

**LAPORAN**  
**DATA SCIENCE CLASS (DSC) 2021**

**ANALISIS PENGARUH INDEK PEMBANGUNAN MANUSIA (IPM) DAN TINGKAT  
PENGANGGURAN TERHADAP KEMISKINAN DI INDONESIA**



Nama : Siti Fatonah  
NIM : 192400001  
Kelas : Statistika 2019 A  
Jurusan : Statistika

PROGRAM STUDI STATISTIKA FAKULTAS SAINS TEKNOLOGI  
UNIVERSITAS PGRI ADI BUANA SURABAYA  
TAHUN 2021/2022

## 1. Deskripsi Dataset

Data yang digunakan pada analisis ini adalah data sekunder yang diperoleh dari data penelitian yang dilakukan oleh peneliti terdahulu. Dalam analisis ini data yang digunakan terdiri dari dua variabel independen dan satu variabel dependen, yaitu data indeks pembangunan masyarakat ( $X_1$ ) dan tingkat pengangguran ( $X_2$ ) serta data kemiskinan ( $Y$ ).

Kemiskinan	IPM	Pengangguran
16.98	68.81	9.02
17.11	69.45	9.93
16.43	70.00	7.57
15.92	70.60	6.57
4.76	72.48	1.90
5.25	73.27	1.99
4.15	73.65	1.89
4.14	74.30	1.48
4.97	68.27	5.14
4.83	69.05	6.29
5.04	69.55	2.60
5.30	69.99	3.78
5.51	69.89	9.07
5.75	70.27	9.55
5.36	70.96	8.92
5.59	71.42	9.28
17.09	68.06	3.47
17.16	68.59	4.91
17.03	69.33	3.30
15.59	69.95	3.74
4.09	78.39	8.47
3.61	78.99	7.23
3.75	79.60	6.12
3.78	80.06	7.14
17.41	65.17	4.18
18.16	65.86	4.65
17.63	66.29	2.76
17.14	67.01	4.28
9.18	68.80	8.45
9.57	69.50	8.72
8.77	70.05	8.89
7.83	70.69	8.22
8.39	68.24	5.08
9.12	68.89	4.34

8.37	69.62	4.00
7.90	69.99	3.87
13.58	68.78	5.68
13.32	69.49	4.99
13.19	69.98	4.63
12.23	70.52	4.57
12.28	68.14	4.19
12.28	68.95	4.47
11.85	69.74	4.21
11.20	70.27	4.00
8.07	64.89	4.04
8.44	65.59	5.15
8.00	65.88	4.23
7.86	66.26	4.36
4.81	67.63	3.80
4.72	68.38	4.92
4.52	69.05	5.45
4.70	69.65	4.77
6.07	67.77	3.24
5.91	68.53	4.54
5.36	69.13	4.82
5.26	69.79	4.23
6.31	73.82	7.38
6.10	74.17	7.50
6.00	74.59	7.95
6.08	75.12	6.91
6.00	68.64	5.50
6.32	68.76	5.68
6.99	69.20	5.23
6.96	69.84	5.54
6.40	73.40	6.69
5.78	73.75	6.20
5.84	73.99	7.69
6.13	74.45	7.16
14.21	66.42	4.79
13.53	66.95	5.14
13.86	67.65	4.62
13.04	68.25	4.33
18.44	66.74	10.51
19.36	67.05	9.93
19.26	67.60	7.05

18.29	68.19	9.29
7.41	65.18	5.29
6.22	65.91	6.05
6.41	66.63	4.01
6.44	67.20	5.33
17.05	64.31	5.75
16.54	65.19	5.69
16.02	65.81	3.94
15.05	66.58	3.32
19.60	62.26	3.26
22.58	62.67	3.83
22.01	63.13	3.25
21.38	63.73	3.27
26.26	61.28	5.02
25.73	61.73	8.08
24.88	62.21	7.46
23.12	62.99	6.49
27.80	56.75	3.44
28.40	57.25	3.99
28.40	58.05	3.35
27.76	59.09	3.62
7.99	70.33	6.56
8.82	70.84	7.83
7.67	71.20	7.43
7.41	71.79	6.22
12.05	62.24	2.08
11.90	62.96	3.35
11.19	63.60	3.33
11.18	64.30	3.21
9.54	68.49	5.08
10.12	69.15	5.95
9.24	69.76	4.80
9.48	70.34	5.61
13.61	66.43	3.68
14.07	66.76	4.10
14.09	67.47	3.29
14.22	68.11	3.81
12.77	68.07	4.43
13.74	68.75	5.55
12.77	69.31	2.72
11.97	69.86	3.30

8.26	66.43	7.54
8.98	70.39	9.03
8.20	71.05	6.18
7.90	71.66	7.18
6.89	69.36	6.50
6.71	69.98	6.89
7.14	70.73	5.09
6.75	71.24	5.58
13.62	66.75	4.96
13.77	67.46	6.07
13.39	68.24	4.31
13.10	68.86	4.39
9.85	68.87	6.23
10.79	69.51	6.71
10.27	70.00	5.84
9.28	70.57	5.60
14.55	76.81	3.33
13.16	77.59	4.07
13.10	78.38	2.72
12.36	78.89	3.02

## 2. Preprocessing Data

### 1) Import library

```
In [2]: import pandas as pd
import numpy as np
import statsmodels
import patsy
import statsmodels.api as sm
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn import metrics
```

Import library yang diperlukan dalam analisis.

### 2) Memanggil data

```
In [3]: data = pd.read_csv('siti_dsc.csv')
data
```

Syntax diatas digunakan untuk memanggil data yang akan di analisis sehingga menghasilkan output sebagai berikut:

Out[3]:

	Kemiskinan	IPM	Pengangguran
0	16.98	68.81	9.02
1	17.11	69.45	9.93
2	16.43	70.00	7.57
3	15.92	70.60	6.57
4	4.76	72.48	1.90
...	...	...	...
131	9.28	70.57	5.60
132	14.55	76.81	3.33
133	13.16	77.59	4.07
134	13.10	78.38	2.72
135	12.36	78.89	3.02

136 rows × 3 columns

### 3) Penentuan variabel

```
In [6]: nama=['IPM', 'Pengangguran']
X=data[nama]
Y=data.Kemiskinan
X
```

Syntax diatas digunakan untuk menentukan variabel yang dijadikan variabel independen dan variabel dependen dan menghasilkan output sebagai berikut:

Out[6]:

	IPM	Pengangguran
0	68.81	9.02
1	69.45	9.93
2	70.00	7.57
3	70.60	6.57
4	72.48	1.90
...	...	...
131	70.57	5.60
132	76.81	3.33
133	77.59	4.07
134	78.38	2.72
135	78.89	3.02

136 rows × 2 columns

#### 4) Data testing dan training

```
[9]: X_train, X_test, Y_train, Y_test = train_test_split(X,Y,random_state=1)
```

Syntax diatas digunakan untuk membagi data menjadi data testing dan data training.

#### 5) Analisis regresi berganda

Berikut ini merupakan syntax-syantax yang digunakan untuk analisis regresi berganda:

- Model regresi linier

```
In [10]: from sklearn.linear_model import LinearRegression  
LinReg=LinearRegression()
```

- Membuat model dengan data latih

```
In [11]: LinReg.fit(X_train,Y_train)
```

```
Out[11]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

- Membuat prediksi pada pengujian data

```
In [13]: y_pred=LinReg.predict(X_test)
```

- Menghitung RMSE

```
In [15]: print(np.sqrt(metrics.mean_squared_error(Y_test,y_pred)))
```

- Membuat model regresi

```
In [16]: model=sm.OLS(Y,X).fit()  
predictions=model.predict(X)  
model.summary()
```

Dari syntax diatas menghasilkan output sebagai berikut:

Out[16]: OLS Regression Results

Dep. Variable:	Kemiskinan	R-squared (uncentered):	0.751
Model:	OLS	Adj. R-squared (uncentered):	0.747
Method:	Least Squares	F-statistic:	202.1
Date:	Sun, 30 Jan 2022	Prob (F-statistic):	3.52e-41
Time:	22:06:45	Log-Likelihood:	-445.61
No. Observations:	136	AIC:	895.2
Df Residuals:	134	BIC:	901.0
Df Model:	2		
Covariance Type:	nonrobust		
	coef	std err	t P> t  [0.025 0.975]
IPM	0.1675	0.024	6.999 0.000 0.120 0.215
Pengangguran	-0.0793	0.289	-0.274 0.784 -0.651 0.492
Omnibus:	21.690	Durbin-Watson:	0.345
Prob(Omnibus):	0.000	Jarque-Bera (JB):	26.352
Skew:	1.019	Prob(JB):	1.90e-06
Kurtosis:	3.706	Cond. No.	36.2

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correct

### 3. Modelling dan evaluating

- Model regresi

Dari hasil output diatas sehingga dapat diperoleh model regresi linier berganda sebagai berikut:

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

$$\hat{Y} = \beta_0 + 0,1675X_1 + (-0,0793) X_2$$

- Uji parsial

- Hipotesis:

H0:  $\beta_i = 0, i = 0, 1, 2$  ( Tidak ada pengaruh secara signifikan antara X dan Y)

H1:  $\beta_i \neq 0, i = 0, 1, 2$  ( Ada pengaruh secara signifikan antara X dan Y )

- Taraf signifikansi = 5% (0,05)



- Daerah kritis : Tolak  $H_0$  jika  $p\text{-value} < 0,05$
- Keputusan  
 Nilai  $p\text{-value}$   $X_1$  sebesar  $0,000 < 0,05$  : Tolak  $H_0$   
 Nilai  $p\text{-value}$   $X_2$  sebesar  $0,784 > 0,05$  : Gagal tolak  $H_0$
- Kesimpulan  
 Variabel  $X_1$  berpengaruh signifikan terhadap variabel  $Y$  sedangkan variabel  $X_2$  tidak berpengaruh signifikan terhadap variabel  $Y$ .
- R – square  
 Nilai R-square sebesar 0,751 dapat disimpulkan bahwa indeks pembangunan masyarakat (IPM) dan tingkat pengangguran berpengaruh terhadap kemiskinan sebesar 75,1% sisanya 24,9% dipengaruhi oleh faktor lain yang tidak masuk dalam model penelitian ini.

#### **4. Kesimpulan**

Variabel  $X_1$  berpengaruh signifikan terhadap variabel  $Y$  sedangkan variabel  $X_2$  tidak berpengaruh signifikan terhadap variabel  $Y$ . Nilai R-square sebesar 0,751 dapat disimpulkan bahwa indeks pembangunan masyarakat (IPM) dan tingkat pengangguran berpengaruh terhadap kemiskinan sebesar 75,1% sisanya 24,9% dipengaruhi oleh faktor lain yang tidak masuk dalam model penelitian ini.