

# MA541\_part 7

August 8, 2021

```
[11]: import numpy as np
import pandas as pd
```

```
[4]: project_data = pd.read_csv('data.csv')
project_data.head()
```

```
[4]:   Close_ETF      oil      gold      JPM
0  97.349998  0.039242  0.004668  0.032258
1  97.750000  0.001953 -0.001366 -0.002948
2  99.160004 -0.031514 -0.007937  0.025724
3  99.650002  0.034552  0.014621  0.011819
4  99.260002  0.013619 -0.011419  0.000855
```

```
[5]: #Consider the entire Gold column as a random sample from the first population,
#and the entire Oil column as a random sample from the second population.
    →Assuming these two samples be
#drawn independently, form a hypothesis and test it to see if the Gold and Oil
    →have equal means in the
#significance level 0.05.
from scipy import stats
significance_level = 0.05
gold_update = project_data['gold'].tolist()
oil_update = project_data['oil'].tolist()
t_test, p_value= stats.ttest_ind(gold_update, oil_update)
print("The p_value is: ", p_value)
if p_value<significance_level:
    print("The test is failed to reject H0")
else:
    print("The test is reject H0")
```

The p\_value is: 0.6274695292874639

The test is reject H0

```
[35]: #Subtract the entire Gold column from the entire Oil column and generate a
    →sample of differences.
#Consider this sample as a random sample from the target population of
    →differences between Gold and Oil.
```

```

#Form a hypothesis and test it to see if the Gold and Oil have equal means in
↳the significance level 0.05.
from scipy import stats
import scipy.stats as st

difference_gold_oil = (project_data['gold'] - project_data['oil'])
#print("The difference of the gold and oil is:",difference_gold_oil)
#diff_gold_oil = difference_gold_oil.tolist()
#print(diff_gold_oil)

sample_100_gold_oil = pd.Series(difference_gold_oil.sample(n=100, replace=True))
st.norm.interval(alpha=0.95, loc=sample_100_gold_oil.mean(),
↳scale=sample_100_gold_oil.std())

print("The sample of the mean is:",sample_100_gold_oil.mean())
print()

mu_diff = 0
std_diff= sample_100_gold_oil.std()
n_diff = 100
mu_0_diff = 100
S_x_diff = std_diff/np.sqrt(n_diff)
print("The result is:",S_x_diff)

T_test_diff = (mu_diff - mu_0_diff)/S_x_diff
print("The value is:",T_test_diff)

pval_diff = stats.t.sf(np.abs(T_test_diff), n_diff-1)*2
print("The p value is:",pval_diff)

significance_level = 0.05

t_test, p_value= stats.ttest_ind(difference_gold_oil,sample_100_gold_oil)
print("The p_value is: ", p_value)
if pval_diff>significance_level:
    print("The test is failed to reject H0")
else:
    print("The test is reject H0")

```

The sample of the mean is: 0.0035872410099999984

The result is: 0.002307532419824566

The value is: -43336.33588021384

The p value is: 0.0

The p\_value is: 0.08119642986083625

The test is reject H0

```
[24]: #Consider the entire Gold column as a random sample from the first population,  
#and the entire Oil column as a random sample from the second population.  
#Assuming these two samples be drawn independently, form a hypothesis and  
#test it to see if the Gold and Oil have equal standard deviations in the  
→significance level 0.05.  
  
import scipy  
significance_level = 0.05  
  
gold_new_update = project_data['gold']  
oil_new_update = project_data['oil']  
sample_100_gold = pd.Series(gold_new_update.sample(n=100, replace=True))  
st.norm.interval(alpha=0.95, loc=sample_100_gold.mean(), scale=sample_100_gold.  
→std())  
sample_100_oil = pd.Series(oil_new_update.sample(n=100, replace=True))  
st.norm.interval(alpha=0.95, loc=sample_100_oil.mean(), scale=sample_100_oil.  
→std())  
  
print("The gold sample of the mean is:",sample_100_gold.mean())  
print()  
  
print("The oil sample of the mean is:",sample_100_oil.mean())  
print()  
  
f = np.var(project_data['gold']) / np.var(project_data['oil'])  
n_oil = 50  
n_gold = 50  
result = 1-scipy.stats.f.cdf(f, n_oil - 1, n_gold -1)  
print("The result is: ",result)  
  
if p_value>significance_level:  
    print("The test is failed to reject H0")  
else:  
    print("The test is reject H0")
```

The gold sample of the mean is: 0.0005041307799999996

The oil sample of the mean is: 0.0023146803099999996

The result is: 0.999987979230873

The test is failed to reject H0

[ ]: