Nama: Ibnu Fajar Setiawan

NIM: 065002000006

UNIVERSITAS TRISAKTI

Modul 6

Nama Dosen: Syandra Sari, S.Kom,

M.Kom

Nama Aslab:

1. Ida Jubaidah

(06500190037)

2. Azzahra Nuranisa

(065001900044)

Hari/Tanggal:

Minggu, 6 November 2022

Praktikum Data Analitik

Praktikum 6 REGRESI BERGANDA DENGAN SEBAGIAN PREDIKTOR BERSIFAT KATEGORIK DAN REGRESI NON LINIER

DESKRIPSI MODUL: Melakukan pengujian asumsi.

No	Elemen Kompetensi	Indikator Kinerja	Jml Jam	hlm
1	Mampu melakukan pengujian regresi berganda dengan sebagian prediktor bersifat kategorik dan regresi non linier	Dapat melakukan pengujian regresi berganda dengan sebagian prediktor bersifat kategorik dan regresi non linier	2	

TEORI SINGKAT

Jika variabel dependen-nya dihubungkan dengan lebih dari satu variabel independen, maka persamaan yang dihasilkan adalah persamaan regresi linier berganda ($multiple\ linier\ regression$). Dalam hal ini kita membatasi pada kasus dua peubah bebas X_1 dan X_2 saja. Dengan hanya dua peubah bebas, persamaan regresi contohnya menjadi :

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2$$

Variabel terikat (respon) dalam regresi linier berganda harus bersifat kontinu sedangkan variabel bebas dapat seluruhnya bersifat kontinu atau sebagian bersifat kategorik. Bila peubah bebas (predictor) bersifat kategori maka dapat diubah menjadi sekumpulan variabel dummy dengan catatan bila dalam variabel kategorik tersebut terdapat n kategori maka harus terdapat sejumlah (n-1) variabel dummy.

Salah satu ukuran kebaikan model adalah dengan melihat koefisien determinasi R² yang menyatakan proporsi keragaman variabel Y yang dapat dijelaskan oleh variabel X. Namun penggunaan yang lebih baik adalah dengan menggunakan nilai **R-Sq(adj)**, yang merupakan nilai estimasi yang tidak bias (*unbiased estimate*) dari populasi.

ELEMEN KOMPETENSI I

Deskripsi :

Dapat melakukan pengujian regresi berganda dengan sebagian

prediktor bersifat kategorik dan regresi non linier

Kompetensi Dasar:

Mampu melakukan pengujian regresi berganda dengan sebagian

prediktor bersifat kategorik dan regresi non linier

Kasus 1 : Regresi dengan satu var independent bersifat kualitatif (2 kelas) 🛘 1 var dummy

Y =vol penjualan;

X1=pengeluaran iklan;

X2=media (1=tv, 0=koran)

Perusahaa n	Volume Penjualan (Y)	Pengeluara n Iklan (X1)	Jenis Media (x2)	
1	620	250	tv	
2	140	120	koran	
3	500	175	tv	

X2	
	1
	0
	1

4	210	135	koran
5	460	180	tv
6	250	140	koran
7	200	130	koran
8	380	170	tv
9	400	200	tv
10	215	150	koran
11	395	175	tv
12	425	215	tv
13	235	145	koran
14	475	190	tv
15	195	130	koran

General Regression (MINITAB)

Regression Equation

X2 (1=tv, 0=koran) 0 (koran) 1 (tv)

Y = -75,412 + 2,07672 X1Y = 53,2126 + 2,07672 X1

Coefficients

Term Coef SE Coef T P
Constant -11,0997 100,926 -0,10998 0,914
X1 2,0767 0,607 3,41996 0,005
X2 (1=tv, 0=koran)
0 -64,3123 21,425 -3,00176 0,011

Summary of Model

S = 46,0191 R-Sq = 91,06% R-Sq(adj) = 89,57% PRESS = 54182,6 R-Sq(pred) = 80,95%

Hasil regresi mengindikasikan bahwa variabel kualitatif jenis media iklan yang digunakan berpengaruh terhadap volume penjualan. Perusahaan yang menggunakan media tv mempunyai volume penjualan lebih tinggi dibandingkan yang menggunakan surat kabar yang ditunjukkan oleh perbedaan koefisien titik potong dengan sumbu Y pada kedua persamaan regresi di atas yaitu -75,412 dan 53,2126 dengan perbedaan sebesar 128,625. Koefisien regresi pengeluaran iklan dan jenis media iklan juga signifikan yang terlihat pada nilai p yang kurang dari 5 %.

0

1 0

Analysis of Variance

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	2	258937	258937	129468	61,1347	0,000001
X1	1	239855	24770	24770	11,6961	0,005079
X2 (1=tv, 0=koran)	1	19082	19082	19082	9,0105	0,011031
Error	12	25413	25413	2118		
Lack-of-Fit	10	19888	19888	1989	0,7199	0,706453
Pure Error	2	5525	5525	2763		
Total	14	284350				

R script:

```
> con = dbConnect(MySQL(), user = 'root', password = '', dbname = 'dblatihan', host = 'localhost')
> myQuery <- "select * from reg;"
> reg <- dbGetQuery(con, myQuery)</pre>
> head(reg)
  perusahaan volume_penjualan pengeluaran_iklan jenis_media x2
                                 620
                                                         250
                                                                         tv 1
                                 140
                                                                      koran 0
              2
                                                         120
3
              3
                                 500
                                                         175
                                                                         tν
                                                                               1
                                 210
                                                         135
                                                                      koran
                                                                              0
5
                                 460
                                                         180
                                                                          tν
6
              6
                                                         140
                                                                      koran
                                 250
```

```
> model=Im(Y \sim X1 + X2, data = reg)
```

> model

OUTPUT

> summary(model)

```
> summary(model)
 lm(formula = volume_penjualan ~ pengeluaran_iklan + jenis_media,
     data = reg)
 Residuals:
              1Q Median
     Min
                               3Q
 -74.707 -23.947
                   5.055 30.094 83.361
 Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
 (Intercept)
                   -75.4120 84.2261 -0.895 0.38821
                                          3.420 0.00508 **
 pengeluaran_iklan 2.0767
                                0.6072
 jenis_mediatv 128.6246
                                42.8498
                                          3.002 0.01103 *
 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
 Residual standard error: 46.02 on 12 degrees of freedom
 Multiple R-squared: 0.9106, Adjusted R-squared: 0.8957
 F-statistic: 61.13 on 2 and 12 DF, p-value: 5.096e-07
Penjelasan:
Dapat kita lihat model regresi yang terjadi adalah
Y = -75.4120 + 2.0767 X1 + 128.6246 X2
Jika Tv (1) : Y = 53.2126 + 2.0767 X1
Jika Koran (0): Y = -75.4120 + 2.0767 X1
Dengan R-Square = 91.06%
```

Kasus 2: Regresi dengan satu var independent bersifat kualitatif (3 kelas) 🛘 2 var dummy

Y=gaji(juta) X1=masa kerja (tahun) X2=pendidikan (sma, diploma, sarjana)

00= sma, 10=diploma, 01=sarjana

Gaji Y	Masa Kerja	Pendidika	
Gaji i	iviasa Kerja	n	
3	1	Sarjana	
2.4	5	Diploma	
2.15	12	SMA	

X2	Х3
0	1
1	0
0	0

3.8	5	Sarjana
5	10	Sarjana
1.75	5	SMA
2	1	Diploma
2.2	3	Diploma
2.05	10	SMA
2.8	8	Diploma
3.2	2	Sarjana
1.5	1	SMA
1.55	2	SMA
3.6	4	Sarjana
1.9	8	SMA
3.5	15	Diploma
2.35	15	SMA
2.7	20	SMA
6	15	Sarjana
2.45	17	SMA

0	1
0	1
0	0
1	0
1	0
0	0
1	0
0	1
0	0
0	0
0	1
0	0
1	0
0	0
0	0
0	1
0	0

General regression (MINITAB)

x20 y = 0.973004 + 0.107144 x11 y = 1.89428 + 0.107144 x1

y = 3.43928 + 0.107144 x1

Coefficients

Term Coef SE Coef T P
Constant 0,97300 0,206329 4,7158 0,000
Masa Kerja X1 0,10714 0,015908 6,7353 0,000
Pendidikan X2 0,92127 0,227207 4,0548 0,001
Pendidikan X3 2,46627 0,216523 11,3904 0,000

Summary of Model

S = 0,394194 R-Sq = 90,08% R-Sq(adj) = 88,22% PRESS = 4,34619 R-Sq(pred) = 82,65%

Analysis of Variance

 Source
 DF
 Seq SS
 Adj SS
 Adj MS
 F
 P

 Regression
 3
 22,5683
 22,5683
 7,5228
 48,413
 0,0000000

 Masa Kerja X1
 1
 2,3014
 7,0490
 7,0490
 45,364
 0,0000048

 Pendidikan X2
 1
 0,1067
 2,5548
 2,5548
 16,441
 0,0009198

 Pendidikan X3
 1
 20,1602
 20,1602
 20,1602
 129,741
 0,0000000

 Error
 16
 2,4862
 2,4862
 0,1554

 Total
 19
 25,0545
 25,0545

```
Obs Gaji Y Fit SE Fit Residual St Resid
19 6 5,04644 0,213644 0,953561 2,87844 R
R denotes an observation with a large standardized residual.
 > con = dbConnect(MySQL(), user = 'root', password = ", dbname =
                           'db_reg', host = 'localhost')
 > myQuery <- "select * from reg2;"
 > View(reg2)
 > View(reg2)
 > head(reg2)
 OUTPUT
   > View(reg2)
   > head(reg2)
      ji_y X1 Pendidikan X2 X3
  1 3.00 1 Sarjana 0 1
2 2.40 5 Diploma 1 0
3 2.15 12 SMA 0 0
4 3.80 5 Sarjana 0 1
5 5.00 10 Sarjana 0 1
6 1.75 5 SMA 0 0
 > model=lm(Y \sim X1 + X2 + X3, data = reg2)
 > model
 OUTPUT
   > model1=lm(ji_y \sim X2 + X2 + X3, data = reg2)
   > model1
   Call:
   lm(formula = ji_y \sim X2 + X2 + X3, data = reg2)
   Coefficients:
   (Intercept)
                             X2
                                              X3
                       0.5356
                                          2.0556
         2.0444
 > summary(model)
 OUTPUT
```

Fits and Diagnostics for Unusual Observations

```
> summary(model1)
Call:
lm(formula = ji_y \sim X2 + X2 + X3, data = reg2)
Residuals:
   Min
            10 Median
                            3Q
                                   Max
-1.1000 -0.4958 -0.1622 0.3306 1.9000
Coefficients:
           Estimate Std. Error t value Pr(>|t|)
                                 8.189 2.65e-07 ***
(Intercept)
             2.0444 0.2496
X2
             0.5356
                        0.4177
                                 1.282
                                          0.217
                                 5.208 7.11e-05 ***
Х3
             2.0556
                        0.3947
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.7489 on 17 degrees of freedom
Multiple R-squared: 0.6194, Adjusted R-squared: 0.5746
F-statistic: 13.83 on 2 and 17 DF, p-value: 0.0002715
```

Kasus 3: Regresi non linier dengan satu variabel independent

An engineer at a computer manufacturing company wants to understand the relationship between strength of plastic and the predictors temperature. He suspects the relationship between temperature and strength is quadratic. He collects 14 samples. The engineer subjects the samples to various temperatures, then measures the strength of the plastic (File: Plastic.mtw)

Temp X	Strength Y
185	5150
183	5125
187	5123
188	5140
189	5195
189	5190
192	5150
195	5155
196	5156
198	5162

193	5172
196	5196
200	5063
202	5025
196	5105
197	5176
195	5160
200	5100
205	4956
206	4960
203	4975
199	5063
200	5090
206	4946
196	5175
207	4905
208	4760

General Regression Analysis: Strength Y versus Temp X (MINITAB)

```
Regression Equation
```

```
Strength Y = -38711,3 + 459,45 Temp X - 1,20233 Temp X*Temp X
```

Coefficients

Term	Coef	SE Coef	Т	P
Constant	-38711,3	4881,77	-7 , 92978	0,000
Temp X	459 , 5	49,83	9,21991	0,000
Temp X*Temp X	-1,2	0,13	-9,46251	0,000

Pembahasan : Kasus di atas diolah dengan menggunakan menu General Regression dalam Minitab. Hal ini karena model yang ingin digunakan adalah model kuadratik ($y = \beta o + \beta 1x + \beta 2x^2$) yang lebih sesuai dengan pola datanya yang melengkung bila dibuat scatter plot. Berdasarkan uji statistik terhadap koefisien regresi di atas dapat dilihat bahwa seluruh koefisien regresi memiliki p value yang kurang dari taraf nyata (α) sebesar 1% yang berarti sangat signifikan. Hal ini menandakan bahwa ketiga koefisien secara sangat signifikan tidak sama dengan nol.

Analysis of Variance

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	2	276023	276023	138012	126,950	0,000000
Temp X	1	178683	92414	92414	85 , 007	0,000000
Temp X*Temp X	1	97341	97341	97341	89 , 539	0,000000
Error	24	26091	26091	1087		
Lack-of-Fit	16	20690	20690	1293	1,915	0,177045

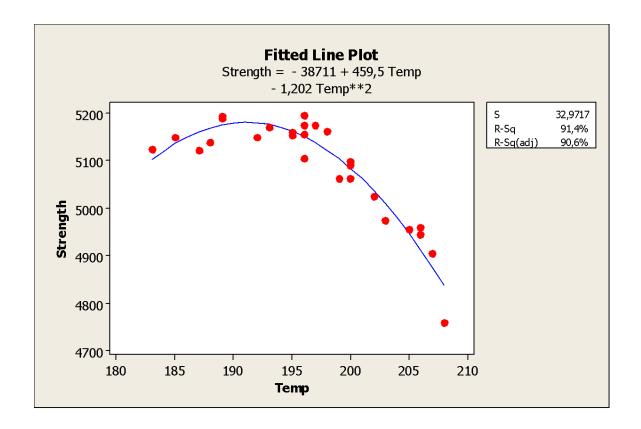
Pure Error 8 5402 5402 675 Total 26 302115

Berdasarkan Analysis of Variance (ANOVA) terlihat Regression memiliki p value < α (1%) berarti minimal terdapat satu koefisien regresi yang tidak sama dengan nol. Uji F pada ANOVA digunakan untuk mengevaluasi pengaruh semua variabel independen terhadap dependen). Hasil Lack of Fit juga menunjukkan tidak signifikan (p value > α) yang berarti model dapat diterima.

Fits and Diagnostics for Unusual Observations

```
Obs Strength Y Fit SE Fit Residual St Resid
2 5125 5103,09 21,3071 21,9136 0,87089 X
27 4760 4836,52 17,3724 -76,5191 -2,73050 R
```

R denotes an observation with a large standardized residual. X denotes an observation whose X value gives it large leverage.



Bagaimana jika kasus 1 dimodelkan secara linier ? Nilai koefisien determinasinya (R-square) akan turun

General Regression Analysis: Strength Y versus Temp X

Regression Equation

Strength Y = 7449,5 - 12,0059 Temp X

Coefficients

```
Term Coef SE Coef T P
Constant 7449,50 392,797 18,9653 0,000
Temp X -12,01 1,996 -6,0159 0,000
```

Summary of Model

```
S = 70,2658   R-Sq = 59,14%   R-Sq(adj) = 57,51%   PRESS = 155356   R-Sq(pred) = 48,58%
```

Analysis of Variance

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	1	178683	178683	178683	36,1905	0,0000028
Temp X	1	178683	178683	178683	36,1905	0,0000028
Error	25	123432	123432	4937		
Lack-of-Fit	17	118030	118030	6943	10,2827	0,0011422
Pure Error	8	5402	5402	675		
Total	26	302115				

Fits and Diagnostics for Unusual Observations

```
Obs Strength Y Fit SE Fit Residual St Resid

2 5125 5252,41 30,5092 -127,414 -2,01297 R

27 4760 4952,27 26,2888 -192,267 -2,95056 R
```

R denotes an observation with a large standardized residual.

Script R

```
> reg=read.delim("clipboard")
> View(reg)
OUTPUT
> head(reg)
OUTPUT
 Console Terminal × Jobs ×
                                                                                    > reg3=read.delim("clipboard")
> View(reg3)
 > head(reg3)
 X Y X2
1 185 5150 34225
2 183 5125 33489
3 187 5123 34969
 4 188 5140 35344
 5 189 5195 35721
 6 189 5190 35721
> model=lm(Y \sim X + x2, data = reg)
> model
OUTPUT
```

```
> model=lm(y \sim x + x2, data = reg)
> model
Call:
lm(formula = y \sim x + x2, data = reg)
Coefficients:
(Intercept)
                                   x2
               459.450 -1.202
 -38711.311
> summary(model)
OUTPUT
> summary(model)
Call:
 lm(formula = y \sim x + x2, data = reg)
Residuals:
             1Q Median
    Min
                             3Q
                                    Max
 -76.519 -26.172 4.662 22.439 46.848
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
 (Intercept) -3.871e+04 4.882e+03 -7.930 3.69e-08 ***
             4.595e+02 4.983e+01 9.220 2.34e-09 ***
Х
x2
            -1.202e+00 1.271e-01 -9.463 1.43e-09 ***
Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
Residual standard error: 32.97 on 24 degrees of freedom
Multiple R-squared: 0.9136, Adjusted R-squared: 0.9064
F-statistic: 127 on 2 and 24 DF, p-value: 1.721e-13
```

ELEMEN KOMPETENSI II

Deskripsi

Dapat melakukan pengujian regresi berganda dengan sebagian prediktor bersifat kategorik dan regresi non linier dengan data directmarketing.csv

Kompetensi Dasar:

Mampu melakukan pengujian regresi berganda dengan sebagian prediktor bersifat kategorik dan regresi non linier dengan data directmarketing.csv

Gunakan data directmarketing.csv. Lakukan analisis regresi untuk melihat **pengaruh dari salary** dan gender terhadap amountspent serta pengaruh dari salary dan age terhadap amountspent. Lakukan pula uji asumsi terhadap kedua model regresi yang didapatkan.

salary dan gender terhadap amountspent

- > tugas1_reg=read.delim("clipboard")
- > head(tugas1_reg)

OUTPUT

```
con = dbConnect(MySQL(), user = 'root', password = '', dbname = 'houseprices', host = 'localhost')
myQuery <- "select * from market_1;"
market_1 <- dbGetQuery(con, myQuery)</pre>
> head(market_1)
     Age Gender OwnHome Married Location Salary Children History Catalogs AmountSpent
     Old Female
                      Own Single
Rent Single
Rent Single
                                           Far 47500
                                                                 0
                                                                       High
                                                                                     6
2 Middle Male
                                         Close 63600
                                                                 0
                                                                                     6
                                                                                                1318
                                                                       Hiah
3 Young Female
                                         Close
                                                 13500
                                                                 0
                                                                                    18
                                                                                                 296
                                                                       Low
                      Own Married
                                                                                                2436
4 Middle
          Male
                                        Close 85600
                                                                       High
                                                                                    18
5 Middle Female
                       Own Single
                                         Close
                                                  68400
                                                                                                1304
                                                                       High
            Male
                       Own Married
                                         Close
                                                 30400
                                                                                                 495
  Young
```

Gender diganti untuk Female = 1 Male = 0

```
> View(tugas1_reg)
> model=Im(Y ~ X1 + X2, data = tugas1_reg)
> model
```

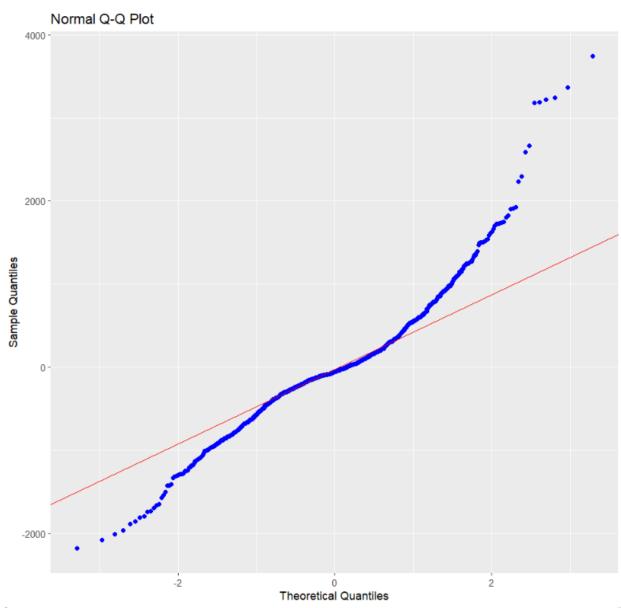
OUTPUT

> summary(model)

```
> summary(model)
Call:
lm(formula = AmountSpent ~ Salary + Gender, data = market_1)
Residuals:
   Min
            1Q Median
                           3Q
-2180.6 -323.1 -53.7
                        282.8 3742.8
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.352e+01 5.645e+01 0.239
                                        0.811
                                          <2e-16 ***
            2.180e-02 7.357e-04 29.626
Salary
Gender1
           -3.867e+01 4.503e+01 -0.859
                                          0.391
Signif. codes: 0 '*** 0.001 '** 0.01 '* 0.05 '. '0.1 ' '1
Residual standard error: 687.2 on 997 degrees of freedom
Multiple R-squared: 0.4898, Adjusted R-squared: 0.4888
F-statistic: 478.6 on 2 and 997 DF, p-value: < 2.2e-16
```

ols_plot_resid_qq(model)





- > Imtest::bptest(model)
- > dwtest(model)
- > ols_vif_tol(model)

```
> ols_plot_resid_qq(model)
> lmtest::bptest(model)
        studentized Breusch-Pagan test
data: model
BP = 165.5, df = 2, p-value < 2.2e-16
> dwtest(model)
        Durbin-Watson test
data: model
DW = 2.002, p-value = 0.5127
alternative hypothesis: true autocorrelation is greater than 0
> ols_vif_tol(model)
  Variables Tolerance
                           VIF
     Salary 0.9316218 1.073397
2
    Gender1 0.9316218 1.073397
> |
```

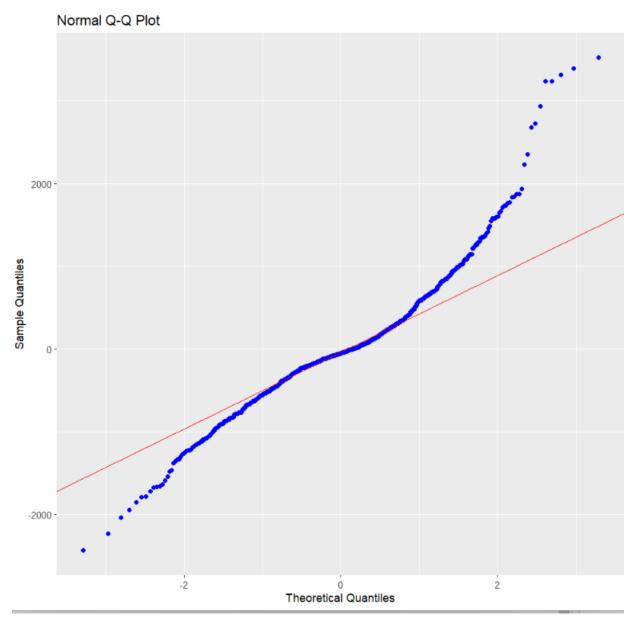
salary dan age terhadap amountspent

```
> tugas2_reg=read.delim("clipboard")
> View(tugas2_reg)
> head(tugas2_reg)
OUTPUT
> model=lm(Y ~ X1 + X2, data = tugas2_reg)
> model
OUTPUT
 > model=lm(AmountSpent ~ Salary + Age, data = market_1)
 > model
 Call:
 lm(formula = AmountSpent ~ Salary + Age, data = market_1)
 Coefficients:
 (Intercept)
                    Salary
                                   Age01d
                                               AgeYoung
  -109.26153
                   0.02236
                               280.87715
                                               48.07869
> summary(model)
```

```
> summary(model)
lm(formula = AmountSpent ~ Salary + Age, data = market_1)
Residuals:
               1Q Median
                                   3Q
    Min
                                           Max
-2434.5 -347.4
                    -50.8 278.2 3520.6
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.093e+02 7.116e+01 -1.535 0.125
Salary 2.236e-02 8.949e-04 24.991 < 2e-16 ***
AgeOld 2.809e+02 5.793e+01 4.849 1.44e-06 ***
AgeYoung 4.808e+01 6.394e+01 0.752 0.452
Signif. codes: 0 '*** 0.001 '** 0.01 '* 0.05 '.' 0.1 ' '1
Residual standard error: 679.2 on 996 degrees of freedom
Multiple R-squared: 0.502, Adjusted R-squared: 0.5005
F-statistic: 334.7 on 3 and 996 DF, p-value: < 2.2e-16
```

ols_plot_resid_qq(model)





- > Imtest::bptest(model)
- > dwtest(model)
- > ols_vif_tol(model)

1. Cek List

Elemen Kompetensi	Penyelesaian		
	Selesai	Tidak	
Elemen Kompetensi I Dapat melakukan pengujian regresi berganda dengan sebagian prediktor bersifat kategorik dan regresi non linier	✓		
Elemen Kompetensi II Dapat melakukan pengujian regresi berganda dengan sebagian prediktor bersifat kategorik dan regresi non linier dengan data directmarketing.csv	✓		

2. Form Umpan Balik

Elemen Kompetensi	Waktu Pengerjaan	Kriteria
Elemen Kompetensi I Dapat melakukan pengujian regresi berganda dengan sebagian prediktor bersifat kategorik dan regresi non linier	60 menit	1
Elemen Kompetensi II Dapat melakukan pengujian regresi berganda dengan sebagian prediktor bersifat kategorik dan regresi non linier dengan data directmarketing.csv	60 menit	1

Kriteria

- 1.Sangat Menarik
- 2.Cukup Menarik
- 3.Kurang Menarik
- 4.Sangat Kurang Menarik