IMPORTANT: The answers are provided with the view to deepen your understanding of the topic and to stimulate more in depth discussions. They should not be shared with others or put on the web. The questions and answers may be revised in the next course delivery.

Answers to Tutorial 6

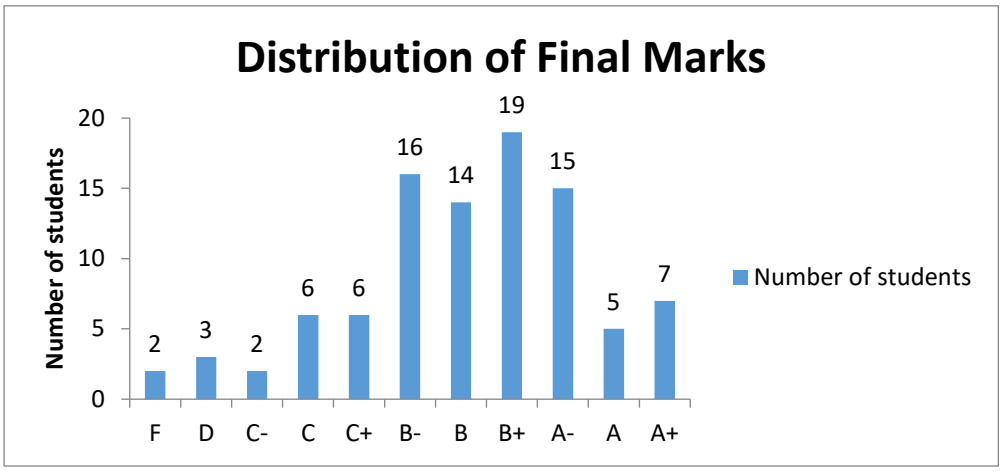
Qn 1

a)

Mean: 62.59663158
Median: 64.08
Mode: 71.88
Standard derivation(.P) : 13.49622408
Standard derivation(.S) : 13.56782259
Variance(.P):182.1480644
Variance(.S): 184.0858098

.p divides by n .S divides by $n-1$

b)



c) SKEW: -1.59404

You should use the raw data when computing the skew.

A rule of thumb is to compute the skew of the distribution. If the absolute value of the skew > $2\sqrt{6/n}$, then the skew is significant and the distribution is not normal.

$2\sqrt{6/95} = 0.50262469$

Hence the distribution is not normal according to the rule of thumb

d) There are various normality tests. See https://en.wikipedia.org/wiki/Normality_test for more information. Any example mentioned in the link is acceptable.

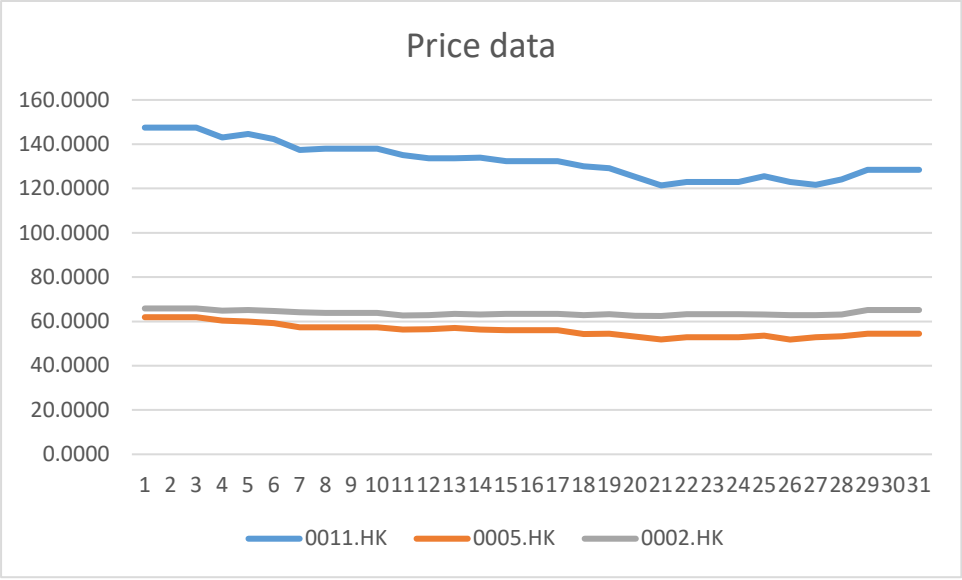
Qn 2

Different results would be obtained depending on the sample taken.

Additional note:
This illustrates the uncertainty due to sampling. Note that if we have the whole population, we would not have the whole truth and do not require statistics. In data engineering applications, often we do not have the time and resources to get the whole population.

Qn 3

a)



b) $r = 0.988724$

c) $r = 0.742161$

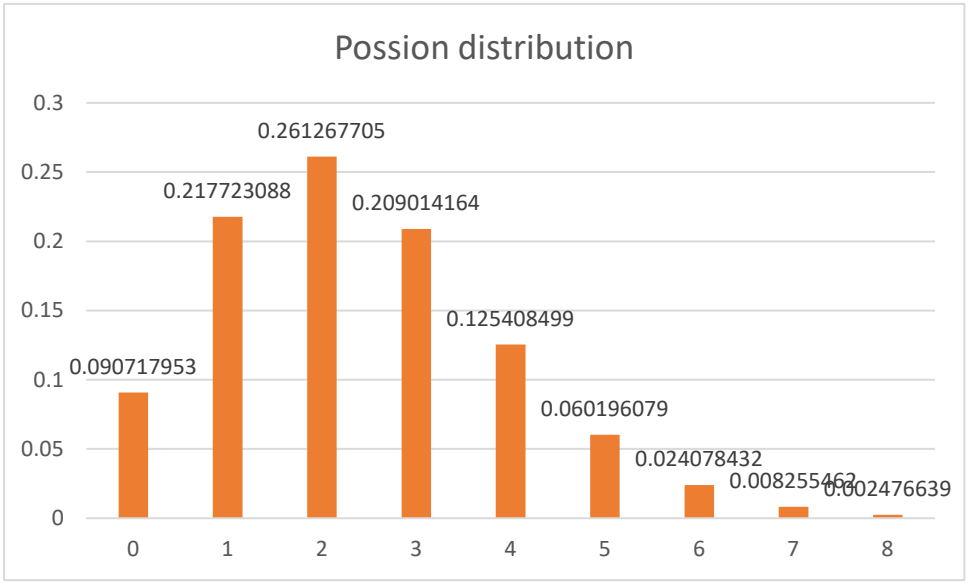
d) The stocks in both scenarios are positively correlated. The correlation between 0005 and 0011 are greater, perhaps they are both banks and they have links.

Assumption: the correlation is a linear relationship.

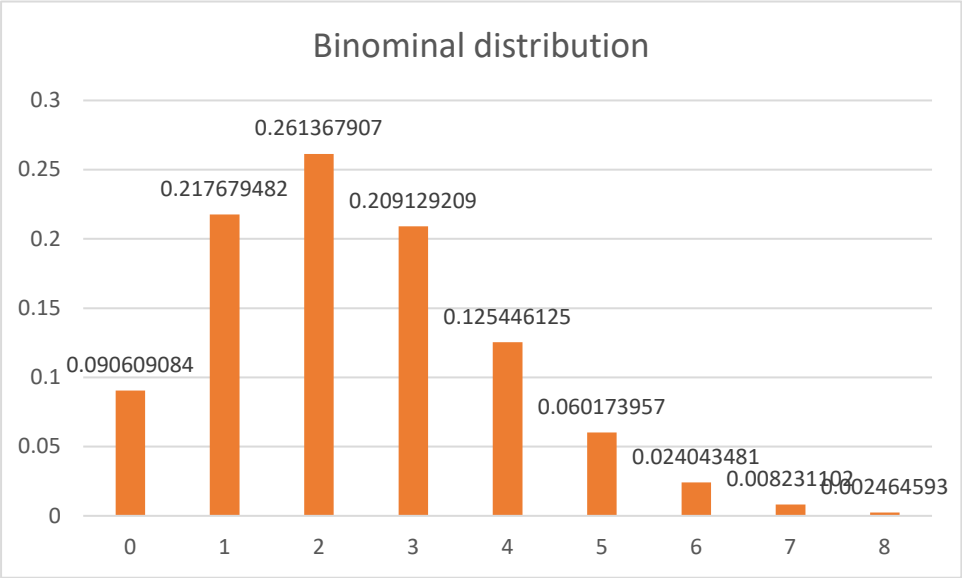
Note: this does not mean that the stocks are both straight lines, it means that when 0005 rises, 0011 also tend to rise and vice versa. The rise (and fall) follows a linear relationship

Qn 4

a)



b)



It agrees with the theoretical predictions. When n is large and p is small, the binomial distribution (n, p) can be approximated by the Poisson distribution.

The expression of the Poisson distribution is simpler and is thus easier to compute and manipulate in an analytical model.