

UECS3213 / UECS3453 DATA MINING

SESSION: January 2019

TUTORIAL 4

Chapter 4 - Regression Analysis

1. The time x in years that an employee spent at a company and the employee's hourly pay, y , for 5 employees are listed in the table below.

x	y
5	25
3	20
4	21
10	35
15	38

- a) Calculate and interpret the correlation coefficient r . Include a plot of the data in your discussion.

Hint:
$$r = \frac{n \sum(xy) - (\sum x)(\sum y)}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}}$$

- b) Find the equation of the *least square regression line* for the abovementioned relationship.
c) Use the equations in part (b) to predict the hourly pay rate of an employee who has worked for 20 years.

2. The table below shows the number of absences, x , in a Calculus course and the final exam grade, y , for 7 students.

a) Find the correlation coefficient, r and interpret your result.

x	1	0	2	6	4	3	3
y	95	90	90	55	70	80	85

b) Find the equation of the *least square regression line* for the abovementioned relationship.

c) Use the equations to part (b) to predict the test score for a student with 5 absences.

3. The table below shows the height, x , in inches and the pulse rate, y , per minute, for 9 people. Find the correlation coefficient, r and interpret your result.

x	68	72	65	70	62	75	78	64	68
y	90	85	88	100	105	98	70	65	72

4. Consider the following set of points: $\{(-2, -1), (1, 1), (3, 2)\}$

a) Find the *least square regression line* for the given data points.

b) Plot the given points and the *regression line* in the same rectangular system of axes.

5. Given the following data: $\{(-1, 0), (0, 2), (1, 4), (2, 5)\}$

a) Find the *least square regression line* for the following set of data

b) Plot the given points and the regression line in the same rectangular system of axes.

6. The values of x and their corresponding values of y are shown in the table below

x	0	1	2	3	4
y	2	3	5	4	6

a) Find the *least square regression line*, $y = a x + b$.

b) Estimate the value of y when $x = 10$.

7. The sales of a company (in million dollars) for each year are shown in the table below.

x (year)	2005	2006	2007	2008	2009
y (sales)	12	19	29	37	45

a) Find the *least square regression line* $y = a x + b$.

b) Use the least squares regression line as a model to estimate the sales of the company in 2012.

The End