Expected value

Discrete

$$\mathrm{E}[X] = \mu = \sum_{i=1}^{n} p_i \cdot x_i = x_1 p_1 + x_2 p_2 + \cdots + x_n p_n \tag{1}$$

Continuous

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x \tag{2}$$

Properties

$$\mathrm{E}[aX + bY + c] = a\,\mathrm{E}[X] + b\,\mathrm{E}[Y] + c \tag{3}$$

Arithmetic mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_n}{n} \tag{4}$$

Variance

Discrete

$$\mathrm{Var}(X) = \sum_{i=1}^{n} p_i \cdot (x_i - \mu)^2 = \sum_{i=1}^{n} (p_i \cdot x_i^2) - \mu^2, \qquad \mu = \sum_{i=1}^{n} p_i \cdot x_i \tag{5}$$

Continuous

$$\mathrm{Var}(X) = \int (x-\mu)^2\, f(x)\, dx \; = \int x^2\, f(x)\, dx \; - \mu^2, \qquad \mu = \int x\, f(x)\, dx \tag{6}$$

Properties

$$\sigma^2 = \mathrm{Var}(X) = \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2 \tag{7}$$

Sample

$$s_x^2 = \frac{1}{n-1}\sum (x_i - \hat{x})^2 \tag{8}$$

Standard deviation

$$\sigma = \sqrt{\mathrm{Var}(X)} = \sqrt{\mathrm{E}[(X - \mathrm{E}[X])^2]} = \sqrt{\mathrm{E}[X^2] - (\mathrm{E}[X])^2} \tag{9}$$

Covariance

$$\mathrm{cov}(X, Y) = \mathrm{E}\left[(X - \mathrm{E}\left[X\right])\left(Y - \mathrm{E}\left[Y\right]\right)\right] = \mathrm{E}\left[XY\right] - \mathrm{E}\left[X\right]\mathrm{E}\left[Y\right]. \tag{10}$$

Pearson correlation coefficient

$$\rho_{X,Y} = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\mathrm{E}\left[(X - \mathrm{E}\left[X\right])\left(Y - \mathrm{E}\left[Y\right]\right)\right]}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}} \tag{11}$$

Regression Analysis

General regression model

$$
\begin{aligned}
&Y \approx f(\mathbf{X}, \boldsymbol{\beta}) \\
&Y = f(\mathbf{X}, \boldsymbol{\beta}) + \boldsymbol{\epsilon} \\
&y_i = f(x_i, \beta) + \epsilon_i, \quad f(x_i, \beta) = \hat{y}
\end{aligned}
$$

$$\mathbf{X} = \left[x_1, x_2, \ldots, x_n\right], \boldsymbol{\beta} = \left[\beta_0, \beta_1, \ldots, \beta_k\right] \boldsymbol{\epsilon} = \left[\epsilon_1, \epsilon_2, \ldots, \epsilon_n\right]$$

Sum of squared errors

$$SSE = \sum_{i=1}^{n} e_i^2. \tag{12}$$

Total sum of squares

$$SST = \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2 \tag{13}$$

Least square sum