

• Load the dataset: birthwt Risk Factors Associated with Low Infant Birth Weight

```
import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt
from sklearn.metrics import mean_squared_error, r2_score
```

```
bwt_df = pd.read_csv('/content/Bwt.csv', index_col=0)
bwt_df
```



	bwt	gestation	parity	age	height	weight	smoke
1	120	284	0	27	62	100	0
2	113	282	0	33	64	135	0
3	128	279	0	28	64	115	1
4	108	282	0	23	67	125	1
5	136	286	0	25	62	93	0
...
1170	113	275	1	27	60	100	0
1171	128	265	0	24	67	120	0
1172	130	291	0	30	65	150	1
1173	125	281	1	21	65	110	0
1174	117	297	0	38	65	129	0

1174 rows × 7 columns

1. Test the association of mother's (age) age and baby's birth weight (bwt) using the correlation test and linear regression.

2. Test the association of mother's weight (weight) and birth weight (bwt) using the correlation test and linear regression.

3. Produce two scatter plots of:

- (i) mother's age by baby's birth weight;
- (ii) mother's weight by baby's birth weight.

1. Testing the association of mother's (age) age and baby's birth weight (bwt) using the correlation test and linear regression and producing scatter plot of it.

```
#Finding correlation coefficient between Mother's age and Baby's birth weight

column_1 = bwt_df["age"]
column_2 = bwt_df["bwt"]
correlation = column_1.corr(column_2)

print("Correlation between Mother's Age and Baby's birth weight :",correlation)
```

Correlation between Mother's Age and Baby's birth weight : 0.0269829

```
#Finding Regression Coefficient

column_1 = bwt_df.age.values
column_2 = bwt_df.bwt.values

length = len(column_1)

column_1 = column_1.reshape(length, 1)
column_2 = column_2.reshape(length, 1)

regr = LinearRegression()
regr.fit(column_1, column_2)

coefficient = regr.coef_[0][0]
print("Regression Coefficient : ",coefficient)
```

```
print("Intercept : ",regr.intercept_[0])
print('Coefficient of determination(r2 score): %.2f'% r2_score(column_1,
```

```
Regression Coefficient : 0.08500766941582519
```

```
Intercept : 117.14790872185156
```

```
Coefficient of determination(r2 score): 0.26131
```

```
#Plotting scatter for Mother's Age vs Baby's birth weight
```

```
plt.scatter(column_1, column_2, color='black')
```

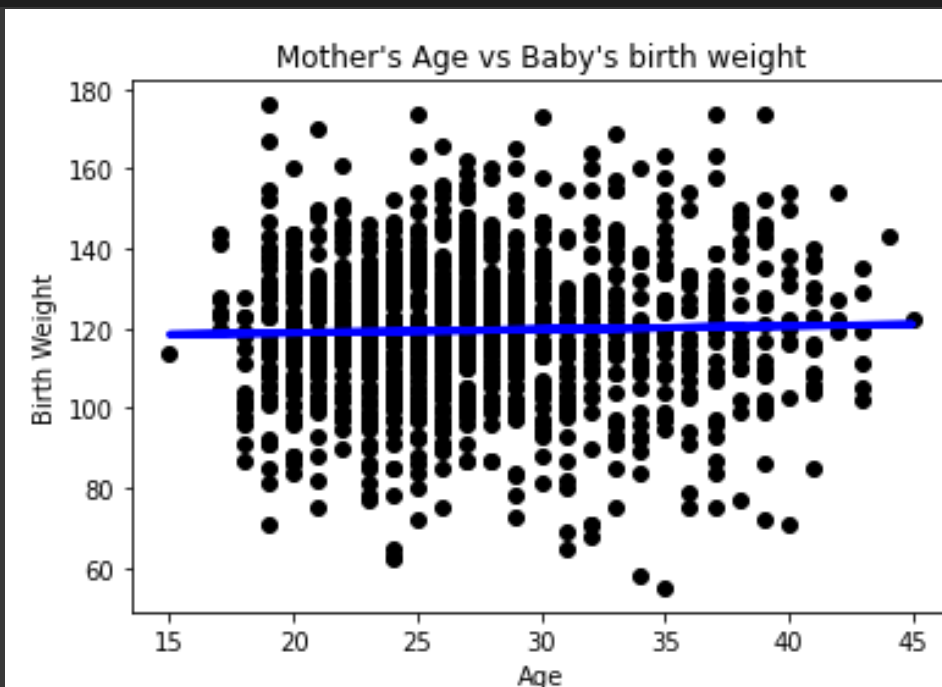
```
plt.plot(column_1, regr.predict(column_1), color='blue',linewidth=3)
```

```
plt.xlabel("Age")
```

```
plt.ylabel("Birth Weight")
```

```
plt.title("Mother's Age vs Baby's birth weight")
```

```
plt.show()
```



▼ Conclusion :

From the plot we can see almost flat regression line this means baby's weight is almost not related to age of mother.

2. Test the association of mother's weight (weight) and birth weight (bwt) using the correlation test and linear regression and producing its scatter plot.

```
#Finding correlation coefficient between Mother's weight and Baby's birth
```

```
column_3 = bwt_df["weight"]
column_4 = bwt_df["bwt"]
correlation = column_3.corr(column_4)

print("Mother's weight and baby's birth weight is correlated with each other")
```

Mother's weight and baby's birth weight is correlated with each other

```
#Finding Regression Coefficient
```

```
column_3 = bwt_df.weight.values
column_4 = bwt_df.bwt.values

length = len(column_1)

column_3 = column_3.reshape(length, 1)
column_4 = column_4.reshape(length, 1)

regr = LinearRegression()
regr.fit(column_3, column_4)

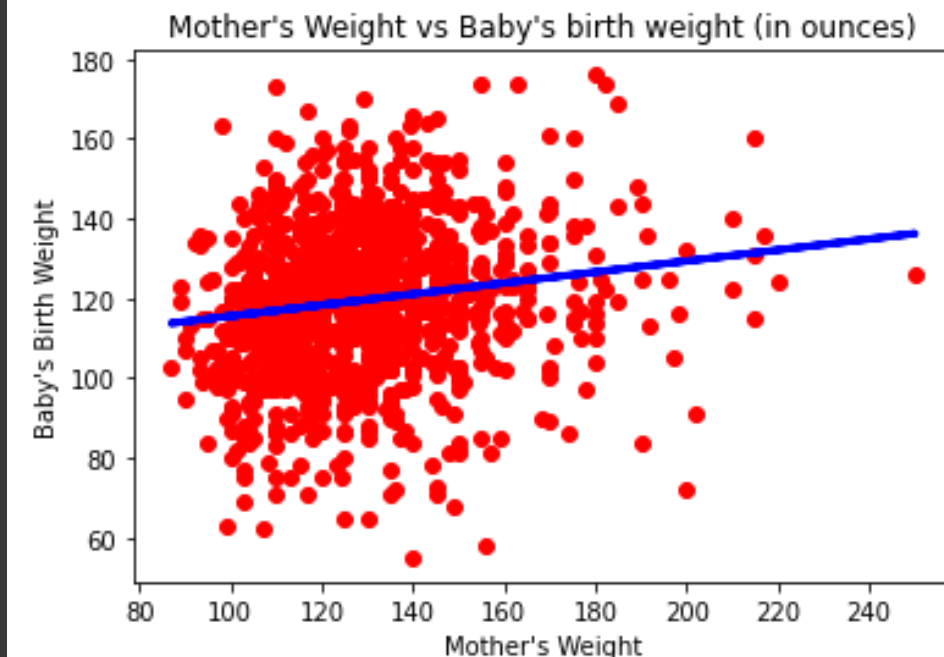
coefficient = regr.coef_[0][0]
print("Regression Coefficient : ",coefficient)
print("Intercept : ",regr.intercept_[0])

print('Coefficient of determination(r2 score): %.2f'% r2_score(column_3, column_4))
```

```
Regression Coefficient : 0.13783290697470632
Intercept : 101.75392786147968
Coefficient of determination(r2 score): -0.70
```

```
#Plotting scatter for Mother's Weight vs Baby's birth weight
```

```
plt.scatter(column_3, column_4, color='red')
plt.plot(column_3, regr.predict(column_3), color='blue', linewidth=3)
plt.xlabel("Mother's Weight")
plt.ylabel("Baby's Birth Weight")
plt.title("Mother's Weight vs Baby's birth weight (in ounces)")
plt.show()
```



Conclusion :

From the plot we can see a non-linear regression as it is following pattern but not in a linear manner. We can say that Mother's weight is slightly correlated with Baby's birth weight.

