INFORMATION POINT:

Coefficient of Determination, R^2

In simple linear regression we are interested in how well one variable, X (the independent or explanatory variable), will predict another variable, Y (the dependent or outcome variable). Simple linear regression assumes that there is a straight line relationship between X and Y and the statistical procedure determines the line that best fits the data.

If there is a perfect linear relationship between X and Y then differences in the individual's values of Y will be accounted for by differences in their values of X. If the relationship is weaker, then variation in the values of the explanatory variable will not account for all the variation in values of the outcome variable. The coefficient of determination measures the proportion of the variation in the outcome variable that is explained by the explanatory variable.

The coefficient of determination is often called R^2 because it is the square of the Pearson correlation coefficient, often denoted by R. The coefficient of determination takes values between 0 and 1 (or 0% to 100%). If $R^2 = 1$, then X is a perfect predictor of Y, and variation in X completely accounts for variation in Y. If $R^2 = 0$ then X does not account for any of the variation in Y. We can use R^2 to help select between two competing regression models: we would prefer the model that explained most variation, that is the one with highest value of R^2.

In the above study, the results for the internal medicine ward suggest that OPC explains only 34% of the variation in PAONCIL, leaving 66% of the variation unaccounted for, so OPC would be likely to give poor predictions of PAONCIL.

Further reading

Polit D.F. (1996) *Data analysis and statistics for nursing research*. Appleton and Lange, Stamford, CT.

NICOLA CRICHTON