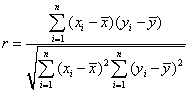
**Correlation and Regression**

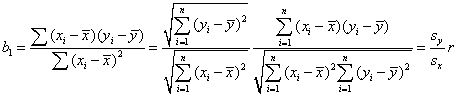
Correlation and regression are intimately related. The sample correlation coefficient between X and Y is



When Y is regressed on X, the regression coefficient of X is

http://www.jerrydallal.com/LHSP/pix/b1.gif

Therefore, the regression coefficient is the correlation coefficent multiplied by the ratio of the standard deviations.



Since the ratio of standard deviatons is always positive, testing whether the population regression coefficient is 0 is equivalent to testing whether the population correlation coefficient is 0. That is, the test of H0: http://www.jerrydallal.com/LHSP/symbol/beta.gif1 = 0 is equivalent to the test of H0: http://www.jerrydallal.com/LHSP/symbol/rho.gif= 0.

While correlation and regression are intimately related, they are not equivalent. The regression equation can be estimated whenever the Y values result from random sampling. The Xs can result from random sampling or they can be specified by the investigator. For example, crop yield can be regressed on the amount of water crops are given regardless of whether the water is rainfall (random) or the result of turning on an irrigation system (by design). The correlation coefficient is a characteristic of the joint distribution of X and Y. In order to estimate the correlation coefficient, both variables must be the result of random sampling. It makes sense to talk about the correlation between yield and rainfall, but it does not make sense to talk about the correlation between yield and amounts of water under the researcher control. This latter correlation will vary according to the specific amounts used in the study. In general, the correlation coefficient will increase or decrease along with the range of the values of the predictor.

*Copyright © 2000* [*Gerard E. Dallal*](mailto:Jerry.Dallal@REMOVEtuftsREMOVE.edu) *Last modified: 07/16/2008 02:54:17.*