Fundamentals of Data Analysis: Assignment #3

Deadline: 10/27/2003 (Monday)

Please post to the mailbox next to the IS management office (2nd floor of IS building)

1. Assume that you are thinking of taking a job at a certain company. You want to know the relationship between "the number of years with the company (*X*)" and "yearly salary (*Y*)", and, ask seven people for these two pieces of information. Answer the following questions.

Number of Years (*X*)	4	1	3	7	6	5	4
Salary (Million Yen) (*Y*)	15	12	14	22	20	18	12

a. Calculate means and standard deviations of *X* and *Y.*
b. Calculate covariance and the correlation coefficient between *X* and *Y*.
c. Find the linear regression equation for predicting *Y* from *X*.
d. Find the linear regression equation for predicting *X* from *Y*.
e. Draw a scatter plot of the data, and draw the lines of equations found in c. and d.
f. Check that the coordinates of intersection of two lines are given by $(\overline{X}, \overline{Y})$ on the plot, where $\overline{X}$ and $\overline{Y}$ are means of *X* and *Y*, respectively. Moreover, prove this relation mathematically.

2. Prove that mean squared error D^2 is given by the difference between the variance of data (S_Y^2) and the variance of predicted data ($S_{\hat{Y}}^2$).

3. Prove that if events A and B are exclusive and $P(A) > 0$, $P(B) > 0$, then *A* and *B* are NOT independent.

4. Calculate the probability that when you throw a dice 12 times, you observe each number of spots (i.e., 1 to 6) twice.

5. Please feel free to write your comments and requests on this lecture (if any).