과제1: Pearson Correlation Coefficient 함수

Pearson Correlation Coefficient 함수

$$r_{XY} = \frac{\frac{\sum_i^n (X_i - \bar{X})(Y_i - \bar{Y})}{n}}{\sqrt{\frac{\sum_i^n (X_i - \bar{X})^2}{n}}\sqrt{\frac{\sum_i^n (Y_i - \bar{Y})^2}{n}}}$$

가 아래의 수식과 동일한 표현이라는 것을 보이시오.

$$r_{XY} = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

1. 분자 변형

$$\sum_i^n (X_i - \bar{X})(Y_i - \bar{Y})$$

$$= \sum (X_i Y_i - \bar{X} Y_i - X_i \bar{Y} + \bar{X}\bar{Y}$$

$$= \sum x_i Y_i - \bar{X} \sum Y_i - \bar{Y} \sum X_i + n \cdot \bar{X}\bar{Y}$$

$$\bar{X} = \frac{1}{n} \sum x_i , \quad \bar{Y} = \frac{1}{n} \sum Y_i$$

$$\therefore \sum x_i Y_i - \frac{1}{n} \sum x_i Y_i - \frac{1}{n} \sum x_i y_i + \frac{1}{n} \sum x_i \sum Y_i$$

$$= \sum x_i y_i - \frac{1}{n} \sum x_i y_i$$

$$= \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n}$$

2. 분모 변형

i) $$\sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} = \sqrt{\frac{\sum (x_i^2 - 2x_i \bar{x} + \bar{x}^2)}{n}} = \sqrt{\frac{1}{n} \sum (x_i)^2 - \frac{2}{n} \bar{x} \sum x_i + \frac{\bar{x}^2}{n}}$$

$$= \sqrt{\frac{1}{n} \sum (x_i)^2 - \frac{1}{n^2} (\sum x_i)^2} \quad (\because \bar{x} = \frac{1}{n} \sum x_i)$$

$$= \sqrt{\frac{n \sum (x_i)^2 - (\sum x_i)^2}{n^2}}$$

$$= \frac{1}{n} \sqrt{n \sum (x_i)^2 - (\sum x_i)^2}$$

ii) $\sqrt{\frac{\Sigma(y_i-\bar{y})^2}{n}}$ 도 같은 과정을 거쳐

$$= \frac{1}{n}\sqrt{n\Sigma(y_i)^2 - (\Sigma y_i)^2}$$

$$\therefore \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n}} \cdot \sqrt{\frac{\Sigma(y_i-\bar{y})^2}{n}} = \frac{\sqrt{\Sigma(x_i)^2-(\Sigma x_i)^2} \cdot \sqrt{\Sigma(y_i)^2-(\Sigma y_i)^2}}{n}$$

$$\frac{\frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n}}{\frac{\sqrt{\Sigma(x_i)^2-(\Sigma x_i)^2} \cdot \sqrt{\Sigma(y_i)^2-(\Sigma y_i)^2}}{n}} = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{\sqrt{\Sigma(x_i)^2-(\Sigma x_i)^2} \cdot \sqrt{\Sigma(y_i)^2-(\Sigma y_i)^2}}$$

$\therefore$ 두 식은 동일하다