
and
Inflation
with simple Linear regression

Background

- Nowadays all price consumer goods is climb higher and higher due to shortage of energy
- We need to give priority to bought most basic need because most of price rise about 10-50%.

Objective

predict formulate how Consumer Price Index and Inflation related

Data sources

https://www.kaggle.com/code/tunguz/us-consumer-price-index-and-inflation-eda-starter/comments

```
df=pd.read_csv('cpiai_csv.csv')
df.head()
```

	Date	Index	Inflation
0	1913-01-01	9.8	NaN
1	1913-02-01	9.8	0.00
2	1913-03-01	9.8	0.00
3	1913-04-01	9.8	0.00
4	1913-05-01	9.7	-1.02

Tabel show Inflation
and
US Consumer Index
from 1913 - 2014

Simple Linear Regression

```
#splitting training dan test set
from sklearn.model_selection import train_test_split #asumsi harus dikasih 2 nilai features dan target jadi dari dataset harus dipisahin

#definisikan features  : pisahkan features dan target
feature = df.drop(columns='Inflation')
target = df[['Inflation']]

#split data proporsi (80 training) :(20 test)  masingmasing target dan features : trained dan test
feature_df_train, feature_df_test, target_df_train, target_df_test = train_test_split(feature, target, test_size=0.20, random_state=42)
```

```
[ ] #definisikan model kosongan nya /model mentah sebelum di trained
    from sklearn.linear_model import LinearRegression

    # define the model mentah yang belum ditrained, bebas nama apa aja
    simple_reg = LinearRegression()

    # train the model :setelah split data proporsi, perlu merubah bentuk data menjadi np array karena sklearn bekerja dengan numpy
    X_df_train = feature_df_train.to_numpy()
    y_df_train = target_df_train.to_numpy()

    #fit proses training
    #modelkosongan/mentah kita minta untuk belajar/mengamati pola hubungan antar x dan y
    simple_reg.fit(X_df_train, y_df_train)
```

Membuat trained model dan test data dengan memisahkan data 80: 20

Setelah trained model, didapatkan setiap kenaikan consumer index 1 point, berasosiasi dengan kenaikan inflasi sebesar 0.000063

```
# retrieve the coefficients hasi paramater yang didapatkan
# show as a nice dataframe

data = feature_df_train

model = simple_reg

coef_df = pd.DataFrame({
    'feature':['intercept'] + data.columns.tolist(),
    'coefficient':[model.intercept_[0]] + list(model.coef_[0])
})
coef_df
```

	feature	coefficient
0	intercept	0.259598
1	Index	0.000063

Model

Inflation=0.259+ 0.000063 Consumer Index Price

test QQ plot

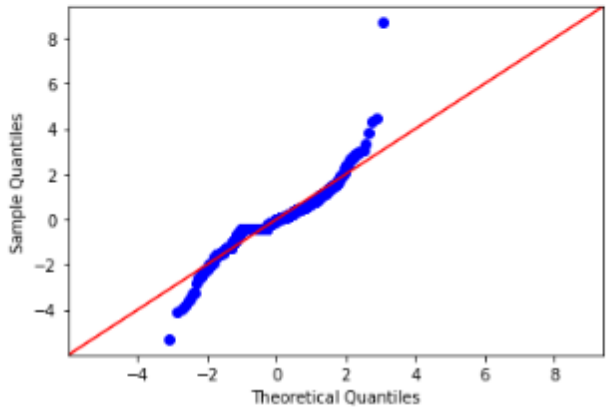
SEARCH STACK OVERFLOW

```
[ ]  from sklearn.preprocessing import StandardScaler


     std_resid = StandardScaler().fit_transform(residual.reshape(-1,1))
     std_resid = np.array([value for nested_array in std_resid for value in nested_array])

     import statsmodels.api as sm
     sm.qqplot(std_resid, line='45')
     plt.show()
```

```
/usr/local/lib/python3.7/dist-packages/statsmodels/tools/_testing.py:19: FutureWarning: pandas.util.testing is deprecated. Use
  import pandas.util.testing as tm
```



Residual berditribusi normal karena berhimpit titik biru dan titik merah (patokan normalitas)

conclusion

Prediksi model antara CPI dan index adalah sebagai berikut

Model

$$Inflation = 0.259 + 0.000063 \text{ Consumer Index Price}$$

Setelah menggunkan uji QQ plot, model bisa berdistribusi normal artinya bisa dibuatkan model

Thank you!

Do you have any follow-up questions for me?