

Quiz: OLS with Python

Due Feb 15 at 3:30pm**Points** 8**Questions** 6**Time Limit** None

Instructions

For this quiz, you will need to first follow the instructions in the document

[Fintech Fracassi Assignment OLS with Python.pdf](#)

(https://app.box.com/embed_widget/s/jqjg5n6vazxbez6m68oju7tjfw36nlh9?view=list&sort=name&direction=ASC&theme=dark).

The data can be found here: [Stock Return Data.csv](#)

(https://app.box.com/embed_widget/s/1tsxvs1cvnmdbpdqj6l1cpuy5hven9k7?view=list&sort=name&direction=ASC&theme=dark).

This quiz has no time limit, and it is an individual assignment. After you write your python code, answer the following questions.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	5 minutes	6 out of 8 *

* Some questions not yet graded

Score for this quiz: **6** out of 8 *

Submitted Feb 4 at 11:14pm

This attempt took 5 minutes.

Question 1

1 / 1 pts

How many observations are in the original database? [hint: use the function `df.info()`]

☐ 6,055

☐ 60,555

☒ 12,110

Correct!

☐ 1000

Question 2

1 / 1 pts

What is the average excess return in the original sample? [hint: use the function `df.describe()`]

☐ 0.036547

☐ 0.046547

☐ 0.056547

☒ 0.026547

Correct!

Question 3

1 / 1 pts

What is the standard deviation of the Agriculture Industry returns in the original sample? [hint: use the function `df.describe()`]

☒ 1.410530

☐ 0

☐ 2.523455

☐ 1.523455

Correct!

Question 4

2 / 2 pts

What is the mean square error of the OLS regression?

Correct!

- ☒ 1.0279229004214308
- ☐ 3
- ☐ 2.0279229004214308
- ☐ 0.0279229004214308

Question 5

1 / 1 pts

What is the Root Mean Square Error?

Correct!

- ☐ 2.0138653265702653
- ☒ 1.0138653265702653
- ☐ 2.0279229004214308
- ☐ 1.0279229004214308

Question 6

Not yet graded / 2 pts

Copy and paste the python code below.

Your Answer:

```
import pandas as pd
import numpy as np
from sklearn.datasets import make_classification
from sklearn.linear_model import LinearRegression
```

```
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import roc_curve
from sklearn.metrics import roc_auc_score
from sklearn.metrics import mean_squared_error
import math
from matplotlib import pyplot
import os
import statsmodels.api as sm
```

```
# In[26]:
```

```
df=pd.read_csv('Stock Return Data.csv')
df.columns
```

```
# In[27]:
```

```
df.head()
```

```
# In[28]:
```

```
df['Mkt_rf_lead'] = df['Mkt_rf'].shift(-1)
df['Mkt_rf_lead']
```

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

```
#df['Mkt_rf_lead'].dropna(how='any', inplace=True)
df.dropna(how='any', inplace=True)
df
```

```
# In[35]:
```

```
X = np.asarray(df.drop(['Date','Mkt_rf','Mkt_rf_lead'],axis=1))
y = np.asarray(df['Mkt_rf_lead'])
```

```
# Create linear regression object
reg = linear_model.LinearRegression()
# Train the model using the training sets
reg.fit(X, y)

# Make predictions using the testing set
pred_x = reg.predict(X)

# The mean absolute error

MSE = mean_squared_error(y, pred_x, squared=True)
RMSE = mean_squared_error(y, pred_x, squared=False)
print(MSE)
print(RMSE)

# In[37]:

dif=y-y_pred
dif_sq=dif*dif
n=len(y)
mse=np.sum(dif_sq)/n
mse2=mean_squared_error(y, y_pred)

rmse=math.sqrt(mse)
print("mse:", mse)
print("mse2:", mse2)
print("rmse:", rmse)

# In[38]:

df.describe()
```

Quiz Score: **6** out of 8