A review of statistical concepts

Let X, Y, and Z be random variables and a, b, and c be constants. Then the **expected value**, or mean, of X is denoted by $\mathrm{E}(X)$ and

$$\mathrm{E}(a + bX + cY) = a + b\mathrm{E}(X) + c\mathrm{E}(Y) \tag{1}$$

The **variance** of X, denoted by $\mathrm{Var}(X)$, is $\mathrm{E}(X - \mathrm{E}(X))^2$. For two variables X and Y, their **covariance** is $\mathrm{Cov}(X, Y) = \mathrm{E}\big((X - \mathrm{E}(X))(Y - \mathrm{E}(Y))\big)$. If X and Y are **independent**, written $X \perp Y$, $\mathrm{Cov}(X, Y) = 0$. Also,

$$\mathrm{Cov}(X, X) = \mathrm{Var}(X) \tag{2}$$

$$\mathrm{Var}(a + bX + cY) = b^2\mathrm{Var}(X) + c^2\mathrm{Var}(Y) + 2bc\mathrm{Cov}(X, Y) \tag{3}$$

$$\mathrm{Cov}(aX, bY + cZ) = ab\mathrm{Cov}(X, Y) + ac\mathrm{Cov}(X, Z) \tag{4}$$

The **correlation** between Y and X,

$$\mathrm{Corr}(Y, X) = \frac{\mathrm{Cov}(Y, X)}{\sqrt{(\mathrm{Var}(Y)\mathrm{Var}(X))}} = \rho_{YX} \tag{5}$$

is a measure of the strength of the **linear regression** of dependent variable Y on independent variable X, that expresses the **conditional expectation** of $Y|X$ as

$$\mathrm{E}(Y|X = x) = \beta_0 + \beta_X X \tag{6}$$

where the **slope** β_X is

$$\beta_X = \frac{\mathrm{Cov}(Y, X)}{\mathrm{Var}(X)} \tag{7}$$

and the **intercept** β_0 is

$$\beta_0 = \mathrm{E}(Y) - \beta_X\mathrm{E}(X) \tag{8}$$