In [24]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn import linear_model
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
```

In [16]:
```
%matplotlib inline
```

In [3]:
```
df1=pd.read_csv("C:\\Users\\Shay\\Documents\Yari\\Data analysis Python\\Data\\MCW.csv")
```

In [4]:
```
df2=pd.read_csv("C:\\Users\\Shay\\Documents\Yari\\Data analysis Python\\Data\\OW.csv")
```

In [14]:
```
dfm = pd.merge(df1,df2,on=['HDDSN','TESTCODEC','LHD','BAND'],  how='outer')
```

In [34]:
```
#dfm
```

In [17]:
```
dfm.to_csv('C:\\Users\\Shay\\Documents\\Yari\\Data Analysis Python\\Data\\dfm.csv')
```

In [4]:
```
MCW_OW_MERGE = pd.read_csv("C:\\Users\\Shay\\Documents\Yari\\Data analysis Python\\Data\\M
```

In [5]:
```
MCW_OW_MERGE
```

Out[5]:

| | Unnamed: 0 | HDDSN | PROCID | TESTCODEC | PFCODE | MFGID | HDDTRIAL | LHD | PHD | BAND | MCW_nm |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 0 | 2FA0491A | 6400 | PDQC3CX | 2525 | KQBT03 | AQ0N | 0 | 0 | 0 | 55.235076 |
| **1** | 1 | 2FA0491A | 6400 | PDQC3CX | 2525 | KQBT03 | AQ0N | 0 | 0 | 1 | 52.245708 |
| **2** | 2 | 2FA0491A | 6400 | PDQC3CX | 2525 | KQBT03 | AQ0N | 0 | 0 | 2 | 55.880085 |
| **3** | 3 | 2FA0491A | 6400 | PDQC3CX | 2525 | KQBT03 | AQ0N | 1 | 1 | 0 | 63.272877 |
| **4** | 4 | 2FA0491A | 6400 | PDQC3CX | 2525 | KQBT03 | AQ0N | 1 | 1 | 1 | 59.961005 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **283** | 283 | 2FA049GA | 6400 | PDQC3CX | 2521 | KQBT03 | AQ0N | 6 | 16 | 1 | 58.670988 |
| **284** | 284 | 2FA049GA | 6400 | PDQC3CX | 2521 | KQBT03 | AQ0N | 6 | 16 | 2 | 61.437083 |
| **285** | 285 | 2FA049GA | 6400 | PDQC3CX | 2521 | KQBT03 | AQ0N | 7 | 17 | 0 | 60.655630 |
| **286** | 286 | 2FA049GA | 6400 | PDQC3CX | 2521 | KQBT03 | AQ0N | 7 | 17 | 1 | 57.442990 |
| **287** | 287 | 2FA049GA | 6400 | PDQC3CX | 2521 | KQBT03 | AQ0N | 7 | 17 | 2 | 58.460120 |

288 rows × 16 columns

In [33]:
```
#MCW_OW_MERGE.dtypes
```

In [47]:
```
x = MCW_OW_MERGE['OW_PERP']
y = MCW_OW_MERGE['MCW_nm']
```

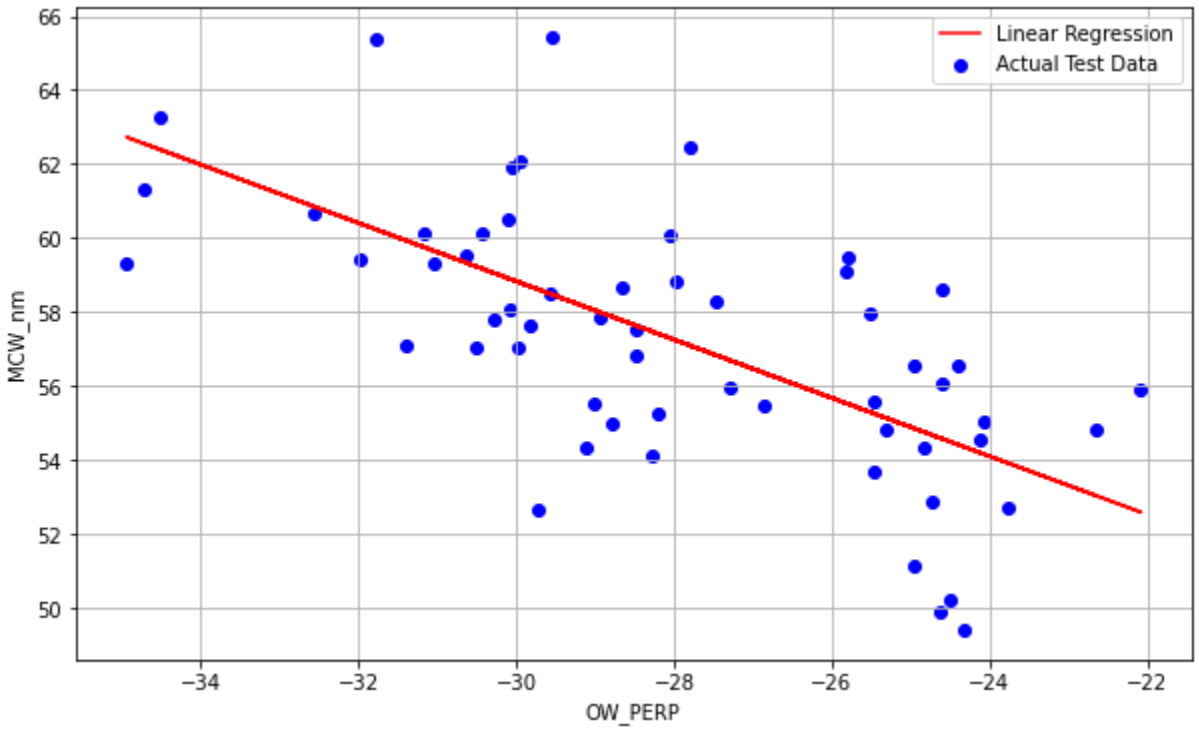
In [48]:
```
#Test tarin split for supervised training
x_train, x_test, y_train, y_test = train_test_split(x,y,test_size = 0.20)
```

In [49]:
```
#Here we create a linear model and train it
reg = LinearRegression()
reg.fit(x_train.values.reshape(-1,1), y_train.values)
```

Out[49]:
```
LinearRegression()
```

In [50]:
```
#Use model to predict o test data
prediction = reg.predict(x_test.values.reshape(-1,1))
```

In [51]:
```
plt.figure(figsize=(10,6))
plt.scatter(x_test,y_test,label = 'Actual Test Data', color='blue')
plt.plot(x_test,prediction, label = 'Linear Regression', color ='r')
plt.xlabel("OW_PERP")
plt.ylabel("MCW_nm")
plt.grid(True)
plt.legend()
plt.show()
```




In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: