

Comments on Statistical Issues in March 2013

Commentary

Yong Gyu Park

Department of Biostatistics, The Catholic University of Korea College of Medicine, Seoul, Korea

In this section, we explain the relationship and distinction between a correlation coefficient and a slope in simple linear regression, which appeared in the article titled, "Do people have healthier lifestyles in greener environments? An analysis of the association between green environments and physical activity in seven large Korean cities," published in January 2013 by Park et al.¹⁾

CORRELATION COEFFICIENT AND SLOPE IN SIMPLE LINEAR REGRESSION

Association and correlation analyses mathematically identify and describe relationships between variables. In general, two variables are considered to be related when a change in one is likely to be accompanied by a change in the other. Although association and correlation are terms with general meanings, in their statistical uses, association is typically reserved for describing relationships between categorical variables; whereas correlation is typically reserved for describing relationships between continuous variables.

Pearson's correlation coefficient, r , is used to assess the linear relationship between two approximately normally distributed, continuous variables. Actually, the variables should vary together in a "bivariate normal distribution." Spearman's correlation coefficient is used to assess the (linear or non-linear) relationship between two continuous variables, at least one of which is not normally distributed.

A correlation coefficient indicates the strength and direction of the relationship between two variables and ranges from -1 to $+1$, where 1 is a perfect correlation and 0 is no correlation. A negative coefficient means that one variable tends to increase as the other decreases, and a positive coefficient means that both variables tend to increase or decrease together. Correlation

analysis does not test for causal relationships, only whether there is a relationship and, to some extent, the strength of this relationship.

Linear regression analysis is used to predict or estimate the value of a normally distributed response variable from the known values of one or more explanatory variables which are not necessarily normally distributed. When the analysis uses a single explanatory variable, the procedure is called simple regression; when it uses a combination of explanatory variables, it is called multiple regression.

An estimated simple linear regression line is described by the equation for a line:

$$Y = a + bX$$

where Y = the value of the response variable to be predicted, a = the point at which the regression line crosses the Y axis (intercept), b = the slope of the regression line, X = the explanatory variable used to predict the value of Y .

Given a value for the explanatory variable, X , the corresponding value of Y can be computed. Thus, Y can be predicted for any value of X within the range of X studied.

Whereas the correlation coefficient, r , indicates the direction and strength of the relationship between two variables, the regression coefficient for the explanatory variable (the slope of the regression line, b) indicates how much the average value of the response variable, Y , varies with each unit change in the explanatory variable, X . In regression analysis, the coefficient of determination, R^2 , indicates how much of the variability in a response variable is explained by knowing the value of an explanatory variable.

There are some caveats when we present the results of correlation analysis and simple linear regression analysis. We only need to present the correlation coefficient, r , and the P -value as results of correlation analysis; however, we should present regression coefficients, their standard errors, and P -values with regression

analysis. Most of all, we should not present the standardized regression coefficient, which is well known as the beta coefficient and equal to the correlation coefficient, r .

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCE

1. Park JY, Shin HK, Choi JS, Oh HS, Choi KH, Park SM, et al. Do people have healthier lifestyles in greener environments? An analysis of the association between green environments and physical activity in seven large Korean cities. *Korean J Fam Med* 2013;34:58-63.