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 10

<u>Qn 1</u>

$r = 0.988724 \qquad n = 31$

$$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = 35.55570417$$

d.f. = 29

p-value = $P\{|T| > 35.55570417\} < 2P\{T > 3.659\} = 2(0.0005) = 0.001$

The trends of the two stocks are related at a 0.1% level of significance.

Note that the use of p-value is more flexible. We can present the data without first specifying the level of significance. Then we can decide to reject or accept the hypothesis later by supplying a level of significance.

We have assumed that

1. The joint distribution of the two stocks is a bi-variate normal distribution.

2. The random variables X and Y representing the two stocks have the following statistical linear relationship

   $E(Y|X = x) = \alpha + \beta x$

Note that it assumes that the price of both stocks has a normal distribution and the relationship between the stocks have an approximate linear trend, i.e., if one stock goes up, the other goes up and vice versa in an approximately linear way, or if one stock goes up, the other goes down and vice versa in an approximately linear way.

Do you think the assumptions are realistic? Discuss.

<u>Qn 2</u>

a)

$$SS = \sum_{i=1}^{n} (y_i - A - Bx_i)^2$$

$$\frac{\partial SS}{\partial A} = -2\sum_{i=1}^{n} (y_i - A - Bx_i) = 0$$

$$\frac{\partial SS}{\partial B} = -2\sum_{i=1}^{n} x_i(y_i - A - Bx_i) = 0$$

which are

$$\sum_{i=1}^{n} y_i - nA - B\sum_{i=1}^{n} x_i = 0$$

$$\sum_{i=1}^{n} x_i y_i - A\sum_{i=1}^{n} x_i - B\sum_{i=1}^{n} {x_i}^2 = 0$$

Putting in the values

$$12 - 3A - 6B = 0$$

$$27.5 - 6A - 14B = 0$$

Solving, $(A, B) = (0.5, 1.75)$

b) Excel's SLOPE function returns $B = 1.75$. INERCEPT function returns $A = 0.5$

c)

$\bar{x} = 2 \quad \bar{y} = 4$

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	2	-1	-2	1	4	2
2	4.5	0	0.5	0	0.25	0
3	5.5	1	1.5	1	2.25	1.5

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{3.5}{\sqrt{2}\sqrt{6.5}} = 0.970725343$$

d) Excel's PEARSON function returns $r = 0.970725$

e)

$$s_{YY} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = 6.5$$

$$SS_R = \sum_{i=1}^{n}(Y_i - A - Bx_i)^2$$

$$= (2 - 0.5 - 1.75)^2 + \left(4.5 - 0.5 - 1.75(2)\right)^2 + \left(5.5 - 0.5 - 1.75(3)\right)^2$$

= 0.375

$$R^2 = \frac{S_{YY} - SS_R}{S_{YY}} = 0.942307692$$

$$|r| = \sqrt{R^2} = 0.970725343$$

The coefficient of determination R^2 represents the proportion of the variation in the response variable explained by the different input values. As R^2 is close to 1 indicates that most of the variation of the response data is explained by the different input values.

Qn 3

a) $0.5 + 1.75(4) = 7.5$

b) i) $\frac{4.5+5.5}{2} = 5$

ii) $\frac{2+4.5+5.5}{3} = 4$

c) $(0.7)(5.5) + (0.7)(0.3)(4.5) + (0.7)(0.3)^2(2) = 4.921$

Qn 4

a) $r \approx 0.24$

b) $r_s = 1$

c) The Pearson's coefficient show that there is not a linear relationship.

The Spearman's coefficient reveals a perfect monotonically increasing relationship.