Linear Correlation

1. Standardization.

   Let X is a random variable with mean EX and variance σ. Recall that the corresponding standardized variable

   $$Z = \frac{X - EX}{\sigma}$$

   satisfies: $E(Z) = 0$, $V(Z) = 1$, and $E(Z^2) = 1$.

2. Linear Correlation

   The *coefficient of linear correlation* ρ between two random variables X and Y defined on the same sample space is defined by

   $$\rho = E(Z_1 Z_2)$$

   where Z_1 and Z_2 are the corresponding standardized variables. Notice that:

   $$\begin{aligned} 0 &\leq E(Z_1 - \rho Z_2)^2 \\ &= EZ_1^2 + \rho^2 EZ_2^2 - 2\rho E(Z_1 Z_2) \\ &= 1 - \rho^2 \end{aligned}$$

   from which follows $-1 \leq \rho \leq 1$.

3. A Geometric Analog. Let Z_1 and Z_2 be unit vectors in $\mathbb{R}^n$. The projection of Z_1 on Z_2 is the vector ρZ_2 where $\rho = \langle Z_1, Z_2 \rangle$ is the scalar product of Z_1 and Z_2. Since Z_1 and Z_2 are unit vectors it follows that ρ is just the cosine of the angle between Z_1 and Z_2.

   Notice that:

   $$\begin{aligned} 0 &\leq |Z_1 - \rho Z_2|^2 \\ &= |Z_1|^2 + \rho^2 |Z_2|^2 - 2\rho\langle Z_1, Z_2 \rangle \\ &= 1 - \rho^2 \end{aligned}$$

   It is seen that the geometric analogue of a standardized random variable is a unit vector. The coefficient of linear correlation plays the same role as the cosine of the angle between two unit vectors.

[1] http://pennance.us