STATISTICS

Instructor: Associated Prof. Dr. Doğan Nadi LEBLEBİCİ

Source: Kaplan, Robert M. <u>Basic Statistics for the Behavioral Sciences</u>, Allyn and Bacon, Inc., Boston, 1987

© Copyright 2005, Doğan N. LEBLEBİCİ

1

SIMPLE REGRESSION ANALYSIS

© Copyright 2005, Doğan N. LEBLEBİCİ

Regression is the determination of a statistical relationship between two or more variables. In simple regression, we have only two variables, one variable (defined as independent) is the cause of the behaviour of another one (defined as dependent variable). Regression can only interpret what exists physically i.e., there must be a physical way in which independent variable X can affect dependent variable Y. The basic relationship between X and Y is given by

$$\hat{Y} = a + bX$$

where the symbol Y denotes the estimated value of Y for a given value of X. This equation is known as the regression equation of Y on X (also represents the regression line of Y on X when drawn on a graph) which means that each unit change in X produces a change of b in Y, which is positive for direct and negative for inverse relationships.

Then generally used method to find the 'best' fit that a straight line of this kind can give is the least-square method. To use it efficiently, we first determine a and b with the following formula.

$$b = \frac{N(\sum XY) - (\sum X).(\sum Y)}{N\sum X^2 - (\sum X)^2}$$
$$a = \overline{Y} - b\overline{X}$$

Sample: Different values of X and Y variables are given as follows:

X	Y	ХҮ	Y ²	X ²
23	30	690	900	529
45	47	2115	2209	2025
12	17	204	289	144
14	19	266	361	196
10	15	150	225	100
20	26	520	676	400
32	37	1184	1369	1024
25	32	800	1024	625
27	34	918	1156	729
14	45	630	2025	196
$\sum X = 222$	$\sum Y = 302$	$\sum XY = 7477$	$\sum Y^2 = 10234$	$\sum X^2 = 5968$

When we used the formula, we found a=13.77 and b=0.74. Thus, the regression formula will be

$$\hat{Y} = a + bX$$
 $\hat{Y} = 13.77 + 0.74X$

© Copyright 2005, Doğan N. LEBLEBİCİ

Regression cooefficent will be

$$r = \frac{b\sqrt{X_i^2}}{\sqrt{Y_i^2}} = \frac{0.74\sqrt{5968}}{\sqrt{10234}} = 0,56$$

But this regression coefficient is not statistically significant. Since N=9, degree of freedom (d.f.) will be 8 (N-1) and the value in t table with d.f.=8 at 95% importance level is 2,306 and our value is below this value. So, it is not statistically significant.