2주차 보고서

Created	@September 15, 2021 6:28 PM
Created by	20190258 김혜린

Pearson Correlation Coefficient 함수

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

위와 같은 식을 만족하는 Pearson Correlation Coefficient 함수는 아래 수식과 동일하다.

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

공분산 $Cov(X,Y)$ 에 대해 아래와 같은 식을 유도할 수 있다.

$$\begin{aligned} Cov(X,Y) = E[(X - \bar{X})(Y - \bar{Y})] &= E(XY - \bar{X}Y - X\bar{Y} + \bar{X}\bar{Y}) \\ &= E(XY) - \bar{X}E(Y) - E(X)\bar{Y} + \bar{X}\bar{Y} \\ &= E(XY) - \bar{X}\bar{Y} - \bar{X}\bar{Y} + \bar{X}\bar{Y} \\ &= E(XY) - \bar{X}\bar{Y} \end{aligned}$$

따라서 공분산 $Cov(X,Y)$ 는 아래와 같은 식을 만족한다.

$$\begin{aligned} Cov(X,Y) &= E(XY) - \bar{X}\bar{Y} \\ &= \frac{\Sigma xy}{n} - \frac{\Sigma x}{n}\frac{\Sigma y}{n} \\ &= \frac{n\Sigma xy - \Sigma x \Sigma y}{n^2} \end{aligned}$$

또한 표준편차 σX는 아래와 같은 식을 만족한다.

$$\sigma X = \sqrt{Var(X)} = \sqrt{E(X^2) - E(X)^2}$$

따라서 $\sigma X \sigma Y$ 를 아래와 같이 서술할 수 있다.

$$\begin{aligned}\sigma X \sigma Y &= \sqrt{E(X^2) - E(X)^2}\sqrt{E(Y^2) - E(Y)^2} \\ &= \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}\sqrt{\frac{\Sigma y^2}{n} - \left(\frac{\Sigma y}{n}\right)^2} \\ &= \sqrt{\frac{n\Sigma x^2 - (\Sigma x)^2}{n^2}}\sqrt{\frac{n\Sigma y^2 - (\Sigma y)^2}{n^2}} \\ &= \frac{\sqrt{n\Sigma x^2 - (\Sigma x)^2}}{n}\frac{\sqrt{n\Sigma y^2 - (\Sigma y)^2}}{n} \\ &= \frac{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}{n^2}\end{aligned}$$

따라서 Pearson Correlation Coefficient r_{XY} 는 아래와 같은 수식을 만족한다.

$$\begin{aligned}r_{XY} &= \frac{\frac{n\Sigma xy - \Sigma x \Sigma y}{n^2}}{\frac{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}{n^2}} \\ &= \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}\end{aligned}$$

따라서 두 수식은 동일한 표현이다.

$$r_{XY} = \frac{\frac{\sum_i^n (X_i - \bar{X})(Y_i - \bar{Y})}{n}}{\sqrt{\frac{\sum_i^n (X_i - \bar{X})^2}{n}}\sqrt{\frac{\sum_i^n (Y_i - \bar{Y})^2}{n}}} = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$