

PEMODELAN REGRESI LINIER DENGAN SOFTWARE R

Regresi linier adalah suatu metode analisis untuk mengetahui hubungan fungsional antar variabel. Regresi linier dibedakan menjadi dua, yaitu regresi linier sederhana dan regresi linier berganda.

A. REGRESI LINIER SEDERHANA

Regresi linier sederhana merupakan metode regresi yang hanya memiliki satu peubah bebas (variabel prediktor). Secara matematis, regresi linier sederhana dituliskan sebagai berikut.

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

$$\hat{Y} = \beta_0 + \beta_1 X$$

Dengan

Y : Nilai pengamatan dari variabel tak bebas (variabel dependen/variabel respon)

\hat{Y} : Nilai ramalan/prediksi dari variabel tak bebas (variabel dependen/variabel respon)

X : Nilai pengamatan dari variabel bebas (variabel independen/variabel prediktor)

β_0 : intersep atau konstanta

β_1 : *slope* atau koefisien kemiringan model regresi

ε : *error* atau nilai kesalahan prediksi

Paket (*package*) untuk analisis regresi dengan bantuan *software* R biasanya telah ter-*install* otomatis dalam *software* (package **stats**).

Contoh aplikasi analisis regresi dengan *software* R menggunakan data: waktu yang diperlukan untuk memproduksi suatu produk (*run time/rt*) dan jumlah produk yang akan di produksi (*run size/rs*), data dipilih secara acak, dan didapatkan 20 jenis produk (Sheather, 2009). Dalam modul ini akan diteliti apakah jumlah produk yang diproduksi (disebut sebagai variabel prediktor: X) memengaruhi waktu produksi (disebut sebagai variabel respon Y).

Langkah-langkah, *syntax*, *output*, dan interpretasi *output* yang dilakukan untuk analisis regresi dengan *software* R adalah sebagai berikut.

1. Membuat data frame

Misalkan data frame yang dibuat diberi nama *production* yang mempunyai dua variabel: variabel *Y*, yaitu run time (rt) dan variabel *X* adalah run size (rs). Perintah-perintah yang digunakan sebagai berikut.

```
> production <- data.frame(rt=c(195, 215, 243, 162, 185, 231,
234, 166, 253, 196, 220, 168, 207, 225, 189, 213, 147, 230, 208, 172),
rs=c(175, 189, 344, 88, 114, 338, 271, 173, 284, 277, 337, 58, 146, 68))
> #untuk memanggil data, gunakan perintah "nama data frame",
yaitu production
> production
```



```
> production <- data.frame(rt=c(195, 215, 243, 162, 185, 231,
, 253, 196, 220, 168, 207, 225, 189, 213, 147, 230, 208, 172), rs=c(17
5, 189, 344, 88, 114, 338, 271, 173, 284, 277, 337, 58, 146, 277, 123,
227, 63, 337, 146, 68))
> production
  rt  rs
1 195 175
2 215 189
3 243 344
4 162  88
5 185 114
6 231 338
7 234 271
8 166 173
9 253  58
10 196 277
11 168 337
12 207  88
13 225 146
14 189 277
15 169 123
16 215 227
17 147  63
18 230 337
19 208 146
20 172  68
```

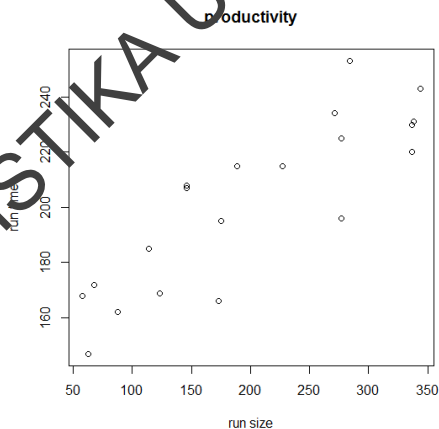
Gambar 1. Function dan Output Data Frame Production

Gambar 1. menampilkan data frame dari data production, yang meliputi data run time (rt) dan run size (rs) masing-masing sebanyak 20 observasi jenis produk.

2. Membuat *scatter plot* dan analisis korelasi untuk data production

Function yang digunakan untuk membuat *scatter plot* adalah **plot**. Tanda \sim merupakan tanda yang digunakan untuk menjalankan proses regresi. Formula yang digunakan adalah variabel respon \sim variabel predictor 1+variabel predictor 2+...+variabel predictor p .

```
> plot(rt ~ rs, data = production, xlab="run size", ylab="run time", main="productivity")
```



Gambar 2. Scatterplot antara Run Size (X) dan Run Time (Y)

```

> #xlab, ylab, dan main merupakan perintah untuk memberi
judul
> cor.test(production$rt, production$rs)
  > cor.test(production$rt,production$rs)

      Pearson's product-moment correlation

data:  production$rt and production$rs
t = 6.9798, df = 18, p-value = 1.615e-06
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.6625316 0.9411514
sample estimates:
      cor
0.8545206

```

Gambar 3. *Function dan Output Matriks Korelasi antara X dan Y*

Gambar 2. menampilkan scatterplot antara variabel X dan variabel Y dari 20 data observasi. Berdasarkan Gambar 3. diketahui nilai korelasi antara X dan Y sebesar 0.854, menunjukkan bahwa terdapat hubungan positif antara variabel X dan Y . p -value dari pengujian *pearson correlation* ($1.615e-06$) lebih kecil dari $\alpha = 5\%$ sehingga hubungan antara variabel X dan Y signifikan secara statistik.

3. Membuat model regresi dari data *production*

Function yang digunakan untuk membuat model regresi adalah **lm**. Misalkan “reg” adalah nama untuk model regresi yang terbentuk.

```

> reg <- lm(rt ~ rs, data=production)
> #gunakan perintah “reg” untuk memanggil output

```

> reg

```
> reg <- lm(rt ~ rs, data=production)
> reg

Call:
lm(formula = rt ~ rs, data = production)

Coefficients:
(Intercept)          rs
    149.7477         0.2592
```

Gambar 4. Formula dan *Output* Model Regresi

Berdasarkan Gambar 4. diketahui bahwa model regresi dari data production, yaitu hubungan antara jumlah produk (X) dan lama produksi (Y) adalah sebagai berikut.

$$\hat{Y} = 149.7477 + 0.259X$$

Dari model regresi tersebut, diketahui bahwa terdapat hubungan positif antara variabel X dan Y yang ditunjukkan oleh nilai slope positif sebesar 0.259. Nilai slope tersebut diartikan sebagai rata-rata kenaikan waktu produksi (Y) sebesar 0.259 satuan waktu akibat kenaikan per satuan jumlah produk (X).

Function summary digunakan untuk melihat hasil regresi data production lebih lengkap. Beberapa informasi yang diberikan adalah nilai residual, koefisien

regresi beserta nilai signifikansi (p-value), dan nilai- nilai untuk uji kebaikan model

```
> summary(reg)
> summary(reg)

Call:
lm(formula = rt ~ rs, data = production)

Residuals:
    Min       1Q   Median       3Q      Max
-28.597 -11.079   3.329   8.302  29.627

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 149.74770     8.32815   17.98 6.00e-13 ***
rs           0.25924     0.03714    6.98 1.61e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

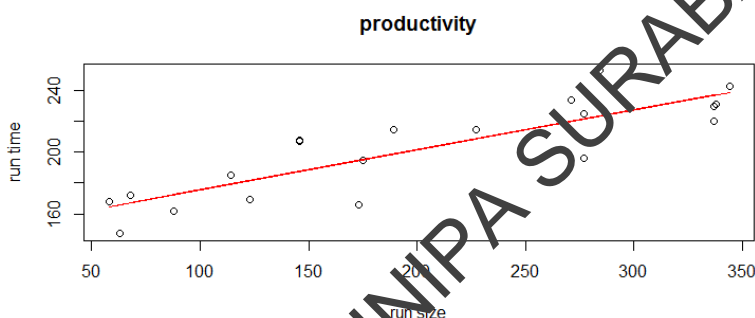
Residual standard error: 16.25 on 18 degrees of freedom
Multiple R-squared:  0.7302,    Adjusted R-squared:  0.7152
F-statistic: 48.72 on 1 and 18 DF,  p-value: 1.615e-06
```

Gambar 5. Function dan Output Summary dari Model Regresi

Nilai minimum dan maksimum residual yang dihasilkan oleh model regresi sebesar -28.597 dan 29.627. Berdasarkan hasil uji t pada Gambar 5., diketahui bahwa jumlah produk (X) berpengaruh signifikan terhadap waktu produksi (Y). Hal tersebut disebabkan p-value dari uji t (sebesar $1.61e-06$) lebih kecil dari $\alpha = 5\%$. Nilai *Adjusted R squared* menunjukkan bahwa variabel prediktor (X) mampu menjelaskan variansi dari variabel

respon (Y) sebesar 71.52%, persen sisanya dijelaskan oleh variabel lain (Gambar 5).

Gambar 6. merupakan plot dengan garis regresi antara X dengan Y , menunjukkan variasi observasi yang kecil dari garis regresi yang terbentuk.



Gambar 6. Plot dengan Garis Regresi

Gambar 6. dapat diperoleh dengan menggunakan formula-formula berikut.

```
>plot(production$rs, production$rt, xlab="run size",  
ylab="run time", main="productivity")  
>lines(production$rs, fitted(reg), col="red") #untuk membuat  
garis regresi antara variabel rs ( $X$ ) dan nilai prediksi  $Y$   
(fitted values  $\hat{Y}$ )
```


4. Selang Kepercayaan (*Confident Interval*) 95%

Function **confint** merupakan *function* untuk menghitung taksiran interval dari satu atau lebih parameter pada model yang diestimasi.

```
> confint(reg, level=0.95)
      > confint(reg, level=0.95)
              2.5 %      97.5 %
(Intercept) 132.2509062 167.2444991
rs           0.1812107  0.3372105
```

Gambar 7. Formula dan *Output* Selang Kepercayaan (*Confident Interval*) Parameter $\hat{\beta}_0$ dan $\hat{\beta}_1$

Gambar 7. menunjukkan nilai $\hat{\beta}_0$ (parameter taksiran intersep) terletak antara 132.251 dan 167.244 sementara nilai $\hat{\beta}_1$ (parameter taksiran X) terletak antara 0.181 dan 0.337.

$$132.251 \leq \hat{\beta}_0 \leq 167.244$$
$$0.181 \leq \hat{\beta}_1 \leq 0.337$$

Nilai selang kepercayaan yang tidak melewati nilai nol menyimpulkan bahwa jumlah produk (rs) berpengaruh terhadap waktu produksi (rt).

5. Menampilkan Analysis of Variance (ANOVA)

anova merupakan *function* yang digunakan untuk menampilkan hasil *Analysis of Variance* (ANOVA) dari

model regresi yang telah terbentuk. Tabel ANOVA menunjukkan kesesuaian model, yang diketahui dari signifikansi parameter terhadap variabel respon yang dilakukan secara bersamaan dengan uji F.

```
> anova(reg)
> anova(reg)
Analysis of Variance Table

Response: rt
      Df Sum Sq Mean Sq F value    Pr(>F)
rs      1 12868.4 12868.4  48.717 1.615e-06 ***
Residuals 18  4754.6    264.1
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Gambar 8. Formula dan *Output* ANOVA

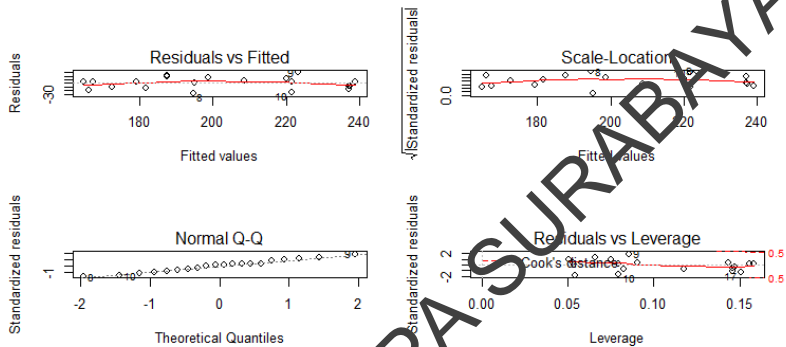
Gambar 8. menunjukkan bahwa model regresi linier sederhana yang telah terbentuk telah mewakili data yang ada. Hal tersebut disebabkan oleh p-value dari uji F (nilai statistik F sebesar 48.717) lebih kecil dari $\alpha = 5\%$, yaitu sebesar $1.615e-06$.

6. *Diagnostic Plot*

Diagnostic plot menampilkan 4 plot, yaitu (1) plot antara *fitted values* (nilai-nilai prediksi) dan residual, (2) plot antara kuantil-kuantil normal dari *standardized residuals* (residual yang distandarisasi), (3) plot antara *fitted values* dan akar *standardized residuals*, dan (4) plot antara *lverage* dan *standardized residual*. Formula-

formula yang digunakan untuk membuat *diagnostic plot* adalah sebagai berikut.

```
> layout(matrix(c(1,2,3,4),2,2)) # untuk membuat 4 plot dalam 1 halaman  
> plot(reg)
```



Gambar 9. *Diagnostic Plot*

Berdasarkan Gambar 9. diketahui bahwa tidak terjadi heteroskedastisitas karena plot antara *fitted values* dan *standardized residual* tidak membentuk pola tertentu.

7. Evaluasi Kesesuaian dan Asumsi Model Regresi

Evaluasi kesesuaian model dilakukan dengan pengecekan multikolinieritas. Asumsi-asumsi yang harus terpenuhi dalam analisis regresi sebelum dilakukan inferensia statistik adalah (1) varians residual homogen (homoscedasticity), (2) autokorelasi, dan (3) residual berdistribusi normal.

Function yang digunakan untuk pengecekan multikolinieritas adalah **vif**. Sebelum menggunakan **vif**, *library car* harus ter-*install* terlebih dahulu.

```
> library(car)
> vif(reg)
-
> library(car)
> vif(reg)
Error in vif.default(reg) : model contains fewer than 2 terms
```

Gambar 10. Formula Pengujian Multikolinieritas

Pada Gambar 10. terlihat adanya error dalam komputasi perhitungan nilai VIF, sebab multikolinieritas dapat terjadi pada model regresi yang memiliki minimal dua variabel predictor *X*.

Sementara itu, formula-formula yang digunakan untuk melakukan pengujian asumsi-asumsi yang harus dipenuhi oleh model regresi sebagai berikut (perlu diingat bahwa pengujian asumsi dilakukan terhadap residual dari model regresi).

```
>residual= residuals(reg) #untuk memanggil residual yang dihasilkan oleh model regresi
>library(stats) #package untuk deteksi normalitas
shapiro=shapiro.test(residual) #deteksi normalitas
>library(lmtest) #package untuk deteksi homoskedas dan autokorelasi
>homoskedas=bptest(reg) #deteksi homoskedas
>autokorr=dwtest(reg) #deteksi autokorelasi
```

Berdasarkan uji shapiro-wilk yang tersaji pada Gambar 11. diketahui bahwa residual telah berdistribusi normal karena p-value (sebesar 0.892) lebih besar dari $\alpha = 5\%$ (H_0 : residual berdistribusi normal, gagal ditolak).

```

> residual=residuals(reg)
> residual
  1          2          3          4          5          6          7          8          9         10
-0.1152469 16.2553496  4.0726679 -10.5610965  5.6987827  1.5580439  2.5580439  2.5580439  2.5580439  2.5580439
  6          7          8          9         10         11         12         13         14         15
-6.3718734 13.9974148 -28.5967607 29.6272544  21.5580439  21.5580439  21.5580439  21.5580439  21.5580439  21.5580439
 16         17         18         19         20
-17.1126303  3.2161967 19.4028033  3.4419561 -12.6346053  6.4041115 -19.0800188 -7.1126303 20.4028033  4.6237657
> library(stats) #package untuk deteksi normalitas
> shapiro=shapiro.test(residual) #deteksi normalitas
> shapiro

      Shapiro-Wilk normality test

data:  residual
W = 0.97712, p-value = 0.89171

> library(lmtest) #package untuk deteksi homoskedas dan autokorelasi
> homoskedas=bptest(reg)
> homoskedas

      studentized Breusch-Pagan test

data:  reg
BP = 0.10128, df = 1, p-value = 0.7503

> autokorr=dwtest(reg)
> autokorr

      Durbin-watson test

data:  reg
dw = 2.7608, p-value = 0.9669
alternative hypothesis: true autocorrelation is greater than 0

```

Gambar 11. Function dan Output Uji Asumsi Residual

Uji homoskedastisitas dilakukan dengan uji Breusch-Pagan (uji BP). P-value yang dihasilkan oleh uji

BP (sebesar 0.750) lebih besar dari $\alpha = 5\%$ sehingga asumsi homoskedastisitas residual terpenuhi.

Uji durbin-watson merupakan uji untuk mendeteksi adanya autokorelasi residual. Gambar 11. menunjukkan bahwa tidak ada autokorelasi pada residual model regresi karena p-value (sebesar 0.967) lebih besar dari $\alpha = 5\%$ (H_0 : tidak ada autokorelasi residual, gagal ditolak). Berdasarkan pengujian asumsi-asumsi yang telah dilakukan, seluruh asumsi dalam model regresi telah terpenuhi, sehingga dapat disimpulkan model regresi telah sesuai menggambarkan data dan inferensia statistik dapat dilakukan.

B. REGRESI LINIER BERGANDA

Perbedaan mendasar antara regresi linier sederhana dan regresi linier berganda terletak pada jumlah variabel prediktor. Regresi linier sederhana hanya mempunyai satu variabel prediktor sementara regresi linier berganda mempunyai lebih dari satu variabel prediktor. Semua pengujian yang dilakukan pada regresi

linier sederhana juga dilakukan pada regresi linier berganda.

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \varepsilon$$

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

1. Memanggil Data yang akan Digunakan

Data yang akan digunakan sebagai contoh aplikasi model regresi berganda adalah data Menu Pricing (Sheather, 2009), yaitu harga menu (Price, Y) di salah satu restoran khas Itali di Manhattan, rating yang diberikan pelanggan untuk makanan (Food, x_1), rating yang diberikan pelanggan untuk dekorasi (Decor, x_2), dan rating yang diberikan pelanggan untuk pelayanan (Service, x_3).

Misalkan data tersimpan di direktori D: dalam folder - UNIPA dengan nama file data nyc.csv. Untuk memanggil data yang tersimpan dalam format csv ke R, *function* yg digunakan adalah **read.csv**. Misalkan data yang akan dipanggil dinamai "data".

```
> data=read.csv("D:\\- UNIPA\\MODUL R\\regresi\\Datasets  
Simon J. Sheather\\Data\\nyc.csv",header=TRUE) #tanda "\\\""
```

digunakan untuk membuka folder yang menyimpan data, header=TRUE merupakan perintah bahwa baris pertama dalam dataset csv merupakan judul variabel

```
> data[1:5,] #jika ingin menampilkan data yang telah di  
eksport ke R hanya pada baris ke-1 sampe ke-5
```

Karena dataset nyc.csv yang telah terekpor ke R dengan nama data, mempunyai memiliki 7 kolom, tetapi tidak semua kolom akan digunakan dalam analisis regresi, maka dibuat dataset baru yang hanya memuat variabel-variabel yang digunakan dalam analisis, yaitu variabel dalam kolom ke-3 sampai kolom ke-6 (Gambar 12).

```
> dt=data[,3:6] #marupakan dataset baru (nama: dt) yang  
diambil dari dataset "data" pada kolom ke-3 sampai ke-6
```

```
> data=read.csv("D:\\- UNIPA\\MODUL R\\regresi\\Datasets Simon J. Shea  
ther\\Data\\nyc.csv",header=TRUE)  
> data[1:5,]
```

Case	Restaurant	Price	Food	Decor	Service	East
1	Daniella Ristorante	43	22	18	20	0
2	Tello's Ristorante	32	20	19	19	0
3	Virichino	34	21	13	18	0
4	Bottino	41	20	20	17	0
5	Da Umberto	54	24	19	21	0

```
> dt=data[,3:6]
```

```
> dt[1:5,]
```

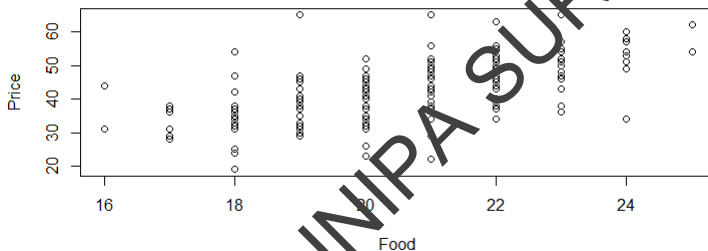
	Price	Food	Decor	Service
1	43	22	18	20
2	32	20	19	19
3	34	21	13	18
4	41	20	20	17
5	54	24	19	21

Gambar 12. Function dan Dataset Menu Pricing

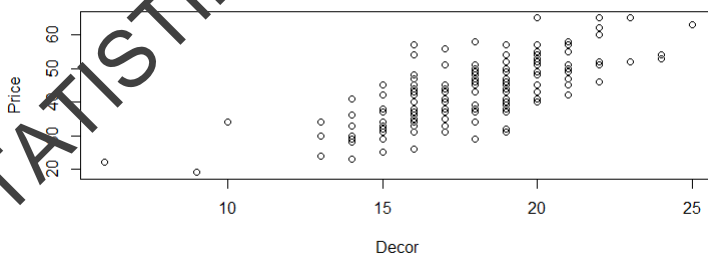
2. Membuat *scatter plot* dan Analisis Korelasi untuk Data dt (Menu Pricing)

Function yang digunakan untuk menampilkan scatterplot antara variabel Y dan X seperti pada Gambar 13. adalah **plot** dengan formula sebagai berikut.

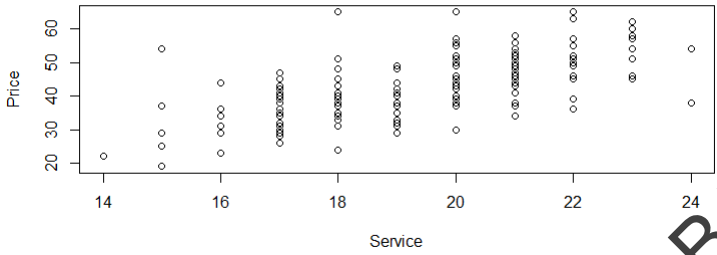
```
> plot(Price ~ Food+Decor+Service, data=dt)
```



Gambar 13a. Scatterplot antara Price Y dan Food x_1



Gambar 13b. Scatterplot antara Price Y dan Decor x_2



Gambar 13c. Scatterplot antara Price Y dan Service X_3

Gambar 14a. menunjukkan nilai korelasi antar variabel X dan Y cukup besar, bernilai diatas 0.5. Nilai korelasi setiap variabel bernilai positif, sehingga dapat disimpulkan bahwa terjadi hubungan searah antar variabel. Misalkan jika rating makanan (food) meningkat maka harga (price) akan meningkat. Informasi dari Gambar 14b. adalah korelasi antar variabel signifikan secara statistik karena nilai p-value yang dihasilkan disetiap variabel prediktor dan respon lebih kecil dari $\alpha = 5\%$ (Tolak H_0).

```
> cor(dt, method="pearson")
      Price      Food      Decor      Service
Price 1.0000000 0.6270435 0.7243525 0.6411402
Food  0.6270435 1.0000000 0.5039161 0.7945248
Decor 0.7243525 0.5039161 1.0000000 0.6453306
Service 0.6411402 0.7945248 0.6453306 1.0000000
```

Gambar 14a. Function dan Output Korelasi

```

> cor.test(dt$Price,dt$Food)

Pearson's product-moment correlation

data: dt$Price and dt$Food
t = 10.371, df = 166, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5255275 0.7109548
sample estimates:
      cor
0.6270435

> cor.test(dt$Price,dt$Decor)

Pearson's product-moment correlation

data: dt$Price and dt$Decor
t = 13.537, df = 166, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.6435202 0.7892068
sample estimates:
      cor
0.7243525

> cor.test(dt$Price,dt$Service)

Pearson's product-moment correlation

data: dt$Price and dt$Service
t = 10.764, df = 166, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5423821 0.7224210
sample estimates:
      cor
0.6411402

```

Gambar 14b. *Function* dan *Output* p-value Uji Korelasi Pearson

3. Membuat Model Regresi dari Data dt

Seperti pada regresi linier sederhana, membuat model regresi linier berganda pada R ditunjukkan pada Gambar 15. Berdasarkan Gambar 15 dapat dituliskan model regresi yang terbentuk sebagai berikut.

$$\hat{Y} = -24.641 + 1.556x_1 + 1.847x_2 + 0.135x_3$$

Model regresi yang dihasilkan menjelaskan bahwa setiap kenaikan 1 satuan rating atas makanan disaat variabel lainnya tetap maka akan menaikkan harga menu makanan sebesar \$ US 1.556. Jika rating atas dekorasi naik sebesar 1 satuan rating maka harga makanan akan bertambah sebesar 1.847 (\$US). Sedangkan jika rating pelayanan bertambah 1 satuan maka harga makanan akan bertambah sebesar 0.135 (\$US).

Pengujian inferensi yang pertama dilakukan adalah pengujian pengaruh variabel predictor secara bersamaan terhadap respon dengan menggunakan uji F. Hipotesis awal (H_0) dari pengujian ini adalah sebagai berikut.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_a : \beta_{food} = \beta_{decor} = \beta_{service} = 0$$

Jika H_0 ditolak maka pengujian inferensi selanjutnya, yaitu uji individu parameter dapat dilakukan. Berdasarkan Gambar 15 diketahui bahwa H_0 ditolak karena p-value dari uji F lebih kecil dari $\alpha = 5\%$. Oleh karena itu, uji individu parameter dapat dilakukan.

Uji individu parameter dilakukan dengan uji t. Berdasarkan uji t, diketahui bahwa variabel x_3 tidak berpengaruh signifikan terhadap Y sebab p-value dari uji t lebih besar dari $\alpha = 5\%$. Nilai *adjusted R-Squared* menunjukkan bahwa variabel prediktor (X) mampu menjelaskan varians dari variabel respon (Y) sebesar 61%. Nilai residual yang dihasilkan dari model regresi berkisar antara -14.8440 dan 19.058 (Gambar 15).

```

> reg1=lm(Price ~ Food+Decor+Service, data=dt)
> reg1

Call:
lm(formula = Price ~ Food + Decor + Service, data = dt)

Coefficients:
(Intercept)      Food      Decor      Service
   -24.641      1.556      1.847      0.135

> summary(reg1)

Call:
lm(formula = Price ~ Food + Decor + Service, data = dt)

Residuals:
    Min       1Q   Median       3Q      Max
-14.8440  -3.7039  -0.1525   3.6218  19.0576

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -24.6409    4.7536  -5.184 6.33e-07 ***
Food          1.5556     0.3731   4.170 4.93e-05 ***
Decor         1.8473     0.2176   8.491 1.17e-14 ***
Service       0.1350     0.3957   0.341  0.733
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.803 on 164 degrees of freedom
Multiple R-squared:  0.617,    Adjusted R-squared:  0.61
F-statistic: 88.06 on 3 and 164 DF,  p-value: < 2.2e-16

```

Gambar 15. Function dan Output Regresi

4. Selang Kepercayaan (*Confident Interval*) 95%

Gambar 16. menunjukkan selang kepercayaan untuk masing-masing parameter taksiran dari variabel predictor yang dapat dituliskan sebagai berikut.

$$-34.027 \leq \hat{\beta}_0 \leq -15.254$$

$$0.819 \leq \hat{\beta}_1 \leq 2.292$$

$$1.418 \leq \hat{\beta}_2 \leq 2.277$$

$$-0.646 \leq \hat{\beta}_3 \leq 0.916$$

```
> confint(reg1, level=0.95)
                2.5 %          97.5 %
(Intercept) -34.0270654 -15.2547257
Food          0.8189773    2.2922347
Decor         1.4177674    2.2769029
Service      -0.6461439    0.9162753
```

Gambar 16. Selang Kepercayaan Parameter Taksiran Regresi

Nilai selang kepercayaan pada parameter taksiran untuk variabel service x_3 yang memuat nilai nol menunjukkan bahwa variabel x_3 tidak berpengaruh terhadap variabel Y .

5. Menampilkan Analysis of Variance (ANOVA)

Analysis of Variance pada regresi berganda dengan *software* R merupakan uji F secara parsial (*partial F-test*). Berdasarkan Gambar 16., diketahui bahwa hasil uji F

secara parsial menyebutkan bahwa variabel x_3 tidak berpengaruh karena hasil uji F memiliki p-value lebih besar dari $\alpha = 5\%$ (gagal tolak H_0).

```
> anova(reg1)
Analysis of Variance Table

Response: Price
  Df Sum Sq Mean Sq F value Pr(>F)
Food  1 5670.3   5670.3  168.3552 <2e-16 ***
Decor  1 3223.7   3223.7   95.7130 <2e-16 ***
Service 1  3.9     3.9    0.1165  0.7333
Residuals 164 5523.6    33.7
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Gambar 17. Tabel Anova

Pada *software* R, dapat diketahui pengaruh penambahan variabel prediktor baru terhadap model. Analisis dilakukan dengan bantuan model regresi lain yang memiliki variabel predictor dengan model yang telah dibentuk. dalam modul ini adalah model **reg1**) yang sama ditambahkan minimal satu variabel prediktor baru. Misalkan model baru yang akan dibentuk dinamai model regresi **reg2** dengan penambahan variabel prediktor **East** yang ada pada dataset awal. ANOVA dengan R dilakukan dengan membandingkan model regresi **reg1** dengan model **reg2**. *Function* yang digunakan sebagai berikut.

```
> anova(reg1, reg2)
```

```

> dt1=data[,3:7]
> dt1[1,5]
[1] 0
> dt1=data[,3:7]
> dt1[1:5,]
  Price Food Decor Service East
1    43   22   18      20     0
2    32   20   19      19     0
3    34   21   13      18     0
4    41   20   20      17     0
5    54   24   19      21     0
> reg2=lm(Price ~ Food+Decor+Service+East, data=dt1)
> anova(reg1,reg2)
Analysis of Variance Table

Model 1: Price ~ Food + Decor + Service
Model 2: Price ~ Food + Decor + Service + East
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
  ---
1     164 5523.6
2     163 5366.5 1     157.1 4.7716 0.03036 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

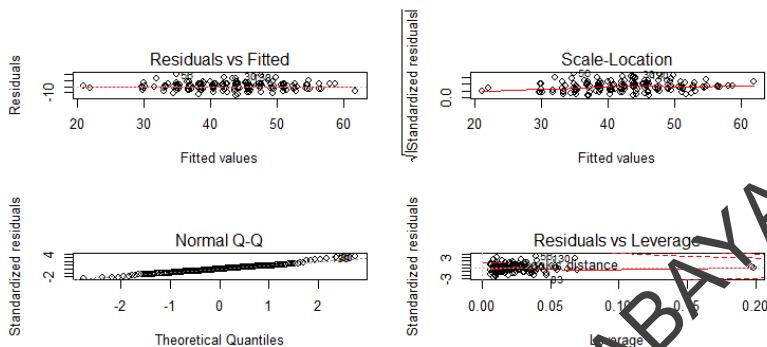
Gambar 18. Langkah-langkah ANOVA Regresi Berganda di R

Berdasarkan Gambar 18. diketahui bahwa terdapat pengaruh yang signifikan yang diberikan oleh variabel **east** terhadap harga menu ketika variabel prediktor x_1 , x_2 , dan x_3 telah diperhitungkan. Hal tersebut disebabkan oleh p-value dari uji F lebih kecil dari $\alpha = 5\%$ (Tolak H_0).

6. Diagnostic Plot, Evaluasi Kesesuaian dan Asumsi

Model Regresi

Berdasarkan Gambar 19. diketahui bahwa tidak terjadi heteroskedastisitas karena plot antara *fitted values* dan *standardized residual* tidak membentuk pola tertentu.



Gambar 19. Diagnostic Plot

Evaluasi kesesuaian dan asumsi model regresi ditampilkan pada Gambar 20. Berdasarkan Gambar 20. Diketahui bahwa tidak terdapat multikolinieritas pada model regresi karena nilai VIF yang dihasilkan kurang dari 10.

Berdasarkan informasi pada Gambar 19., disimpulkan bahwa residual telah berdistribusi normal sebab p-value bernilai 0.079 (lebih besar dari $\alpha = 5\%$). Asumsi homoskedastisitas residual juga terpenuhi karena p-value dari uji BP lebih besar dari $\alpha = 5\%$. Akan tetapi, asumsi tidak ada autokorelasi terlanggar karena p-value dari uji Durbin-Watson lebih kecil dari $\alpha = 5\%$ (H_0 : tidak ada autokorelasi residual, ditolak).

```

> library(car)
> vif(reg1)
  Food   Decor  Service
2.712989 1.714273 3.468564
> residual1= residuals(reg1) #untuk memanggil residual yang dihasilkan
oleh model regresi
> library(stats) #package untuk deteksi normalitas
> shapiro=shapiro.test(residual1) #deteksi normalitas
> shapiro

      Shapiro-Wilk normality test

data:  residual1
W = 0.98553, p-value = 0.07927

> library(lmtest) #package untuk deteksi homoskedas dan autokorelasi
> homoskedas=bptest(reg1)
> homoskedas

      studentized Breusch-Pagan test

data:  reg1
BP = 2.9501, df = 3, p-value = 0.3994

> autokorr=dwtest(reg1)
> autokorr

      Durbin-Watson test

data:  reg1
DW = 1.7081, p-value = 0.02593
alternative hypothesis: true autocorrelation is greater than 0

```

Gambar 20. Evaluasi Kesesuaian Model dan Deteksi Asumsi

Daftar Pustaka

Draper, N.R. dan Smith, H. 1981. *Applied Regression Analysis. Second Edition*. JohnWiley & Sons, Inc.

Help R

Sheater, Simon. J. 2009. *A Modern Approach to Regression with R*. USA. Springer.

Suhartono. 2008. *Analisis Data Statistik dengan R*.
Surabaya. Lab. Statistik Komputasi, ITS.

STATISTIKA UNIPA SURABAYA