Project 1: Implementing Descriptive Statistics in Python and Excel

Panko Aliaksandr

January 21, 2018

Contents

1 Excel

1.1 Data Download

JP Morgan stock historical data was downloaded from Yahoo Finance.

1.2 Average Stock Value

Average stock value was calculated using Excel build-in function. Adjusted close price was used.

1.3 Stock volatility

Since it was not absolutely clear for me price or return volatility is supposed to be found, I decided to find both.

For volatility calculation standard build-in excel function is used.

1.4 Daily Stock Returns

Daily return can be calculated using the equation below:

$$daily\_return(in\%) = \frac{today\_price - yesterday\_price}{yesterday\_price} * 100\%$$

1.5 Scatter Plot

A scatter plot was built using the stock price data and standard excel scatter diagram. A trend line was added to show price evolution.

1.6 Sharpe ratio

The Sharpe ratio is the average return earned in excess of the risk-free rate per unit of volatility or total risk. It can be calculated using the formula below:

$$SharpeRatio = \frac{PortfolioAverateReturn - RiskFreeRate}{StDevPortfolioReturn}$$

1.7 Regression Analysis

Regression is a statistical measure used in finance, investing and other disciplines that attempts to determine the strength of the relationship between one dependent variable and a series of other changing variables (known as independent variables). Regression helps investment and financial managers to value assets and understand the relationships between variables, such as commodity prices and the stocks of businesses dealing in those commodities.

Using this regression with explained variable: JP Morgan stock (close price) and explanatory variable: S&P500, I find how S&P500 index movement influence(explains) movement of JP Morgan stock price.

Since there are only 2 variables, a type of the regression is linear.

More precisely: the LINEST function calculates the statistics for a line by using the "least squares" method to calculate a straight line that best fits your data, and then returns an array that describes the line.

According to the coefficient of determination, SP500 index explains only 23% of JPM stock movements.

Regression results:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.480201131
R Square	0.230593126
Adjusted R Square	0.217327491
Standard Error	2.253677599
Observations	60

ANOVA

	df	*SS*	*MS*
Regression	1	88.28803782	88.28803782
Residual	58	294.5856377	5.07906272
Total	59	382.8736756	

	Coefficients	*Standard Error*	*t Stat*
Intercept	-92.77109982	37.87991109	-2.449084413
X Variable 1	0.075088132	0.018009936	4.169261588

F	Significance F
17.38274219	0.000103256

P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
0.01736466	-168.5959801	-16.94621949	-168.5959801	-16.94621949
0.000103256	0.03903733	0.111138935	0.03903733	0.111138935

Regression Results (LINEST)		
Slope, intersect	0.075088132	-92.77109982
Errors	0.018009936	37.87991109
R2 , y-error	0.230593126	2.253677599
F-statistics, lvls of freedom	17.38274219	58
SS-reg, SS-resid	88.28803782	294.5856377
Regression Results (ToolPak)		

2 Python

2.1 Data Download

Required data has been downloaded from Yahoo Finance using next python function:

```
from pandas_datareader import data as pdr
pdr.get_data_yahoo()
```

2.2 Average Stock value

Average stock value can be calculated using pandas **DataFrame.mean()** function.

2.3 Stock volatility

Stock volatility can be calculated using pandas **DataFrame.std()** function.

2.4 Daily return

Daily returns can be calculated using pandas **pct_change(1)** function.

2.5 Regression in Python

To implement a regression in python **statsmodels** package can be used:

```
import statsmodels.api as sm
sm.add_constant()
model =sm.OLS()
model.fit()
```

This is ordinary least squares model.

2.6 Python results

```
Average Stock Value: 60.7189366167

Stock Volatility: 2.37396714637

Average Daily Stock return: 0.00233532731515

Daily Return Volatility: 0.00855633905235
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  Close   R-squared:                       0.231
Model:                            OLS   Adj. R-squared:                  0.217
Method:                 Least Squares   F-statistic:                     17.38
Date:                Sun, 21 Jan 2018   Prob (F-statistic):           0.000103
Time:                        16:05:05   Log-Likelihood:                -132.87
No. Observations:                  60   AIC:                             269.7
Df Residuals:                      58   BIC:                             273.9
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const        -92.7711     37.880     -2.449      0.017    -168.596     -16.946
Close          0.0751      0.018      4.169      0.000       0.039       0.111
==============================================================================
Omnibus:                        7.622   Durbin-Watson:                   0.095
Prob(Omnibus):                  0.022   Jarque-Bera (JB):                3.390
Skew:                           0.310   Prob(JB):                        0.184
Kurtosis:                       2.014   Cond. No.                     2.74e+05
==============================================================================
```

As we can see the results of all models are basically the same. The low p-values indicate to us that the constant, as well as the coefficient, are statistically different from zero.

An assumption of the linear regression is that the residuals are normally distributed. Skewness should be close to zero while kurtosis should be close to 3 when the residuals are normally distributed. In our case, the registered values are not far from the theoretical ones. Moreover, based on Jarque-Bera test (Prob JB $= 0.184 > 0.05$) we cannot reject the hypothesis that the residuals come from a normal distribution at 5 percent level of significance.