

Tutorial: Regression

1. Some data appeared in a newspaper a few years ago showing the change in the use of Cable TV in households in China.

Year (x)	2001	2002	2003	2004
Households with cable/millions(y)	68	72	80	83

- a. Draw a scatter diagram for these data and comment on any noticeable pattern.
 - b. Obtain by hand, the equation of the least squares regression line of y on x.
 - c. Interpret the regression coefficients.
 - d. Plot the regression line on your scatter diagram and comment about its suitability.
 - e. What is the expected Cable usage in 2008? Is this plausible?
2. An Egyptian village was used as the site of a study of nutrition in developing countries. The data were obtained by measuring the heights (cm) of all 161 children in the village each month over several years. The data shows the mean heights for each age.

age	height (cm)
18	76.1
19	77
20	78.1
21	78.2
22	78.8
23	79.7
24	79.9
25	81.1
26	81.2
27	81.8
28	82.8
29	83.5

- a. Obtain by hand, the equation of the least squares regression line.
 - b. Plot the regression line and use it to determine the average height for an average 27 year old.
 - c. Calculate the correlation coefficient for this data. What does it tell you about this data set?
3. It is natural to expect that the larger the house, the higher the price. That is, we expect price and size of the house to be positively correlated. Ten houses were randomly selected among A local Danish newspaper ads for houses. The relationship between area and price is vaguely *suggesting* that it is not only the size of a house that determines the price. Your task is to investigate if there is any evidence to this suggestion.

size (m ²)	price 1000s DKK
104	1128
96	926
121	1284
145	701

62	712
96	851
149	976
196	1214
121	888
125	1272

4. Crickets make their chirping sounds by rapidly sliding one wing over the other. The faster they move their wings, the higher the chirping sound that is produced. Scientists have noticed that crickets move their wings faster in warm temperatures than in cold temperatures. Therefore, by listening to the pitch of the chirp of crickets, it is possible to tell the temperature of the air. The table below gives the recorded pitch (in vibrations per second) of a cricket chirping recorded at 15 different temperatures.

Chirps per second	20	16	20	18	17	16	15	17	15	16	15	17	16	17	14
Temperature (°C)	89	72	93	84	81	75	70	82	69	83	80	83	81	84	76

Write a Java Program to compute the least squares regression line for an unknown quantity of data (n).

Also use the program to compute the correlation coefficient (and if you are really keen and able, get your program to plot the data and the line on the same graph).