- Pengujian parameter (**β_0**) :
  Hipotesis :
  H_0 : Biaya iklan yang dikeluarkan oleh perusahaan swasta asing untuk membuat iklan mengenai produknya dan tipe perusahaan tidak mempengaruhi laba perusahaan.
  H_1 : Biaya iklan yang dikeluarkan oleh perusahaan swasta asing untuk membuat iklan mengenai produknya dan tipe perusahaan mempengaruhi laba perusahaan.

  Keputusan :

Predictor	Coef	SE Coef	T	P
Constant	2.1629	0.7239	2.99	0.008
Iklan	0.0699	0.1067	0.65	0.521
Tipe	5.9212	0.8622	6.87	0.000

  Pvalue = 0.008 < α = 0.05, jadi dapat diputuskan untuk menolak H_0.

Kesimpulan :

Karena tolak H_0 maka dapat disimpulkan bahwa biaya iklan yang dikeluarkan oleh perusahaan untuk membuat iklan mengenai produknya dan tipe perusahaan mempengaruhi laba perusahaan.

- Pengujian parameter (β_1) :

Hipotesis :

H_0 : Biaya iklan yang dikeluarkan oleh perusahaan untuk membuat iklan mengenai produknya tidak mempengaruhi laba perusahaan.

H_1 : Biaya iklan yang dikeluarkan oleh perusahaan swasta asing untuk membuat iklan mengenai produknya mempengaruhi laba perusahaan.

Keputusan :

Predictor	Coef	SE Coef	T	P
Constant	2.1629	0.7239	2.99	0.008
Iklan	0.0699	0.1067	0.65	0.521
Tipe	5.9212	0.8622	6.87	0.000

Pvalue = 0.521 > α = 0.05, jadi tidak ada cukup bukti untuk menolak H_0.

Kesimpulan :

Karena tidak ada cukup bukti untuk menolak H_0. maka dapat disimpulkan bahwa biaya iklan yang dikeluarkan oleh perusahaan untuk membuat iklan mengenai produknya tidak mempengaruhi laba perusahaan.

- Pengujian parameter (α) :

Hipotesis :

H_0 : Tipe perusahaan tidak mempengaruhi laba perusahaan.

H_1 : Tipe perusahaan mempengaruhi laba perusahaan.

Keputusan :

Predictor	Coef	SE Coef	T	P
Constant	2.1629	0.7239	2.99	0.008
Iklan	0.0699	0.1067	0.65	0.521
Tipe	5.9212	0.8622	6.87	0.000

Pvalue = 0.000 < α = 0.05, jadi dapat diputuskan untuk menolak H_0.

Kesimpulan :

Karena tolak H_0 maka dapat disimpulkan bahwa tipe perusahaan mempengaruhi laba perusahaan.

✓ Prosedur menggunakan R

Langkah pertama adalah dengan membangun model. Lalu mencari summary untuk semua model. Pembahasan dan interpretasi model serta pengujian sama halnya dengan prosedur Minitab.

```
> data
   Iklan  Laba Tipe
1     10  9.17    1
2      1  1.32    0
3     12  8.54    1
4     12  7.68    1
5      5  7.15    1
6      4  2.54    0
7      8 10.85    1
8      4  2.39    0
9      8  1.50    0
10     8  5.13    0
11     5  9.08    1
12    14  8.77    1
13     2 10.85    1
14     2  1.49    0
15    12  7.92    1
16     9  5.87    0
17    13  8.97    1
18     9  7.07    1
19     3  0.32    0
20     3  1.84    0

> model=lm(Laba~., data=data)
> model

Call:
lm(formula = Laba ~ ., data = data)

Coefficients:
(Intercept)        Iklan         Tipe
    2.16290      0.06986      5.92117

> summary(model)

Call:
lm(formula = Laba ~ ., data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-2.0525 -1.0572 -0.3372  0.4522  3.0784

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.16290    0.72388   2.988  0.00826 **
Iklan        0.06986    0.10671   0.655  0.52145
Tipe         5.92117    0.86223   6.867 2.73e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.576 on 17 degrees of freedom
Multiple R-squared:  0.8212,  Adjusted R-squared:  0.8002
F-statistic: 39.04 on 2 and 17 DF,  p-value: 4.414e-07
```

Latihan:

1. USGA ingin membandingkan jarak rata-rata penggerak dari 4 jenis bola golf yang berbeda (A, B, C, D). Iron Byron, robot pemain golf USGA digunakan untuk memukul sebuah sampel 10 bola dari setiap jenis. Data jarak dapat dilihat pada tabel di bawah ini.
   Tabel jarak penggerak (dalam kaki) untuk empat jenis bola kaki.

Jenis A	Jenis B	Jenis C	Jenis D
251,2	263,2	269,7	251,6
245,1	262,9	263,2	240,6
240,0	265,0	277,5	249,4
251,1	254,5	267,4	242,0
260,5	264,3	270,3	246,5
250,0	257,0	265,5	251,3
253,9	262,8	270,7	261,8
244,6	264,4	272,9	249,0
254,6	260,6	275,6	247,1
248,8	255,9	266,5	245,9

   a. Tentukan model persamaan regresi dan interpretasikan modelnya.
   b. Lakukan pengujian parameter dan model dengan menggunakan $\alpha = 0{,}05$.

2. Maneger personalia suatu perusahaan tekstil melakukan penelitian terhadap penghasilan rata-rata stiap bulan karyawannya. Diduga bahwa penghasilan rata-rata setiap bulan karyawan di perusahaan tersebut dipengaruhi oleh masa kerja dan tingkat pendidikan. Penghasilan rata-rata setiap bulan karyawan dalam dolar ($), masa kerja dalam satuan tahun dan tingkat pendidikan yang terdiri dari 3 kategori yaitu bukan sarjana, sarjana muda, dan sarjana penuh. Datanya sebagai berikut:

Penghasilan	Masa kerja	Tingkat Pendidikan
400	4	Sarjana penuh
200	3	Sarjana muda
600	8	Sarjana muda
450	5	Sarjana penuh
380	4	Sarjana muda
720	8	Sarjana muda
540	5	Bukan sarjana
430	3	Sarjana muda
640	8	Sarjana muda

320	4	Bukan sarjana

a. Tentukan model persamaan regresi dan interpretasikan modelnya.
b. Lakukan pengujian parameter dan model dengan menggunakan $\alpha = 0,05$.

Modul Praktikum VII

Judul : Tranformasi Data

Tujuan : Untuk mengatasi jika asumsi-asumsi dalam analisis regresi tidak terpenuhi.

Dasar Teori

Dalam analisis regresi ada beberapa asumsi yang harus dipenuhi oleh data yang tersedia. Terkadang pada pratiknya asumsi-asumsi tersebut tidak dapat dipenuhi. Jalan keluar yang dapat ditempuh untuk mengatasi hal tersebut antara lain dengan mengembangkan metode alternatif (seperti penggunaan statistik non parametrik) atau melakuan tranformasi data. Transformasi data merupakan suatu cara yang efektif untuk menghasilkan model yang lebih baik. Tranformasi data berguna untuk:

1. Mengatasi adanya korelasi antar galat
2. Mengatasi galat yang tidak menyebar normal
3. Mengatasi ragam galat yang tidak homogen
4. Mengubah model yang tak linear menjadi linear.

Terdapat beberapa jenis transformasi yang sering digunakan, yaitu:

1. Transformasi Akar (*square-root)*
2. Transformasi Logaritma
3. Transformasi Arcsin
4. Transformasi Kuadrat (*square*)
5. Transformasi pangkat Tiga (*cubic)*
6. Transformasi Kebalikan (*inverse*)
7. Transformasi Kebalikan Akar (*inverse-square root)*
8. Transformasi Kebalikan Kuadrat (*inverse-square)*
9. Tranformasi Kebalikan Pangkat Tiga (*inverse-cubic)*
10. Transformasi Balik Skor (*inverse-score)*

Prosedur Kerja

Data

Contoh:

Diketahui data tentang umur (X) dan kandungan plasma poliamina (Y) untuk 25 anak sehat sebagai berikut:

X	*Y*	*X*	*Y*
0	13.44	3	7.94
0	12.84	3	6.01
0	11.91	3	5.14
0	20.09	3	6.90
0	15.60	3	6.77
1	10.11	4	4.86
1	11.38	4	5.10
1	10.28	4	5.67
1	8.96	4	5.75
1	8.59	4	6.23
2	9.83		
2	9.00		
2	8.65		
2	7.85		
2	8.88		

Lakukan tranformasi untuk:

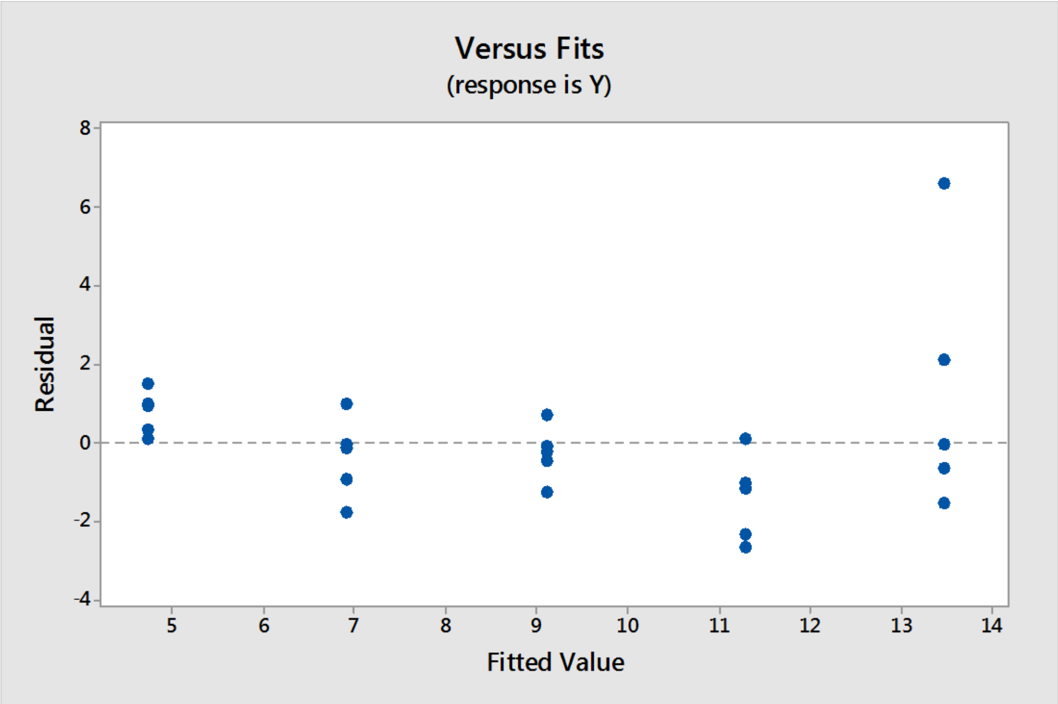
1. Mengatasi adanya korelasi antar galat
2. Mengatasi galat yang tidak menyebar normal
3. Mengatasi ragam galat yang tidak homogen

✓ Prosedur Minitab

1. Transformasi untuk mengatasi adanya korelasi antar galat.
   Pilih menu **Stat>Regression>Regression.** Pada **Graph**, klik **residual versus fits.** Kemudian pada **Option,** klik **durbin-Watson.**

Output

Plot Residual vs Fits

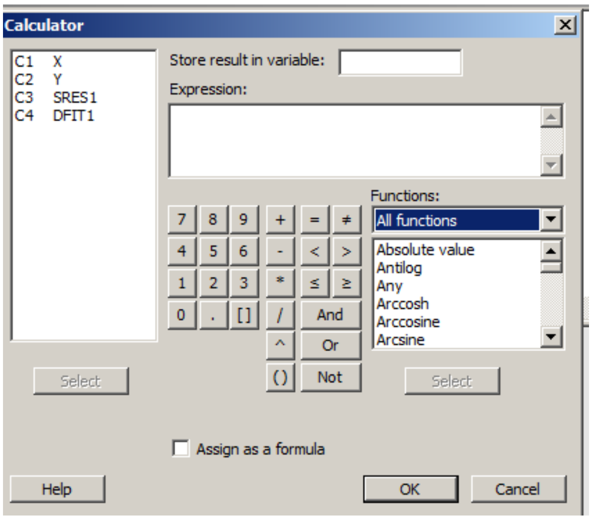


Uji Durbin-Watson

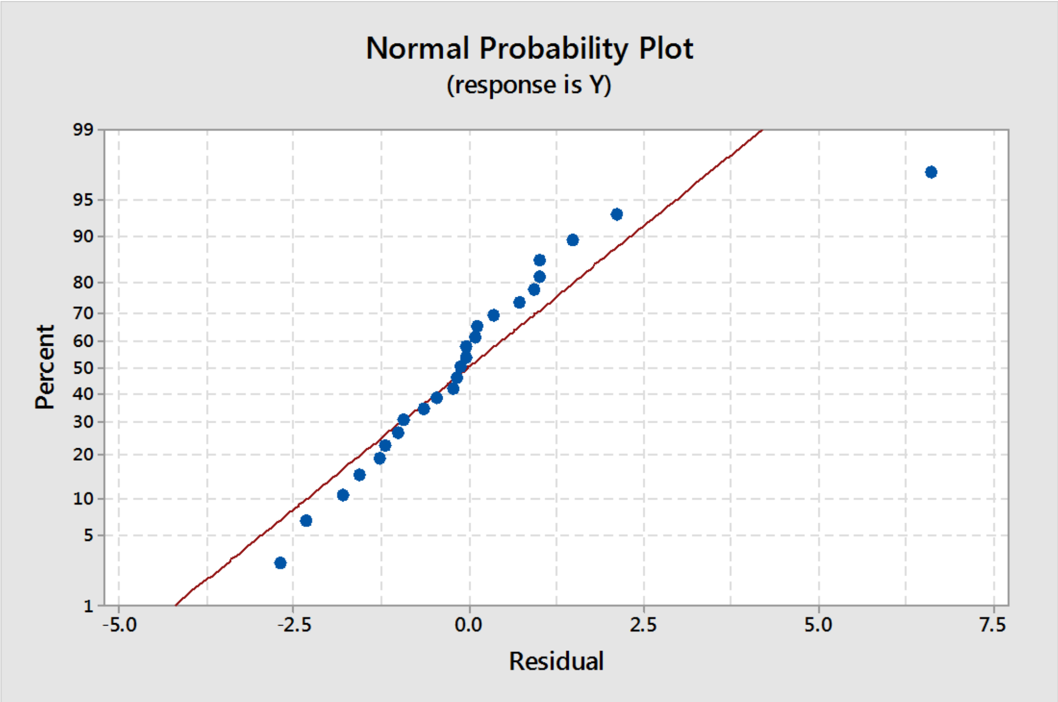
```
Durbin-Watson Statistic

Durbin-Watson Statistic =  1.64134
```

Berdasarkan plot di atas, dapat dilihat bahwa data tidak menyebar acak, yang artinya data tidak menyebar normal. Pada pengujian autokorelasi, juga terdapat autokorelasi antar galat, karena nilai Durbin-Watson berada antara 2,5463 dan 2,7121. Sehingga dapat disimpulkan bahwa asumsi data menyebar normal dan tidak terdapat autokorelasi tidak terpenuhi.

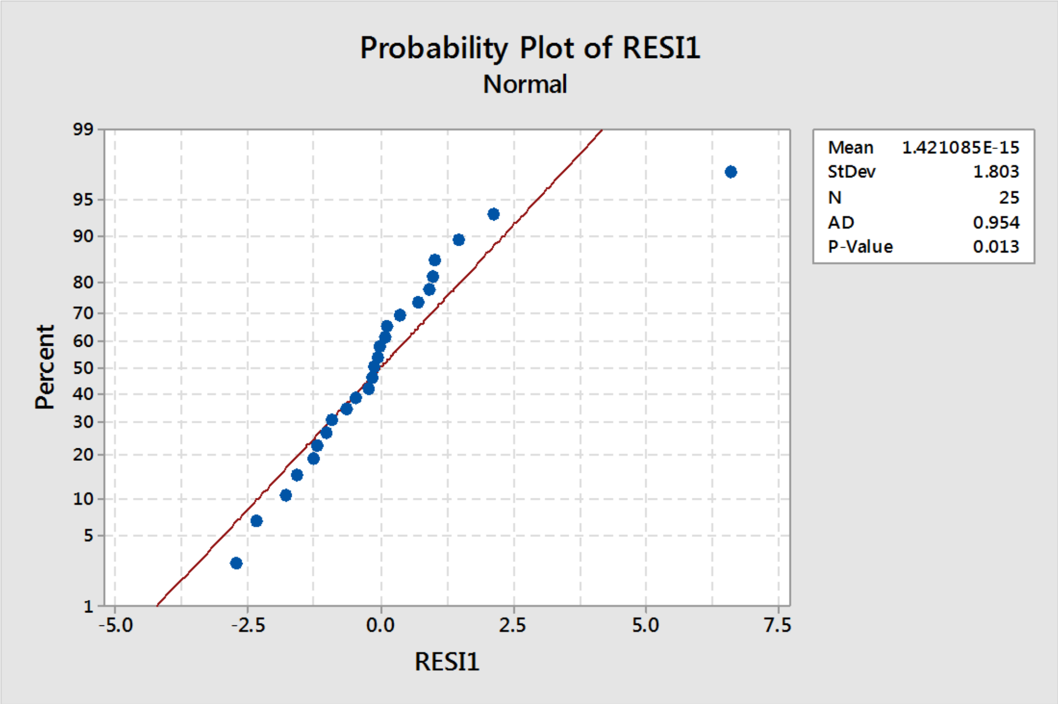
2. Transformasi untuk mengatasi keheterogenan ragam galat.
   Setelah terdeteksi ragam galat tidak homogen, maka dilakukan transformasi. Transformasi dapat dilakukan baik pada **Ms. Excel** maupun **Minitab**. Pada Minitab, pilih menu **Calc>Calculator**, kemudian pilih **function**.



3. Transformasi untuk mengatasu ketidaknormalan galat.
   Pilih menu **Stat>Regression>Regression>pada Graph klik normal plot of residuals.**



Jika hasil plot normal tidak meyakinkan, maka dilakukan uji secara formal. Uji kenormalan dapat dilakukan dengan uji Anderson-Darling. Pilih menu **Stat>Basic Statistics>normality test.** Pada kotak Normally test, pilih pengujian menggunakan uji Anderson Darling.



Dengan menggunakan hipotesis sebagai berikut:

H_0: Data mengikuti distribusi normal

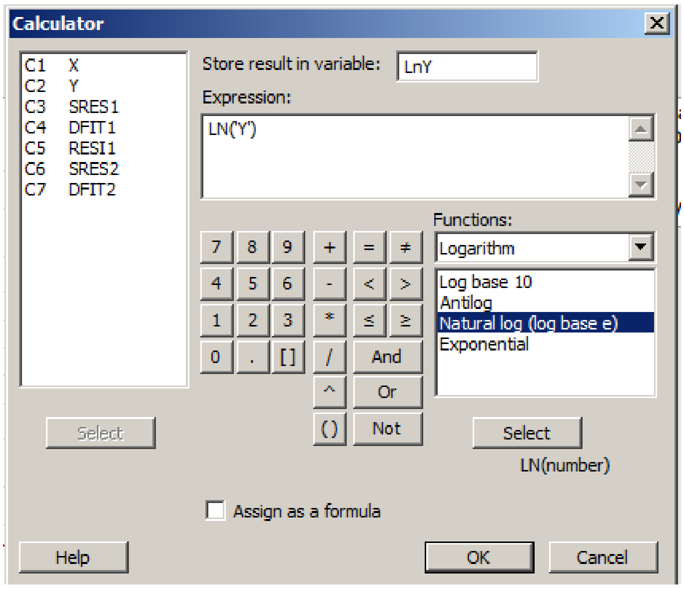
H_1: Data tidak mengikuti distribusi normal

Terlihat bahwa data tidak mengikuti distribusi normal dengan $\alpha = 0{,}05$, nilai *pvalue* lebih kecil dari nilai α. Sehingga asumsi kenormalan data tidak terpenuhi.

Transformasi Data

Transformasi dapat dilakukan dengan menggunakan **Minitab**.

1. Pilih **Calc>Calculator.**
2. Jika ingin dilakukan transformasi **ln,** maka pada bagian **functions**, pilih **Natural log (log base e).**



3. Lalu tranformasi untuk data Y, dan data X. setelah didapat ln dari data Y dan data X, maka diuji asumsi persamaan regresi yang didapat dari hasil transformasi. Jika memenuhi asumsi, maka data dapat menggunakan data yang telah ditransformasi dengan interpretasi hasil transformasi. Jika hasil transformasi juga tidak dapat memenuhi asumsi, maka dapat menggunakan metode statistik lainnya.

✓ Prosedur R

Pada bagian transformasi, pada R kita dapat lakukan dengan menggunaka sintak-sintak sebagai berikut:

1. Tranformasi akar
```
sqrt(data)
```
2. Transformasi kuadrat
```
data^2
```
3. Transformasi kubik
```
data^3
```
4. Transformasi eksponen
```
exp(data)
```
5. Transformasi arcsin
```
arcsin(data)
```
6. Dan lain sebagainya.

Untuk langkah lainnya, sama seperti pada modul sebelumnya.

Modul Praktikkum VIII

Judul : Regresi Bertatar (*Stepwise Regression*)

Tujuan : 1. Mengetahui tahapan dalam regresi bertatar

2. Mendapatkan model terbaik berdasarkan metode regresi bertatar

3. Mampu menginterpretsikan model

Dasar Teori

Regresi Bertatar

Regresi bertatar ini mungkin secara luas sangat berguna dalam teknik pemilihan peubah. Prosedur tersebut membentuk sebuah deret model regresi dengan menambahkan atau mengurangkan peubah-peubah pada rmasing-masing tahapan. Kriteria untuk menambahkan atau mengurangkan pada setiap tahap biasanya digambarkan dalam susunan sebuah pengujian F parsial, misal F_{masuk} nilai statistik F untuk menambah sebuah variabel pada model dan misalkan F_{luar} nilai dari statistik F untuk mengurangi sebuah peubah dari model. Kita harus mempunyai $F_{masuk} \geq F_{luar}$ dan biasanya $F_{masuk} = F_{keluar}$.

Langkah-langkah dalam Regresi Bertatar

- Pada tahap pertama, membentuk sebuah model dengan membentuk sebuah model dengan satu variable bebasnya mempunyai korelasi yang sangat kuat pengaruhnya dengan terhadap *y*. Ini juga akan menjadi variable yang menghasilkan statistik *F* yang sangat kuat. Misal *x* dipilih, maka pada tahap kedua sisa calon variable-peubah k-1 di uji dan statistik untuk variable tersebut, yaitu :

$$F_j = \frac{SS_R(\beta_j|\beta_1, \beta_0)}{MS_E(x_j, x_1)}$$

  adalah maksimum, maka ditambahkan pada persamaan, memberikan $F_j > F_{masuk}$. MS_E menyatakan rata-rata error kuadrat untuk model yang berisikan x_1 dan x_j.

- Misalkan prosedur ini sekarang menunjukkan x_2 akan ditambahkan pada model tersebut, sekarang regresi bertahap menentukan apakah x_1 ditambahkan pada tahap pertama yang akan dihilangkan. Ini dengan mengerjakan perhitungan statistik *F*

$$F_1 = \frac{SS_R(\beta_1|\beta_2, \beta_0)}{MS_E(x_1, x_2)}$$

Jika $F_j < F_{luar}$, maka variabel x_1 dihilangkan.

- Umumnya pada setiap tingkatan himpunan calon variable-peubah sisa adalah diuji dan variable tersebut dengan statistic F terbesar dimasukkan yang memberikan bahwa nilai F yang diamati melebihi F_{masuk}. Maka statistic F parsial untuk setiap variable dalam model tersebut dihitung dan variable dengan nilai F terkecil dihilangkan jika $F_j < F_{luar}$. Prosedur tersebut terus berjalan hingga tidak ada lagi variable yang ditambahkan atau dikurangkan dari model tersebut.

Prosedur Kerja

Data

Suatu survey dilakukan terhadap 17 rumah sakit di sekitar Banda Aceh. Peubah-peubah yang diamati dalarn survey tersebut adalah :

X_1 = banyaknya pasien rata-rata per hari

X_2 = banyaknya pelayanan sinar-X per hari

X_3 = tempat tidur yang terisi per bulan

X_4 = banyaknya penduduk disekitarnya yang mungkin memerlukan fasilitas

X_5 = rata-rata lamanya pasien dirawat (opname) dalam hari

Y = banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut.

Secara lengkap data hasil survey tersebut disajikan sebagai berikut :

X_1	X_2	X_3	X_4	X_5	Y
15.6	2463	472.9	4.5	18	566.5
44	2048	1339.7	6.9	9.5	696.8
20.4	3940	620.2	4.3	4.3	1033.2
18.7	6505	568.3	3.9	36.2	1603.6
49.2	5723	1497.6	5.5	35.2	1611.4
44.9	11520	1365.8	4.6	24	1613.3

45.5	5779	1687	5.6	43.3	1854.2
59.3	5969	1639.9	5.2	46.7	2160.6
94.4	8461	2873.2	6.2	78.7	2305.9
182	21106	366.1	6.2	180.5	3503.9
96	13313	2912	5.9	60.9	3571.9
131.4	10771	3921	4.9	103.7	3741.4
127.2	15543	3865.7	5.5	126.8	4026.5
252.9	36194	7684.1	7	157.7	10343.8
409.2	34703	12446.3	10.8	169.4	11732.2
463.7	39204	14098.4	7.1	331.4	15414.9
510.2	86533	15524	6.4	371.6	18845.4

A. Tanpa melihat apakah asumsi-asumsi regresidi penuhi, tentukan:
   1. Persamaan Regresi linear berganda untuk data diatas, serta interpetasikan parameter-parameternya
   2. Uji model yang didapatkan (Uji serempak dan Parsial)

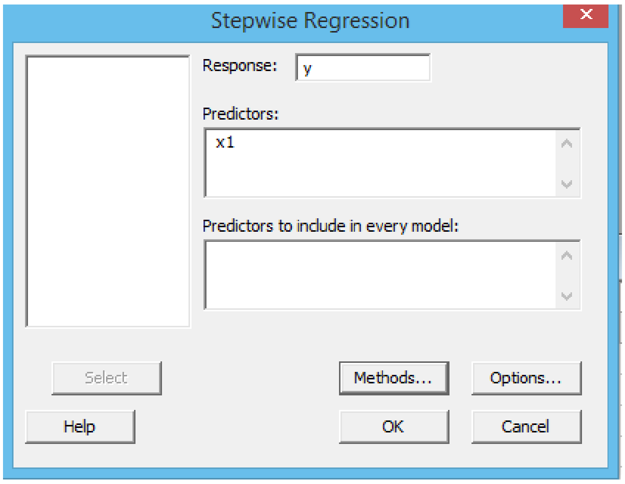
B. Apakah terjadi multikolinearitas pada data tersebut? Jika iya, lakukan *stepwise regression* untuk menghilangkan multikolinearitas pada data tersebut. (interpretasikan hasilnya).

Penyelesaian

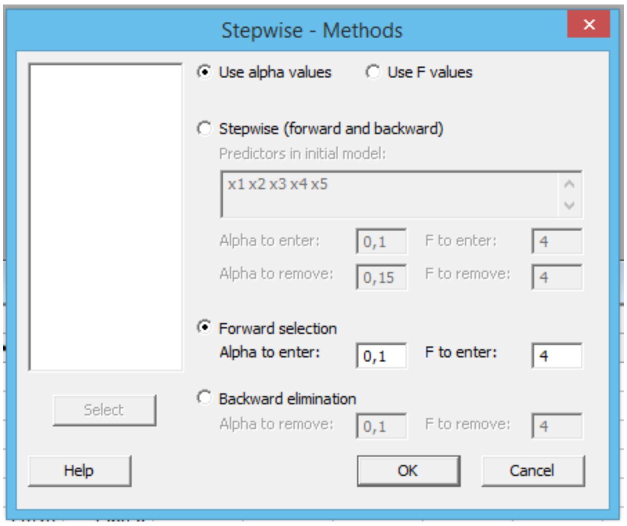
Berikut adalah langkah-langkah penyelesaian dengan menggunakan minitab untuk metode regresi bertatar

a. Forward Selection

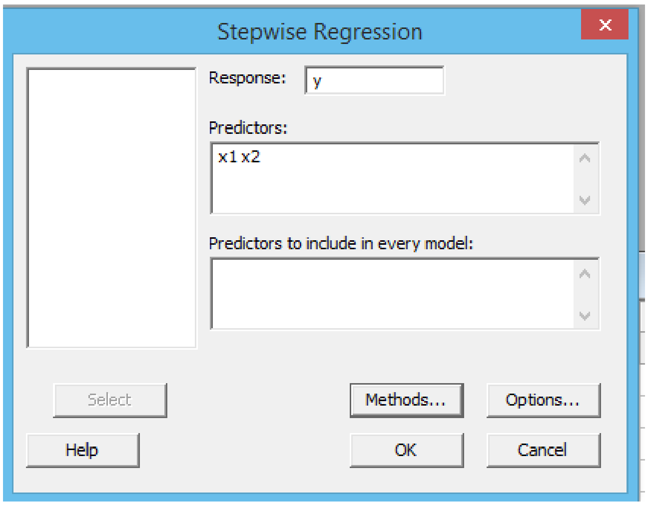
1. Pilih menu Stat>Regression>Stepwise Regression. Lalu pada kotak predictors, masukkan variabel x1 dan pada bagian Response masukkan variabel y.



2. Klik menu Methods. Pilih metode yang digunakan Forward Selections.



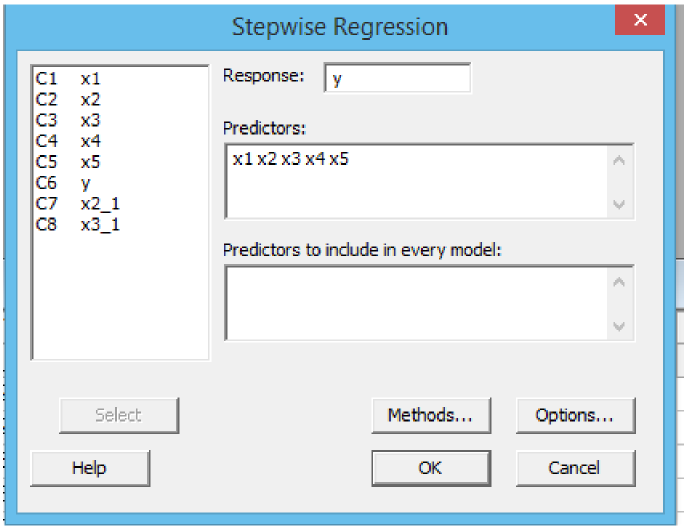
3. Lalu Klik Ok. Klik Ok.
4. Tambahkan variabel x2 pada kotak Predictors pada kotak dialog Stepwise Regression sebelumnya.



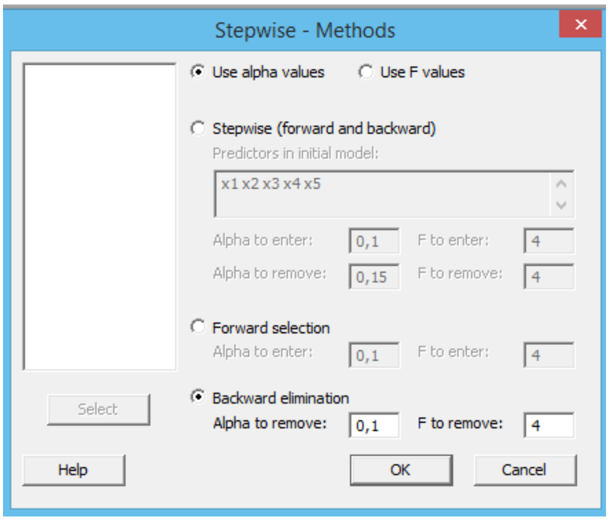
5. Ulangi langkah 2 dan 3. Lakukan langkah selanjutnya seperti langkah 1-5 sehingga kelima variabel bebas dimasukkan.

b. Backward Selection

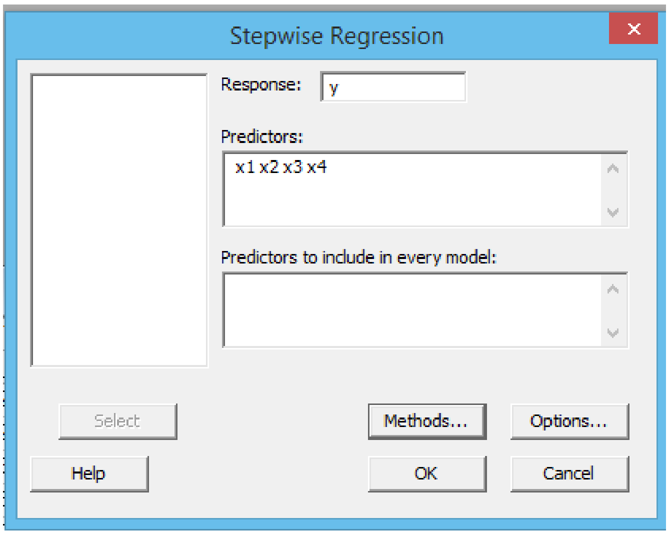
1. Pilih menu Stat>Regression>Stepwise Regression. Lalu pada kotak predictors, masukkan variabel x1, x2, x3, x4, x5 dan pada bagian Response masukkan variabel y.



2. Klik menu Methods. Pilih metode yang digunakan Backward Selections.



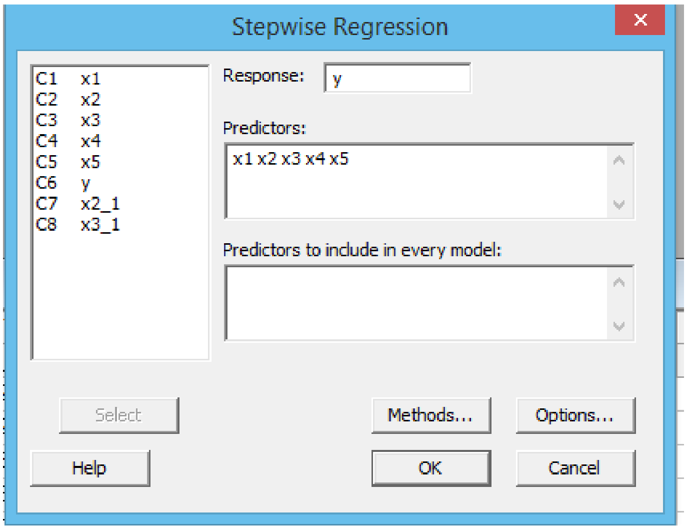
3. Lalu Klik Ok. Klik Ok.
4. Hilangkan variabel x5 pada kotak Predictors pada kotak dialog Stepwise Regression sebelumnya.



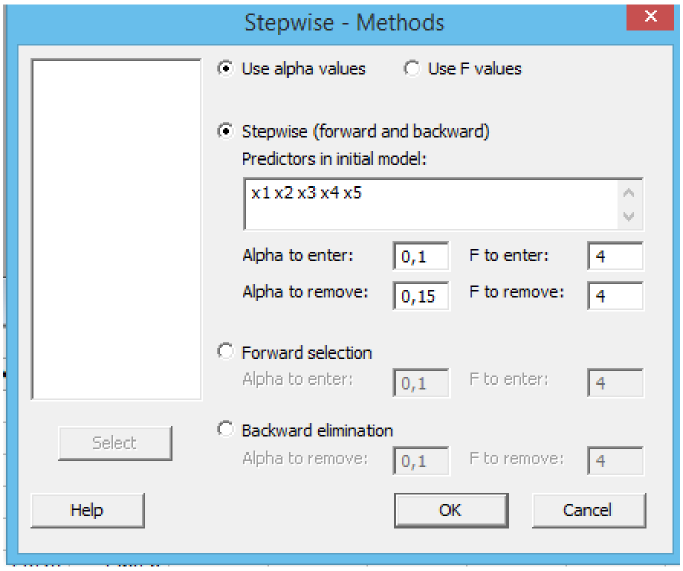
5. Ulangi langkah 2 dan 3. Lakukan langkah selanjutnya seperti langkah 1-5 sehingga hanya tinggal variabel x1 yang tersisa.

c. Stepwise Selection

1. Pilih menu Stat>Regression>Stepwise Regression. Lalu pada kotak predictors, masukkan variabel x1, x2, x3, x4, x5 dan pada bagian Response masukkan variabel y.



6. Klik menu Methods. Pilih metode yang digunakan Stepwise. Masukkan kelima variabel bebas pada kota Predictors in initial model.



7. Lalu Klik Ok. Klik Ok.

1. Persamaan regresi linear berganda :

Persamaan regresi adalah bentuk hubungan antara variabel yang akan tak bebas (*dependent variable*) dengan variabel bebas (*independent variable*).

```
The regression equation is
y = 1302 + 12.5 x1 + 0.0592 x2 + 0.514 x3 - 265 x4 + 0.63 x5
```

$Y = 1302 + 12.5 X_1 + 0.0592 X_2 + 0.514 X_3 - 265 X_4 + 0.63 X_5$

Interpretasi :

- Untuk setiap pertambahan banyaknya pasien rata-rata per hari (X_1) maka banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y) akan bertambah sebesar 12.5 jam dengan asumsi asumsi faktor lain dianggap konstan.
- Untuk setiap pertambahan banyaknya pelayanan sinar-X per hari (X_2) maka banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y) akan bertambah sebesar 0.0592 jam dengan asumsi asumsi faktor lain dianggap konstan.
- Untuk setiap pertambahan tempat tidur yang terisi per bulan (X_3) maka banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y) akan bertambah sebesar 0.514 jam dengan asumsi asumsi faktor lain dianggap konstan.
- Untuk setiap pertambahan banyaknya penduduk disekitarnya yang mungkin memerlukan fasilitas (X_4) maka banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y) akan berkurang sebesar 265 jam dengan asumsi asumsi faktor lain dianggap konstan.
- Untuk setiap pertambahan rata-rata lamanya pasien dirawat (opname) dalam hari (X_5) maka banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y) akan bertambah sebesar 0.63 jam dengan asumsi asumsi faktor lain dianggap konstan.

Dengan R^2 sebesar 99.1% dapat disimpulkan bahwa model sudah bagus dan mampu menjelaskan 99.1% dari total keragaman data sedangkan sisanya dipengaruhi oleh faktor lain.

2. a. Uji Serempak (Uji F)

Uji F dikenal dengan Uji serentak atau uji Model/Uji Anova, yaitu uji untuk melihat bagaimanakah pengaruh semua variabel bebasnya secara bersama-sama terhadap variabel terikatnya.. Atau untuk menguji apakah model regresi yang kita buat baik/signifikan atau tidak baik/non signifikan

Predictor	Coef	SE Coef	T	P	VIF
Constant	1302	1323	0.98	0.346	
x1	12.55	11.32	1.11	0.291	133.693
x2	0.05920	0.02119	2.79	0.017	8.136
x3	0.5144	0.1616	3.18	0.009	26.173
x4	-265.1	256.1	-1.04	0.323	6.591
x5	0.634	8.893	0.07	0.944	37.222

Interpretasi :

- Hipotesis :

  H_0 : minimal ada satu variabel X yang tidak mempengaruhi variabel Y.

  H_1 : semua variabel X secara serempak mempengaruhi variabel Y.

- Keputusan :

  Pada output diatas dapat dilihat bahwa P*value* untuk *constant* adalah 0.346.

  P*value* = 0.346 > α = 0.05 , jadi tidak cukup bukti untuk menolak H_0.

- Kesimpulan :

  Karena tidak cukup bukti untuk menolak H_0, maka dapat disimpulkan bahwa semua variabel X secara serempak mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y).

b. Uji parsial :

Uji t dikenal dengan uji parsial, yaitu untuk menguji bagaimana pengaruh masing-masing variabel bebasnya secara sendiri-sendiri terhadap variabel terikatnya. Uji ini dapat dilakukan dengan mambandingkan t hitung dengan t tabel atau dengan melihat kolom signifikansi pada masing-masing t hitung.

Predictor	Coef	SE Coef	T	P	VIF
Constant	1302	1323	0.98	0.346	
x1	12.55	11.32	1.11	0.291	133.693
x2	0.05920	0.02119	2.79	0.017	8.136
x3	0.5144	0.1616	3.18	0.009	26.173
x4	-265.1	256.1	-1.04	0.323	6.591
x5	0.634	8.893	0.07	0.944	37.222

Interpretasi :

➢ Untuk banyaknya pasien rata-rata per hari (X_1)

- Hipotesis :

  H_0 : banyaknya pasien rata-rata per hari (X_1) tidak mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y).

  H_1 : banyaknya pasien rata-rata per hari (X_1) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y).

- Keputusan :

Pada output diatas dapat dilihat bahwa P*value* untuk x1 adalah 0.291.
P*value* = 0.291 > α = 0.05 , jadi tidak cukup bukti untuk menolak H_0.

- Kesimpulan :
  Karena tidak cukup bukti untuk menolak H_0, maka dapat disimpulkan bahwa banyaknya pasien rata-rata per hari (X_1) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y).

➢ Untuk banyaknya pelayanan sinar-X per hari (X_2)

- Hipotesis :
  H_0 : banyaknya pelayanan sinar-X per hari (X_2) tidak mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y).
  H_1 : banyaknya pelayanan sinar-X per hari (X_2) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y).

- Keputusan :
  Pada output diatas dapat dilihat bahwa P*value* untuk x2 adalah 0.017.
  P*value* = 0.017 < α = 0.05 , jadi dapat diputuskan untuk menolak H_0.
- Kesimpulan :
  Karena diputuskan untuk menolak H_0, maka dapat disimpulkan bahwa banyaknya pelayanan sinar-X per hari (X_2) tidak mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y).

➢ Untuk tempat tidur yang terisi per bulan (X_3)

- Hipotesis :
  H_0 : tempat tidur yang terisi per bulan (X_3) tidak mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y).
  H_1 : tempat tidur yang terisi per bulan (X_3) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y).

- Keputusan :
  Pada output diatas dapat dilihat bahwa P*value* untuk x3 adalah 0.009.
  P*value* = 0.009 < α = 0.05 , jadi dapat diputuskan untuk menolak H_0.

- Kesimpulan :
  Karena diputuskan untuk menolak H_0, maka dapat disimpulkan bahwa tempat tidur yang terisi per bulan (X_3) tidak mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y).

➢ Untuk banyaknya penduduk disekitarnya yang mungkin memerlukan fasilitas(X_4)

- Hipotesis :

H_0 : banyaknya penduduk disekitarnya yang mungkin memerlukan fasilitas(X_4) tidak mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y).

H_1 : banyaknya penduduk disekitarnya yang mungkin memerlukan fasilitas(X_4) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y).

- Keputusan :
  Pada output diatas dapat dilihat bahwa P*value* untuk x4 adalah 0.323.
  P*value* = 0.323 > α = 0.05 , jadi tidak cukup bukti untuk menolak H_0.

- Kesimpulan :
  Karena tidak cukup bukti untuk menolak H_0, maka dapat disimpulkan bahwa banyaknya penduduk disekitarnya yang mungkin memerlukan fasilitas(X_4) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y).

➢ Untuk rata-rata lamanya pasien dirawat (opname) dalam hari (X_5)

- Hipotesis :
  H_0 : rata-rata lamanya pasien dirawat (opname) dalam hari (X_5) tidak mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y).

  H_1 : rata-rata lamanya pasien dirawat (opname) dalam hari (X_5) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y).

- Keputusan :
  Pada output diatas dapat dilihat bahwa P*value* untuk x5 adalah 0.944.

  P*value* = 0.944 > α = 0.05 , jadi tidak cukup bukti untuk menolak H_0.

- Kesimpulan :
  Karena tidak cukup bukti untuk menolak H_0, maka dapat disimpulkan bahwa rata-rata lamanya pasien dirawat (opname) dalam hari (X_5) tidak mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y).

A. Multikolinearitas dan *stepwise regression*.

1. Multikolinearitas

Multikolinearitas atau Kolinearitas Ganda adalah adanya hubungan linear antara peubah bebas X dalam Model Regresi Ganda. Jika nilai VIF ≥ 10, maka dapat disimpulkan bahwa pada data tersebut terdapat multikolinaritas.

Predictor	Coef	SE Coef	T	P	VIF
Constant	1302	1323	0.98	0.346	
x1	12.55	11.32	1.11	0.291	133.693
x2	0.05920	0.02119	2.79	0.017	8.136
x3	0.5144	0.1616	3.18	0.009	26.173
x4	-265.1	256.1	-1.04	0.323	6.591
x5	0.634	8.893	0.07	0.944	37.222

Interpretasi :

- Dari output diatas dapat dilihat bahwa x1, x3 dan x5 memiliki nilai VIF > 10 jadi dapat disimpulkan bahwa pada banyaknya pasien rata-rata per hari (X_1), tempat tidur yang terisi per bulan (X_3) dan rata-rata lamanya pasien dirawat (opname) dalam hari (X_5) terdapat multikolinearitas. Karena adanya multikolinearitas maka dilakukan analsis regresi bertatar (*stepwise regression*).

2. **Regresi bertatar (*stepwise regression*)**

Metode Stepwise, Backward Elimination dan Forward Selection merupakan suatu metode untuk mengurangi kemungkinan adanya muktikolinearitas dari persamaan/model yang dihasilkan. *Regresi stepwise* melibatkan dua jenis proses yaitu: forward selection dan backward elimination. Teknik ini dilakukan melalui beberapa tahapan. Pada masing-masing tahapan, kita akan memutuskan variabel mana yang merupakan prediktor terbaik untuk dimasukkan ke dalam model. Variabel ditentukan berdasarkan uji-F, variabel ditambahkan ke dalam model selama nilai p-valuenya kurang dari nilai kritik α (biasanya 0,15). Kemudian variabel dengan nilai p-value lebih dari nilai kritik α akan dihilangkan. Proses ini dilakukan terus menerus hingga tidak ada lagi variabel yang memenuhi kriteria untuk ditambahkan atau dihilangkan.

a. Metode stepwise (Forward and Backward)

Stepwise Regression: y versus x1, x2, x3, x4, x5

Alpha-to-Enter: 0.1 Alpha-to-Remove: 0.15

Response is y on 5 predictors, with N = 17

Step	1	2	3	4
Constant	-106.058	-118.422	1.989	1375.578
x1	33.7	24.0	9.1	13.3
T-Value	18.72	6.97	1.99	2.76
P-Value	0.000	0.000	0.068	0.017
x2		0.081	0.079	0.059
T-Value		3.10	4.26	2.92
P-Value		0.008	0.001	0.013

```
x3                                 0.51       0.51
T-Value                            3.87       4.20
P-Value                           0.002      0.001

x4                                            -279
T-Value                                      -1.81
P-Value                                      0.096

S                 1163       927    656        606
R-Sq             95.90     97.56  98.87      99.11
R-Sq(adj)        95.62     97.22  98.61      98.81
Mallows Cp        37.7      19.1    5.0        4.0
```

Interpretasi :

- **Step 1**
  Pada step 1 hanya ada satu variable yaitu banyaknya pasien rata-rata per hari (X_1) yang memiliki nilai Pvaleu = 0.000 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pasien rata-rata per hari (X_1) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y) dengan R^2 = 95.90 yang berarti bahwa model sudah bagus dan mampu menjlaskan 95.90% dari total keragaman data sedangkan sisanya dipengaruhi oleh faktor lain.

- **Step 2**

Pada step 2 ada dua variable yaitu banyaknya pasien rata-rata per hari (X_1) dan banyaknya pelayanan sinar-X per hari (X_2) yang memiliki nilai :

- banyaknya pasien rata-rata per hari (X_1) memiliki Pvaleu = 0.000 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pasien rata-rata per hari (X_1) mempengaruhi Y
- banyaknya pelayanan sinar-X per hari (X_2) memiliki Pvaleu = 0.000 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pelayanan sinar-X per hari (X_2) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)

dengan R^2 = 97.56yang berarti bahwa model sudah bagus dan mampu menjlaskan 97.56% dari total keragaman data sedangkan sisanya dipengaruhi oleh faktor lain

- **Step 3**

Pada step 3 ada tiga variable yaitu banyaknya pasien rata-rata per hari (X_1), x2 dan x3 yang memiliki nilai :

- banyaknya pasien rata-rata per hari (X_1) memiliki Pvaleu = 0.068 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pasien rata-rata per hari (X_1) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)

- banyaknya pelayanan sinar-X per hari (X_2) memiliki Pvaleu = 0.001 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pelayanan sinar-X per hari (X_2) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)
- tempat tidur yang terisi per bulan (X_3) memiliki Pvaleu = 0.002 < α = 0.1, sehingga dapat disimpulkan bahwa tempat tidur yang terisi per bulan (X_3) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)

Dengan R^2 = 98.87 yang berarti bahwa model sudah bagus dan mampu menjlaskan 98.87% dari total keragaman data sedangkan sisanya dipengaruhi oleh faktor lain.

- **Step 4**

Pada step 4 ada empat variable yaitu banyaknya pasien rata-rata per hari (X_1), x2 , x3 dan x4 yang memiliki nilai :

- banyaknya pasien rata-rata per hari (X_1) memiliki Pvaleu = 0.017 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pasien rata-rata per hari (X_1) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)
- banyaknya pelayanan sinar-X per hari (X_2) memiliki Pvaleu = 0.013 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pelayanan sinar-X per hari (X_2) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)
- tempat tidur yang terisi per bulan (X_3) memiliki Pvaleu = 0.001 < α = 0.1, sehingga dapat disimpulkan bahwa tempat tidur yang terisi per bulan (X_3) mempengaruhi Y banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)
- banyaknya penduduk disekitarnya yang mungkin memerlukan fasilitas(X_4) memiliki Pvaleu = 0.001 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya penduduk disekitarnya yang mungkin memerlukan fasilitas(X_4) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)

Dengan R^2 = 99.11 yang berarti bahwa model sudah bagus dan mampu menjlaskan 99.11% dari total keragaman data sedangkan sisanya dipengaruhi oleh faktor lain.

b. forward stepwise

- **1 variabel**

```
Stepwise Regression: y versus x1

Forward selection.  Alpha-to-Enter: 0.1


Response is y on 1 predictors, with N = 17


Step              1
Constant     -106.1

x1             33.7
T-Value       18.72
P-Value       0.000

S              1163
R-Sq          95.90
R-Sq(adj)     95.62
Mallows Cp      2.0
```

Interpretasi :

Dari output diatas dapat dilihat bahwa ketika hanya dimasukkan banyaknya pasien rata-rata per hari (X_1) pada step 1 P*value* berniali 0.000 yang lebih kecil dari *alpha to enter* sebesar 0.1, berarti bahwa banyaknya pasien rata-rata per hari (X_1) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y). dengan R^2 = 95.90 yang berarti bahwa model sudah bagus dan mampu emnjelasskan 95.90% dari total keragaman data sedangkan sisanya dipengaruhi oleh faktor lain.

- **2 variabel :**

```
Stepwise Regression: y versus x1, x2
Forward selection.  Alpha-to-Enter: 0.1
Response is y on 2 predictors, with N = 17

Step              1        2
Constant     -106.1   -118.4

x1             33.7     24.0
T-Value       18.72     6.97
P-Value       0.000    0.000

x2                     0.081
T-Value                 3.10
P-Value                0.008

S              1163      927
R-Sq          95.90    97.56
R-Sq(adj)     95.62    97.22
Mallows Cp     10.6      3.0
```

Interpretasi :

Ketika di amsukkan 2 variabel x1 dan x2 maka terdapat 2 step sebagai berikut :

- Step 1
  Pada step 1 hanya ada satu variable yaitu banyaknya pasien rata-rata per hari (X_1) yang memiliki nilai Pvaleu = 0.000 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pasien rata-rata per hari (X_1) mempengaruhi Y dengan R^2 = 95.90 yang berarti bahwa model

sudah bagus dan mampu menjlaskan 95.90% dari total keragaman data sedangkan sisanya dipengaruhi oleh faktor lain.

- Step 2

Pada step 2 ada dua variable yaitu banyaknya pasien rata-rata per hari (X_1) dan banyaknya pelayanan sinar-X per hari (X_2) yang memiliki nilai :

- banyaknya pasien rata-rata per hari (X_1) memiliki Pvaleu = 0.000 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pasien rata-rata per hari (X_1) mempengaruhi Y
- banyaknya pelayanan sinar-X per hari (X_2) memiliki Pvaleu = 0.000 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pelayanan sinar-X per hari (X_2) mempengaruhi Y

dengan R^2 = 97.56yang berarti bahwa model sudah bagus dan mampu menjlaskan 97.56% dari total keragaman data sedangkan sisanya dipengaruhi oleh faktor lain

- **3 variabel**

Stepwise Regression: y versus x1, x2, x3

```
Forward selection.  Alpha-to-Enter: 0.1
Response is y on 3 predictors, with N = 17

Step              1        2       3
Constant   -106.058 -118.422   1.989

x1             33.7     24.0     9.1
T-Value       18.72     6.97    1.99
P-Value       0.000    0.000   0.068

x2                     0.081   0.079
T-Value                 3.10    4.26
P-Value                0.008   0.001

x3                              0.51
T-Value                         3.87
P-Value                        0.002
```

Interpretasi :

Ketika di amsukkan 3 variabel x1, x2 dan x3 maka terdapat 3 step sebagai berikut :

- Step 1

  Pada step 1 hanya ada satu variable yaitu banyaknya pasien rata-rata per hari (X_1) yang memiliki nilai Pvaleu = 0.000 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pasien rata-rata per hari (X_1) mempengaruhi Y dengan R^2 = 95.90 yang berarti bahwa model sudah bagus dan mampu menjlaskan 95.90% dari total keragaman data sedangkan sisanya dipengaruhi oleh faktor lain.

- Step 2

Pada step 2 ada dua variable yaitu banyaknya pasien rata-rata per hari (X_1) dan banyaknya pelayanan sinar-X per hari (X_2) yang memiliki nilai :

- banyaknya pasien rata-rata per hari (X_1) memiliki Pvaleu = 0.000 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pasien rata-rata per hari (X_1) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)
- banyaknya pelayanan sinar-X per hari (X_2) memiliki Pvaleu = 0.000 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pelayanan sinar-X per hari (X_2) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)

dengan R^2 = 97.56yang berarti bahwa model sudah bagus dan mampu menjlaskan 97.56% dari total keragaman data sedangkan sisanya dipengaruhi oleh faktor lain

- Step 3

Pada step 3 ada tiga variable yaitu banyaknya pasien rata-rata per hari (X_1), x2 dan x3 yang memiliki nilai :

- banyaknya pasien rata-rata per hari (X_1) memiliki Pvaleu = 0.068 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pasien rata-rata per hari (X_1) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)
- banyaknya pelayanan sinar-X per hari (X_2) memiliki Pvaleu = 0.001 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pelayanan sinar-X per hari (X_2) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)
- tempat tidur yang terisi per bulan (X_3) memiliki Pvaleu = 0.002 < α = 0.1, sehingga dapat disimpulkan bahwa tempat tidur yang terisi per bulan (X_3) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)

Dengan R^2 = 98.87 yang berarti bahwa model sudah bagus dan mampu menjlaskan 98.87% dari total keragaman data sedangkan sisanya dipengaruhi oleh faktor lain

- **4 variabel**

Stepwise Regression: y versus x1, x2, x3, x4

Forward selection. Alpha-to-Enter: 0.1

Response is y on 4 predictors, with N = 17

Step	1	2	3	4
Constant	-106.058	-118.422	1.989	1375.578
x1	33.7	24.0	9.1	13.3
T-Value	18.72	6.97	1.99	2.76
P-Value	0.000	0.000	0.068	0.017
x2		0.081	0.079	0.059
T-Value		3.10	4.26	2.92
P-Value		0.008	0.001	0.013

x3			0.51	0.51
T-Value			3.87	4.20
P-Value			0.002	0.001
x4				-279
T-Value				-1.81
P-Value				0.096
S	1163	927	656	606
R-Sq	95.90	97.56	98.87	99.11
R-Sq(adj)	95.62	97.22	98.61	98.81
Mallows Cp	42.3	21.8	6.3	5.0

Interpretasi

Ketika di amsukkan 4 variabel x1, x2, x3 dan x4 maka terdapat 4 step sebagai berikut:

- Step 1

  Pada step 1 hanya ada satu variable yaitu banyaknya pasien rata-rata per hari (X_1) yang memiliki nilai Pvaleu = 0.000 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pasien rata-rata per hari (X_1) mempengaruhi Y dengan R^2 = 95.90 yang berarti bahwa model sudah bagus dan mampu menjlaskan 95.90% dari total keragaman data sedangkan sisanya dipengaruhi oleh faktor lain.

- Step 2

Pada step 2 ada dua variable yaitu banyaknya pasien rata-rata per hari (X_1) dan banyaknya pelayanan sinar-X per hari (X_2) yang memiliki nilai :

- banyaknya pasien rata-rata per hari (X_1) memiliki Pvaleu = 0.000 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pasien rata-rata per hari (X_1) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)
- banyaknya pelayanan sinar-X per hari (X_2) memiliki Pvaleu = 0.000 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pelayanan sinar-X per hari (X_2) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)

dengan R^2 = 97.56yang berarti bahwa model sudah bagus dan mampu menjlaskan 97.56% dari total keragaman data sedangkan sisanya dipengaruhi oleh faktor lain

- Step 3

Pada step 3 ada tiga variable yaitu banyaknya pasien rata-rata per hari (X_1), banyaknya pelayanan sinar-X per hari (X_2), tempat tidur yang terisi per bulan (X_3) yang memiliki nilai :

- banyaknya pasien rata-rata per hari (X_1) memiliki Pvaleu = 0.068 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pasien rata-rata per hari (X_1) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)

- banyaknya pelayanan sinar-X per hari (X_2) memiliki Pvaleu = 0.001 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pelayanan sinar-X per hari (X_2) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)
- tempat tidur yang terisi per bulan (X_3) memiliki Pvaleu = 0.002 < α = 0.1, sehingga dapat disimpulkan bahwa tempat tidur yang terisi per bulan (X_3) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)

Dengan R^2 = 98.87 yang berarti bahwa model sudah bagus dan mampu menjlaskan 98.87% dari total keragaman data sedangkan sisanya dipengaruhi oleh faktor lain.

- Step 4

Pada step 4 ada empat variable yaitu banyaknya pasien rata-rata per hari (X_1), banyaknya pelayanan sinar-X per hari (X_2), tempat tidur yang terisi per bulan (X_3) dan banyaknya penduduk disekitarnya yang mungkin memerlukan fasilitas(X_4) yang memiliki nilai :

- banyaknya pasien rata-rata per hari (X_1) memiliki Pvaleu = 0.017 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pasien rata-rata per hari (X_1) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)
- banyaknya pelayanan sinar-X per hari (X_2) memiliki Pvaleu = 0.013 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pelayanan sinar-X per hari (X_2) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)
- tempat tidur yang terisi per bulan (X_3) memiliki Pvaleu = 0.001 < α = 0.1, sehingga dapat disimpulkan bahwa tempat tidur yang terisi per bulan (X_3) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)
- banyaknya penduduk disekitarnya yang mungkin memerlukan fasilitas(X_4) memiliki Pvaleu = 0.001 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya penduduk disekitarnya yang mungkin memerlukan fasilitas(X_4) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)

Dengan R^2 = 99.11 yang berarti bahwa model sudah bagus dan mampu menjlaskan 99.11% dari total keragaman data sedangkan sisanya dipengaruhi oleh faktor lain.

- **5 variabel**

Stepwise Regression: y versus x1, x2, x3, x4, x5

```
Forward selection.  Alpha-to-Enter: 0.1
Response is y on 5 predictors, with N = 17


Step              1         2       3         4
Constant   -106.058  -118.422   1.989  1375.578

x1             33.7      24.0     9.1      13.3
T-Value       18.72      6.97    1.99      2.76
P-Value       0.000     0.000   0.068     0.017
```

x2		0.081	0.079	0.059
T-Value		3.10	4.26	2.92
P-Value		0.008	0.001	0.013
x3			0.51	0.51
T-Value			3.87	4.20
P-Value			0.002	0.001
x4				-279
T-Value				-1.81
P-Value				0.096
S	1163	927	656	606
R-Sq	95.90	97.56	98.87	99.11
R-Sq(adj)	95.62	97.22	98.61	98.81
Mallows Cp	37.7	19.1	5.0	4.0

Interpretasi :

Ketika di amsukkan 5 variabel banyaknya pasien rata-rata per hari (X_1), banyaknya pelayanan sinar-X per hari (X_2), tempat tidur yang terisi per bulan (X_3) dan banyaknya penduduk disekitarnya yang mungkin memerlukan fasilitas(X_4) dan rata-rata lamanya pasien dirawat (opname) dalam hari (X_5) maka tetap terdapat 4 step sebagai berikut:

- Step 1
  Pada step 1 hanya ada satu variable yaitu banyaknya pasien rata-rata per hari (X_1) yang memiliki nilai Pvaleu = 0.000 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pasien rata-rata per hari (X_1) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y) dengan R^2 = 95.90 yang berarti bahwa model sudah bagus dan mampu menjlaskan 95.90% dari total keragaman data sedangkan sisanya dipengaruhi oleh faktor lain.

- Step 2

Pada step 2 ada dua variable yaitu banyaknya pasien rata-rata per hari (X_1) dan banyaknya pelayanan sinar-X per hari (X_2) yang memiliki nilai :

- banyaknya pasien rata-rata per hari (X_1) memiliki Pvaleu = 0.000 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pasien rata-rata per hari (X_1) mempengaruhi Y
- banyaknya pelayanan sinar-X per hari (X_2) memiliki Pvaleu = 0.000 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pelayanan sinar-X per hari (X_2) mempengaruhi Y

dengan R^2 = 97.56yang berarti bahwa model sudah bagus dan mampu menjlaskan 97.56% dari total keragaman data sedangkan sisanya dipengaruhi oleh faktor lain

- Step 3

Pada step 3 ada tiga variable yaitu banyaknya pasien rata-rata per hari (X_1), banyaknya pelayanan sinar-X per hari (X_2) , tempat tidur yang terisi per bulan (X_3) yang memiliki nilai :

- banyaknya pasien rata-rata per hari (X_1) memiliki Pvaleu = 0.068 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pasien rata-rata per hari (X_1) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)
- banyaknya pelayanan sinar-X per hari (X_2) memiliki Pvaleu = 0.001 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pelayanan sinar-X per hari (X_2) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)
- tempat tidur yang terisi per bulan (X_3) memiliki Pvaleu = 0.002 < α = 0.1, sehingga dapat disimpulkan bahwa tempat tidur yang terisi per bulan (X_3) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)

Dengan R^2 = 98.87 yang berarti bahwa model sudah bagus dan mampu menjlaskan 98.87% dari total keragaman data sedangkan sisanya dipengaruhi oleh faktor lain.

- Step 4

Pada step 4 ada empat variable yaitu banyaknya pasien rata-rata per hari (X_1), banyaknya pelayanan sinar-X per hari (X_2) , tempat tidur yang terisi per bulan (X_3) dan banyaknya penduduk disekitarnya yang mungkin memerlukan fasilitas(X_4) yang memiliki nilai :

- banyaknya pasien rata-rata per hari (X_1) memiliki Pvaleu = 0.017 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pasien rata-rata per hari (X_1) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)
- banyaknya pelayanan sinar-X per hari (X_2) memiliki Pvaleu = 0.013 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pelayanan sinar-X per hari (X_2) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)
- tempat tidur yang terisi per bulan (X_3) memiliki Pvaleu = 0.001 < α = 0.1, sehingga dapat disimpulkan bahwa tempat tidur yang terisi per bulan (X_3) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)
- banyaknya penduduk disekitarnya yang mungkin memerlukan fasilitas(X_4) memiliki Pvaleu = 0.096 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya penduduk disekitarnya yang mungkin memerlukan fasilitas(X_4) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)

Dengan R^2 = 99.11 yang berarti bahwa model sudah bagus dan mampu menjlaskan 99.11% dari total keragaman data sedangkan sisanya dipengaruhi oleh faktor lain.

c.backward stepwise

- **5 variabel**

```
Stepwise Regression: y versus x1, x2, x3, x4, x5

Backward elimination.  Alpha-to-Remove: 0.1


Response is y on 5 predictors, with N = 17


Step              1       2
Constant       1302    1376

x1             12.5    13.3
T-Value        1.11    2.76
P-Value       0.291   0.017

x2            0.059   0.059
T-Value        2.79    2.92
P-Value       0.017   0.013

x3             0.51    0.51
T-Value        3.18    4.20
P-Value       0.009   0.001

x4             -265    -279
T-Value       -1.04   -1.81
P-Value       0.323   0.096

x5              0.6
T-Value        0.07
P-Value       0.944

S               632     606
R-Sq          99.11   99.11
R-Sq(adj)     98.71   98.81
Mallows Cp      6.0     4.0
```

Interpretasi ;

Ketika dimassukkan 5 variabel pada step 1 terdapat nilai sebagai berikut :

- banyaknya pasien rata-rata per hari (X_1) memiliki Pvaleu = 0.291 > α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pasien rata-rata per hari (X_1) tidak mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)
- banyaknya pelayanan sinar-X per hari (X_2) memiliki Pvaleu = 0.017 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pelayanan sinar-X per hari (X_2) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)
- tempat tidur yang terisi per bulan (X_3) memiliki Pvaleu = 0.009 < α = 0.1, sehingga dapat disimpulkan bahwa tempat tidur yang terisi per bulan (X_3) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)
- banyaknya penduduk disekitarnya yang mungkin memerlukan fasilitas(X_4) memiliki Pvaleu = 0.323 >α = 0.1, sehingga dapat disimpulkan bahwa banyaknya penduduk disekitarnya yang mungkin memerlukan fasilitas(X_4) tidak mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)
- rata-rata lamanya pasien dirawat (opname) dalam hari (X_5) memiliki Pvaleu = 0.323 >α = 0.1, sehingga dapat disimpulkan bahwa rata-rata lamanya pasien dirawat (opname) dalam

hari (X_5) tidak mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)

Dengan R^2 = 99.11 yang berarti bahwa model sudah bagus dan mampu menjlaskan 99.11% dari total keragaman data sedangkan sisanya dipengaruhi oleh faktor lain.

Kareana rata-rata lamanya pasien dirawat (opname) dalam hari (X_5) memiliki nilai Pvalue teresar maka rata-rata lamanya pasien dirawat (opname) dalam hari (X_5) dihapus.

- **4 variabel**

```
Stepwise Regression: y versus x1, x2, x3, x4

Backward elimination.  Alpha-to-Remove: 0.1

Response is y on 4 predictors, with N = 17

Step              1
Constant       1376

x1             13.3
T-Value        2.76
P-Value       0.017

x2            0.059
T-Value        2.92
P-Value       0.013

x3             0.51
T-Value        4.20
P-Value       0.001

x4             -279
T-Value       -1.81
P-Value       0.096

S               606
R-Sq          99.11
R-Sq(adj)     98.81
Mallows Cp      5.0
```

Interpretasi :

Setelah di hapus rata-rata lamanya pasien dirawat (opname) dalam hari (X_5) maka step 1 memiliki nilai sebagai berikut :

- banyaknya pasien rata-rata per hari (X_1) memiliki Pvaleu = 0.017 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pasien rata-rata per hari (X_1) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)
- banyaknya pelayanan sinar-X per hari (X_2) memiliki Pvaleu = 0.013 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya pelayanan sinar-X per hari (X_2) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)

- tempat tidur yang terisi per bulan (X_3) memiliki Pvaleu = 0.001 < α = 0.1, sehingga dapat disimpulkan bahwa tempat tidur yang terisi per bulan (X_3) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y)
- banyaknya penduduk disekitarnya yang mungkin memerlukan fasilitas(X_4) memiliki Pvalue = 0.096 < α = 0.1, sehingga dapat disimpulkan bahwa banyaknya penduduk disekitarnya yang mungkin memerlukan fasilitas(X_4) mempengaruhi banyaknya jam kerja per bulan yang dipakai di rumah sakit tersebut (Y).

Dengan R^2 = 99.11 yang berarti bahwa model sudah bagus dan mampu menjlaskan 99.11% dari total keragaman data sedangkan sisanya dipengaruhi oleh faktor lain.

Dari output diatas dapat dilihat bahwa pada step 1 tidak ada lagi P*value* yang bernilai lebih besar dari *alpha to remove* yang bernilai 0.1. maka tidak ada lagi faktor yang harus dihapus.

Prosedur R

Pada prosedur R, kita dapat menggunakan sintaks

```
step(object, scope, scale = 0,
     direction = c("both", "backward", "forward"),
     trace = 1, keep = NULL, steps = 1000, k = 2, ...)
```

`object` merupakan formula dari persamaan regresi linear berganda.

`direction` merupakan metode yang ingin digunakan, yaitu

`both` ➔ Stepwise

`backward` ➔ Backward Selection

`forward` ➔ Forward Selection

Berikut adalah proses stepwise dari kasus di atas

```
> #memanggil paket
> library(DescTools)
> library(lmtest)
Loading required package: zoo

Attaching package: ‘zoo’

The following objects are masked from ‘package:base’:

    as.Date, as.Date.numeric

> #memanggil data
> bertatar=read.delim("clipboard")
```

```
> View(bertatar)
> #melihat korelasi antar variabel bebas
> cor(bertatar[,-6])
          x1        x2        x3        x4        x5
x1 1.0000000 0.9089931 0.9680132 0.6796277 0.9492448
x2 0.9089931 1.0000000 0.8829603 0.4529689 0.9114322
x3 0.9680132 0.8829603 1.0000000 0.6561524 0.8863133
x4 0.6796277 0.4529689 0.6561524 1.0000000 0.4737482
x5 0.9492448 0.9114322 0.8863133 0.4737482 1.0000000
> #membuat model
> modelber=lm(y~., data=bertatar)
> modelber

Call:
lm(formula = y ~ ., data = bertatar)

Coefficients:
(Intercept)           x1           x2           x3           x4            x
5
  1301.6169      12.5461       0.0592       0.5144    -265.1212        0.634
5

> #melihat koefisien determinan
> summary(modelber)

Call:
lm(formula = y ~ ., data = bertatar)

Residuals:
    Min      1Q  Median      3Q     Max 
-629.91 -403.71   10.33  267.14 1530.00 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)   
(Intercept) 1301.61693 1323.35763   0.984  0.34647   
x1            12.54606   11.31743   1.109  0.29127   
x2             0.05920    0.02119   2.794  0.01746 * 
x3             0.51436    0.16161   3.183  0.00872 **
x4          -265.12118  256.09156  -1.035  0.32277   
x5             0.63447    8.89277   0.071  0.94440   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 632.4 on 11 degrees of freedom
Multiple R-squared:  0.9911,	Adjusted R-squared:  0.9871 
F-statistic:   245 on 5 and 11 DF,  p-value: 6.852e-11

> #melakukan pengujian multikolinearitas
> VIF(modelber)
        x1         x2         x3         x4         x5 
133.692590   8.136370  26.172947   6.590953  37.221692 
> #jika terdapat multikolinearitas, maka untuk mencegahnya dengan dilakukan
stepwise
> #untuk model fordward
> modelstep=step(modelber, direction = "forward")
Start:  AIC=223.89
y ~ x1 + x2 + x3 + x4 + x5

> summary(modelstep)
```

```
Call:
lm(formula = y ~ x1 + x2 + x3 + x4 + x5, data = bertatar)

Residuals:
    Min      1Q  Median      3Q     Max 
-629.91 -403.71   10.33  267.14 1530.00 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)   
(Intercept) 1301.61693 1323.35763   0.984  0.34647   
x1            12.54606   11.31743   1.109  0.29127   
x2             0.05920    0.02119   2.794  0.01746 * 
x3             0.51436    0.16161   3.183  0.00872 **
x4          -265.12118  256.09156  -1.035  0.32277   
x5             0.63447    8.89277   0.071  0.94440   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 632.4 on 11 degrees of freedom
Multiple R-squared:  0.9911,	Adjusted R-squared:  0.9871 
F-statistic:   245 on 5 and 11 DF,  p-value: 6.852e-11

> #untuk model backward
> modelstep2=step(modelber, direction = "backward")
Start:  AIC=223.89
y ~ x1 + x2 + x3 + x4 + x5

       Df Sum of Sq     RSS    AIC
- x5    1      2036 4401864 221.89
- x4    1    428688 4828515 223.47
- x1    1    491543 4891371 223.69
<none>              4399828 223.89
- x2    1   3122493 7522321 231.00
- x3    1   4051495 8451322 232.98

Step:  AIC=221.89
y ~ x1 + x2 + x3 + x4

       Df Sum of Sq      RSS    AIC
<none>               4401864 221.89
- x4    1   1197087  5598951 223.98
- x1    1   2799982  7201846 228.26
- x2    1   3121649  7523513 229.01
- x3    1   6483193 10885057 235.28
> summary(modelstep2)

Call:
lm(formula = y ~ x1 + x2 + x3 + x4, data = bertatar)

Residuals:
    Min      1Q  Median      3Q     Max 
-634.64 -420.94   11.17  265.89 1530.29 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)   
(Intercept) 1375.57787  787.75587   1.746  0.10630   
x1            13.26991    4.80306   2.763  0.01719 * 
x2             0.05913    0.02027   2.917  0.01291 * 
```

```
x3             0.50713    0.12063   4.204  0.00122 **
x4          -279.30418  154.61165  -1.806  0.09597 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 605.7 on 12 degrees of freedom
Multiple R-squared:  0.9911,   Adjusted R-squared:  0.9881
F-statistic:   334 on 4 and 12 DF,  p-value: 3.458e-12

> #untuk model gabungan
> modelstep3=step(modelber, direction = "both")
Start:  AIC=223.89
y ~ x1 + x2 + x3 + x4 + x5

       Df Sum of Sq     RSS    AIC
- x5    1      2036 4401864 221.89
- x4    1    428688 4828515 223.47
- x1    1    491543 4891371 223.69
<none>              4399828 223.89
- x2    1   3122493 7522321 231.00
- x3    1   4051495 8451322 232.98

Step:  AIC=221.89
y ~ x1 + x2 + x3 + x4

       Df Sum of Sq      RSS    AIC
<none>               4401864 221.89
+ x5    1      2036  4399828 223.89
- x4    1   1197087  5598951 223.98
- x1    1   2799982  7201846 228.26
- x2    1   3121649  7523513 229.01
- x3    1   6483193 10885057 235.28
> summary(modelstep3)

Call:
lm(formula = y ~ x1 + x2 + x3 + x4, data = bertatar)

Residuals:
    Min      1Q  Median      3Q     Max
-634.64 -420.94   11.17  265.89 1530.29

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 1375.57787  787.75587   1.746  0.10630
x1            13.26991    4.80306   2.763  0.01719 *
x2             0.05913    0.02027   2.917  0.01291 *
x3             0.50713    0.12063   4.204  0.00122 **
x4          -279.30418  154.61165  -1.806  0.09597 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 605.7 on 12 degrees of freedom
Multiple R-squared:  0.9911,   Adjusted R-squared:  0.9881
F-statistic:   334 on 4 and 12 DF,  p-value: 3.458e-12
```