BASIC STATISTICS IN MEDICAL PRACTICE

Pages with reference to book, From 168 To 170

Syed Ejaz Alam (PMRC Research Centre, Jinnah Postgraduate Medical Centre, Karachi.)

Describing Quantitative relationship

Scientific studies often require a description of the relationship between two variables. Usually in such circumstances we think of one variable as being influenced by the other. It has become conventional to denote the dependent variable, i.e. the one being influenced, by "Y" and independent variable by "X". We are interested in describing the association between X & Y. To do this we have to measure jointly X and Y on a series of subjects.[1]. The simplest way of describing the relationship between X and Y is by a graph called a scatter diagram. To construct a scatter diagram, the level of Y is plotted against the level of X for each subject. The resulting scattering of points indicates how Y varies with differing levels of X. Although the scatter diagram is very useful for gaining a visual impression of the relationship, a more quantitative description ‘is often needed. Two kinds of statistical techniques are used to further specify the relationship between X and Y:

1. Regression
2. Correlation

The Regression Equation:

The regression approach is appropriate when our main purpose is to develop a predictive model i.e. a device that will enable us to predict Y against a given specified level of X. (See example).

The regression equation has the form: Y a + b X.

Where a = the intercept, i.e. the value of Y when X is zero

b = the slope, i.e. the change in Y resulting from a change in X of one unit. The constant a and-b are found by the Least Square procedure[2].

The Correlation Coefficient

The correlation coefficient, usually denoted by r, is an index of the extent to which two variables are associated. It can take on values between + 1.0 and -1.0, depending out the strength of the association. A correlation coefficient of zero indicates that the two variables are not related.

The following can serve as a general guide to interpreting the magnitude of the correlation coefficient:

Degree of association	
0.8 to 1.0	Strong
0.5 to 0.8	Moderate
0.2 to 0.5	Weak
0 to 0.2	Negligible

Example

Making the Scatter diagram.(Figure 1)

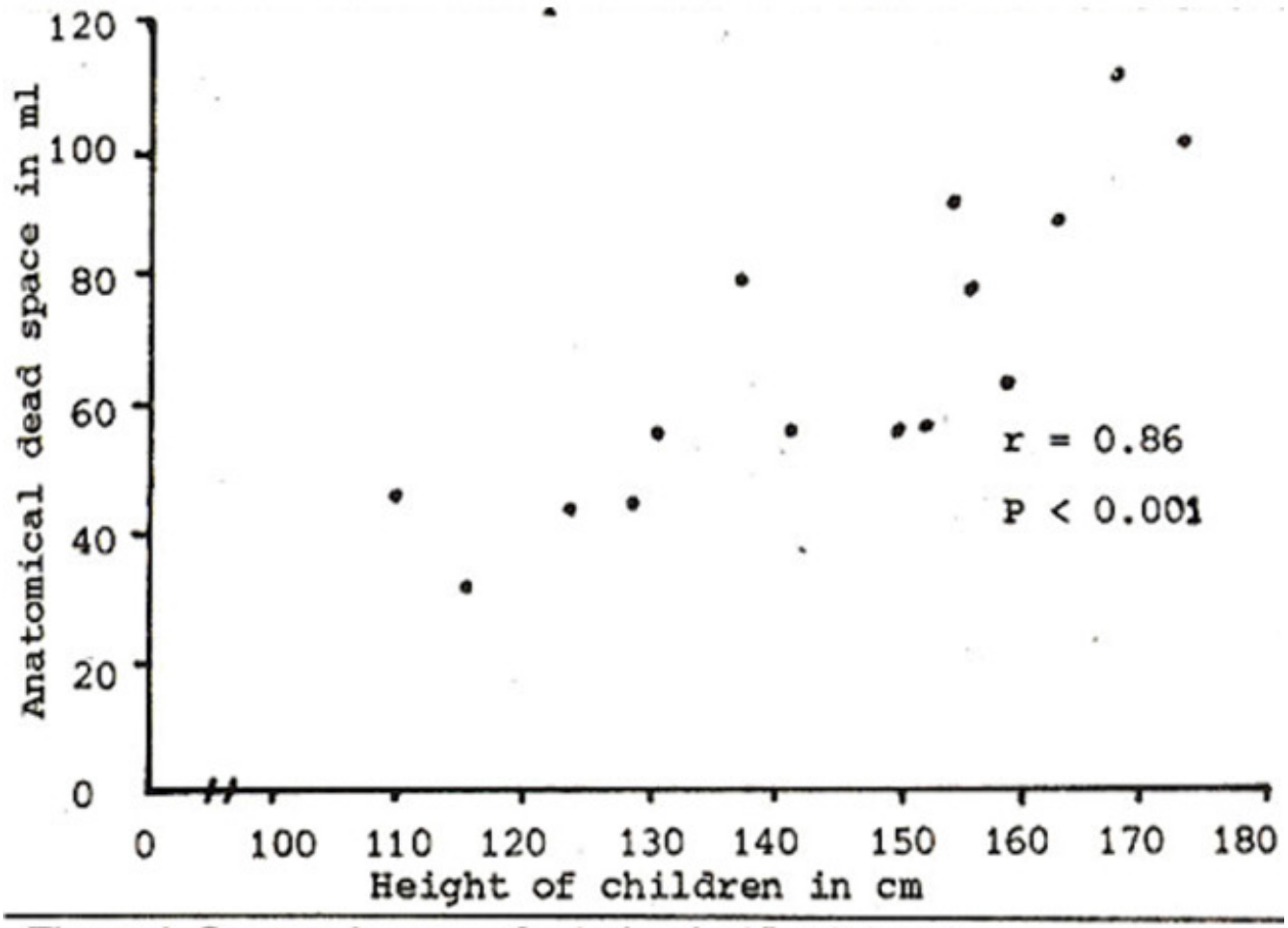

Figure 1. Scatter dagram of relation in 15 children between hight and pulmonary anatomical deal space.

to show the heights and pulmonary anatomical dead spaces in the 15 children. Dr. Green set out the figures as in column 1,2,3. It is helpfull to arrange the observations[3], as he has done in serial order on the independent variable when one of the two variables is clearly identified as independent. The corresponding figures for the dependent variable can be examined in relation to the increasing series for the independent variable. In this way we get the same picture, but in numerical form, as appears in the scatter diagram. The calculation of the correlation coefficient is as follows. With X representing the values of the independent variable (in this case height) and Y representing the value of the dependent variable (in this case anatomical dead space).

The Correlation coefficient of 0.846 indicates a strong positive correlation between size of pulmonary anatomical dead space and height of child. However, to test the deviation of r from 0, or nil correlation, it is better to use the t test in the following calculation:

The table is entered at n-2 degrees of freedom. For example, the correlation coefficient for Dr. Green's figures was 0.846. The number of pairs of observations was 15. Applying the above formula,

The formula to be used is:

$$r = \frac{\Sigma XY - \dfrac{(\Sigma X)\ (\Sigma Y)}{n}}{\sqrt{[\Sigma X^2 - (\Sigma X)^2/n]\,[\Sigma Y^2 - (\Sigma Y)^2/n]}}$$

$$r = \frac{150605 - (2169)\,(1004)\,/\,15}{\sqrt{[318889 - (2169)^2/15]\,[75030 - (1004)^2/15]}}$$

$$= \frac{5426.6}{\sqrt{5251.6 \times 7828.9}}$$

$$r = 0.846$$

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

$$t = 0.846 \times \sqrt{\frac{15-2}{1-0.846^2}} = 5.72$$

we have Entering the t table3 at 15-2 = 13 degrees of freedom we find that, at t = 5.72, p <0.001. So the correlation coefficient maybe regarded as highly significant.

The Regression Equation

Y = a + bX

With this equation we can find a series of values of Y, the dependent variable, that corresponds to each of, a series of values of X, the independent variable. The letters a & b have to be calculated from the data. The letter 'a' signifies the distance above the base line at which the regression line cuts the vertical Y-axis the letter b (the regression coefficient) signifies the amount by which a change in X must be multiplied to give the corresponding average change in Y. In this way it represents the degree to which the line slopes upwards or downwards. Once the correlation coefficient has been computed regression coefficients are easy to work out.

(1)	(2)	(3)	Col (2)	Col (3)	Col (2) xCol (3)
Child Number	Height in cms X	Dead space in ml Y	Square X^2	Square Y^2	XxY
1	110	44	12100	1936	4840
2	116	31	13456	961	3596
3	124	43	16576	1849	5332
4	129	45	16641	2025	5805
5	131	56	17161	3136	7336
6	138	79	19044	6241	10902
7	142	57	20164	3249	8094
8	150	56	22500	3136	8400
9	153	58	23409	3346	8874
10	155	92	24025	8464	14260
11	156	78	24336	6084	12168
12	159	64	25281	4096	10178
13	164	88	26896	7744	14432
14	168	112	28224	12544	18816
15	174	101	30276	10201	17574
n = 15	ΣX 2169	ΣY = 1004	ΣX^2 = 318889	ΣY^2 = 75030	ΣXY 150605

The line representing the

$$\Sigma Y = na + b\Sigma X \qquad (1)$$

$$\Sigma XY = a\Sigma X + b\Sigma X^2 \qquad \ldots\ldots(2)$$

$$Y = a + bX$$

From eq (1) and (2)

$$b = \frac{\Sigma XY - (\Sigma X)(\Sigma Y)/n}{\Sigma X^2 - (\Sigma X)^2/n} \qquad b = \frac{150605 - (2169 \times 1004/15}{318889 - (2169)^2/15}$$

$$b = 1.033$$

$$a = \Sigma Y / n - b(\Sigma X)/n \qquad a = 1004/15 - 1.033\ (2169/15)$$

$$a = -82.4$$

$$Y = -82.4 + 1.033\ X \qquad \text{(Regression equation)}$$

equation is shown superimposed on the scatter diagram in Figure 2.

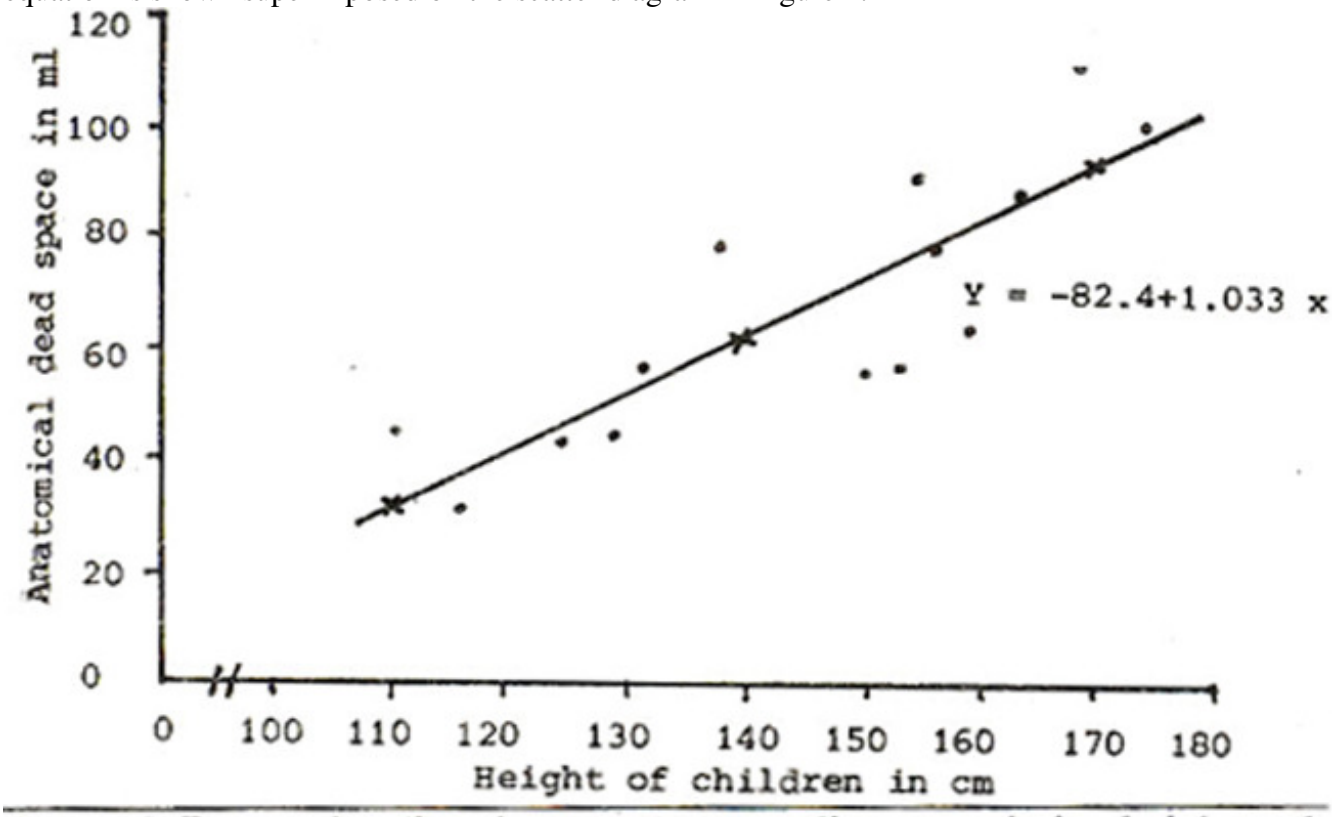

Figure 2. Regression line drawn on scatter diagram relating height and pulmonary anatomical dead space in 15 children (Figure 1).

The way to draw the line is to take three values of X, one on the left side of the scatter diagram one in

the middle, and one on the right, and substitute these in equation.

If X = 110 Y = -82.4 + (1.033 x 110) = 31.2

X = 140 Y = -82.4 + (1.033 x 140) = 62.2

X = 170 Y = -82.4 + (1.033 x 170) = 93.2

REFERENCES

1. Morton, R. F. and Hebel, J. R. A Study Guide to Epidemiology and Biostatistics. University Park Press, Baltimore 1983, pp. 81.84.

2. Chaudhry, S. M.. Introduction to Statistical Theory Part.!, Markazi Kutub Khana Urdu Bazar, Lahore 1975, pp. 178-182.

3. Swinscow, T.D.V. Statistics at Square One. British Medical As. sociation 1978, pp 62-70 & 78.