Statistics Methods in Finance
Homework 2

0786009 廖家鴻
2020/10/7

Outline (HW2 questions)

Use the Dow John index you collected in the previous week.

Collect three additional stock prices that is in the Down John, of the same period.

1.(40%) For each firm:

- Estimate it's daily mean of the return.
- Do a hypothesis test with null be zero mean, assuming we know the variance.
- Do a hypothesis test with null be zero mean, assuming we don't know the variance.
- Construct it's 95% confidence interval.

2.(40%) For each pair of firms:

- Do a hypothesis test with null be no differences in mean, assuming we know all the variances.
- Do a hypothesis test with null be no differences in mean, assuming we don't know any of the variances.

3.(10%) Calculate the covariance and correlation matrix of these three firms.

4.(10%) How do you test whether these correlations are significantly differ from 0?

For each firm

1-1. Estimate mean of the daily return

In the following website, I choose AAPL, PG, and WMT as the three firms in the historical components of the Dow Jones Industrial Average during 2000/1/1 to 2006/12/31

https://www.wikiwand.com/en/Historical_components_of_the_Dow_Jones_Industrial_Average

From the following screen shots, the mean daily returns are 0.000638, -0.000038, and -0.000038 for AAPL, PG, and WMT, respectively.

```
In [3]: ret_aapl.head()
Out[3]:
datadate
20000104   -0.084310
20000105    0.014634
20000106   -0.086538
20000107    0.047368
20000110   -0.017588
Name: AAPL, dtype: float64
```

```
In [4]: ret_aapl.describe()
Out[4]:
count    1758.000000
mean        0.000638
std         0.037121
min        -0.518692
25%        -0.017231
50%         0.000000
75%         0.017895
max         0.136859
Name: AAPL, dtype: float64
```

```
In [5]: ret_pg.head()
Out[5]:
datadate
20000104   -0.019242
20000105   -0.019025
20000106    0.046061
20000107    0.079954
20000110   -0.004828
Name: PG, dtype: float64
```

```
In [6]: ret_pg.describe()
Out[6]:
count    1758.000000
mean       -0.000038
std         0.020385
min        -0.496325
25%        -0.006549
50%         0.000138
75%         0.007147
max         0.095238
Name: PG, dtype: float64
```

```
In [7]: ret_wmt.head()
Out[7]:
datadate
20000104   -0.039252
20000105   -0.019455
20000106    0.010913
20000107    0.075564
20000110   -0.018248
Name: WMT, dtype: float64
```

```
In [8]: ret_wmt.describe()
Out[8]:
count    1758.000000
mean       -0.000038
std         0.018606
min        -0.090517
25%        -0.009829
50%        -0.000548
75%         0.008822
max         0.094340
Name: WMT, dtype: float64
```

For each firm

1-2. Hypothesis test with null be zero mean (known variance)

Test procedure:

1) H_{01}: the mean daily return of AAPL is zero

   H_{02}: the mean daily return of PG is zero

   H_{03}: the mean daily return of WMT is zero

2) Assume variance is known => Z-test

3) $\alpha = 0.05$ for all three firms

4) p_values: p_AAPL = 0.4711

   p_PG = 0.9383

   p_WMT = 0.9313

5) According to Z-test, We accept (do not reject) H_{01}, H_{02}, and H_{03}.

```
'''Problem1-2: One sample Z-test'''
def One_sp_ztest(ret):
    ztest, pval = stests.ztest(ret, x2=None, value=0,
                               alternative='two-sided')
    if pval<0.05:
        print("reject null hypothesis")
    else:
        print("accept null hypothesis")
    return pval

pval_z_aapl = One_sp_ztest(ret_aapl)
pval_z_pg = One_sp_ztest(ret_pg)
pval_z_wmt = One_sp_ztest(ret_wmt)
```

pval_z_aapl	float64	1	0.4711116852208511
pval_z_pg	float64	1	0.9383416747753637
pval_z_wmt	float64	1	0.9313206152247298

For each firm

1-3. Hypothesis test with null be zero mean (unknown variance)

Test procedure:

1) H_{01}: the mean daily return of AAPL is zero

   H_{02}: the mean daily return of PG is zero

   H_{03}: the mean daily return of WMT is zero

2) Assume variance is unknown => T-test

3) $\alpha = 0.05$ for all three firms

4) p_values: p_AAPL = 0.4712

   p_PG = 0.9383

   p_WMT = 0.9313

5) According to T-test, We accept (do not reject)

   H_{01}, H_{02}, and H_{03}

```
'''Problem1-3: One sample t-test'''
def One_sp_ttest(ret):
    tset, pval = ttest_1samp(ret, 0)
    print('p-values', pval)
    if pval < 0.05:    # alpha value is 0.05
        print(" reject null hypothesis")
    else:
        print("accept null hypothesis")
    return pval

pval_t_aapl = One_sp_ttest(ret_aapl)
pval_t_pg = One_sp_ttest(ret_pg)
pval_t_wmt = One_sp_ttest(ret_wmt)
```

pval_t_aapl	float64	1	0.4712075549100313
pval_t_pg	float64	1	0.9383504822918602
pval_t_wmt	float64	1	0.9313304350076016

For each firm

1-4. Construct 95% confidence interval

```
'''Problem1-4: Confidence Interval'''
def conf_interval_t(data,alpha,mean,sem): #std是sem的根號n倍，即std=sem*np.sqrt(n)
    interval_t = stats.t.interval(alpha, df=(len(data)-1),loc=mean, scale=sem)
    return np.round(interval_t,8)

CIt95_aapl = conf_interval_t(ret_aapl, 0.95, ret_aapl.mean(), stats.sem(ret_aapl))
print(f"95% Confidence Interval (AAPL): {CIt95_aapl}")
CIt95_pg = conf_interval_t(ret_pg, 0.95, ret_pg.mean(), stats.sem(ret_pg))
print(f"95% Confidence Interval (PG): {CIt95_pg}")
CIt95_wmt = conf_interval_t(ret_wmt, 0.95, ret_wmt.mean(), stats.sem(ret_wmt))
print(f"95% Confidence Interval (WMT): {CIt95_wmt}")
```

Output results:

```
95% Confidence Interval (AAPL): [-0.00109838  0.00237444]
95% Confidence Interval (PG): [-0.00099119  0.00091597]
95% Confidence Interval (WMT): [-0.00090858  0.00083209]
```

For each pair of firms

2-1. Hypothesis test with null be no differences in mean (known variance)

Test procedure:

1) H_{01}: the mean daily returns of AAPL and PG are equal

   H_{02}: the mean daily return of PG and WMT are equal

   H_{03}: the mean daily return of WMT and AAPL are equal

2) Assume variance is known => Z-test

3) $\alpha = 0.05$ for all three firms

4) p_values: p_AAPL_PG = 0.4934

   p_PG_WMT = 0.9991

   p_WMT_AAPL = 0.4592

5) According to Z-test, We accept (do not reject)

   H_{01}, H_{02}, and H_{03}

```
'''Problem2-1: pair sample Z-test'''
# 檢定兩組樣本平均數是否相等=檢定兩組樣本平均數相減後是否等於0
def pair_sp_ztest(pair_diff):
    ztest, pval = stests.ztest(pair_diff, x2=None, value=0,
                               alternative='two-sided')
    print(float(pval))
    if pval<0.05:
        print("reject null hypothesis")
    else:
        print("accept null hypothesis")
    return pval

pval_pair_z_aapl_pg = pair_sp_ztest(ret_aapl-ret_pg)
pval_pair_z_pg_wmt = pair_sp_ztest(ret_pg-ret_wmt)
pval_pair_z_wmt_aapl = pair_sp_ztest(ret_wmt-ret_aapl)
```

pval_pair_z_aapl_pg	float64	1	0.493417151545646 25
pval_pair_z_pg_wmt	float64	1	0.999129569298987
pval_pair_z_wmt_aapl	float64	1	0.4591769883731964

For each pair of firms

2-2. Hypothesis test with null be no differences in mean (unknown variance)

Test procedure:

1) H_{01}: the mean daily returns of AAPL and PG are equal

   H_{02}: the mean daily return of PG and WMT are equal

   H_{03}: the mean daily return of WMT and AAPL are equal

2) Assume variance is unknown => T-test

3) $\alpha = 0.05$ for all three firms

4) p_values: p_AAPL_PG = 0.4934

   p_PG_WMT = 0.9991

   p_WMT_AAPL = 0.4592

5) According to T-test, We accept (do not reject)

   H_{01}, H_{02}, and H_{03}

```python
'''Problem2-2: Pair sample t-test'''
def pair_sp_ttest(ret1, ret2):
    ttest, pval_pair = stats.ttest_rel(ret1, ret2)
    print(pval_pair)
    if pval_pair<0.05:
        print("reject null hypothesis")
    else:
        print("accept null hypothesis")
    return pval_pair

pval_pair_t_aapl_pg = pair_sp_ttest(ret_aapl, ret_pg)
pval_pair_t_pg_wmt = pair_sp_ttest(ret_pg, ret_wmt)
pval_pair_t_wmt_aapl = pair_sp_ttest(ret_wmt, ret_aapl)
```

pval_pair_t_aapl_pg	float64	1	0.4935074882213947
pval_pair_t_pg_wmt	float64	1	0.9991296931421033
pval_pair_t_wmt_aapl	float64	1	0.4592758842132876

3. Covariance and correlation matrix of these three firms

```
'''Problem3: covariance and correlation matrix'''
df_3firms = pd.concat([ret_aapl, ret_pg, ret_wmt], axis=1)
df_3firms.cov()
df_3firms.corr(method='pearson')
```

Output covariance matrix:

```
In [21]: df_3firms.cov()
Out[21]:
          AAPL        PG       WMT
AAPL  0.001378  0.000041  0.000128
PG    0.000041  0.000416  0.000083
WMT   0.000128  0.000083  0.000346
```

Output correlation matrix:

```
In [23]: df_3firms.corr(method='pearson')
Out[23]:
          AAPL        PG       WMT
AAPL  1.000000  0.054598  0.185805
PG    0.054598  1.000000  0.219764
WMT   0.185805  0.219764  1.000000
```

4. Test whether these correlations are significantly differ from 0

```
'''Problem4: test whether these correlations are significantly differ from 0?'''
corr_aapl_pg, pval_aapl_pg = stats.pearsonr(ret_aapl, ret_pg)
corr_pg_wmt, pval_pg_wmt = stats.pearsonr(ret_wmt, ret_pg)
corr_wmt_aapl, pval_wmt_aapl = stats.pearsonr(ret_aapl, ret_wmt)
```

Test procedure for each pair of firms:

1) H_{01}: true correlation of AAPL and PG are equal to zero.

   H_{02}: true correlation of PG and WMT are equal to zero.

   H_{03}: true correlation of WMT and AAPL are equal to zero.

2) Use T-test to do significance testing.

3) $\alpha = 0.001$ for all three firms

4) p_values: p_AAPL_PG = 0.0221 > α

   p_PG_WMT = 1.14e-20 < α

   p_WMT_AAPL = 4.05e-15 < α

5) According to T-test, we accept H_{01},

   but we reject H_{02} and H_{03}.

Output correlations:

corr_aapl_pg	float64	1	0.05459849165875296
corr_pg_wmt	float64	1	0.21976350977398057
corr_wmt_aapl	float64	1	0.18580509237021273

Output p-values:

pval_aapl_pg	float64	1	0.02206118152740129
pval_pg_wmt	float64	1	1.1417867608969583e-20
pval_wmt_aapl	float64	1	4.046793715050026e-15