Q: Is the Property of Being Positively Correlated Transitive?

This question was raised in an article of the same name by Eric Langford, Neil Schwertman, and Margaret Owens in The American Statistician, vol. 55, #4, November 2001.

The basic question is easily posed: If X, Y, and Z are three random variables and if X and Y are positively correlated, and Y and Z are positively correlated, can X and Z be negatively correlated. The counter intuitive answer is yes! The importance of this result for understanding regression analysis involving more than two variables is clear.

The authors state theorem 1: Let U, V, and W be three independent random variables and define X, Y, and Z by: X=U+V, Y=W+V, Z=W-U. Then ρ_{XY} and ρ_{YZ} are positive, but ρ_{XZ} is negative.

The article gives an example involving the number of triples, base hits, and home runs by the New York Yankees; triples and base hits are positively correlated, base hits and home runs are positively correlated, but triples and home runs are negatively correlated. Can you suggest why this might be expected?

Using the theorem; suggest triples of random variables that you suspect will exhibit this property. If you can test your conjecture using actual data, do so.

For the student comfortable using trigonometric relationships. The correlation coefficient can be interpreted as the cosine of the angle between the vectors. If X, Y, and Z are three n dimensional vectors, we can define:

$$\rho_{XY} = \frac{\sum xy}{\sqrt{\sum x^2}\sqrt{\sum y^2}} = \cos(\theta_{XY})$$

where $\cos(\theta_{XY})$ is the cosine of the angle between the vectors X and Y. If the correlation and the cosine are positive the corresponing angle is acute, but if the the correlation and the cosine are negative the corresponding angle is obtuse.

Using this definition, prove that for any random variables such that ρ_{XY} and ρ_{YZ} are positive and such that

$$\rho_{XY}^2 + \rho_{YZ}^2 > 1$$

then ρ_{XZ} is positive. Illustrate your proof in terms of a three dimensional diagram. Give a statistical interpretation of this result. Can you discover any real examples?