고급소프트웨어 실습1

Week2 - 과제1 학번: 20171669 이름: 이재영

# 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 분자의 경우

$$\frac{\sum_i^n (X_i - \bar{X})(Y_i - \bar{Y})}{n} = \frac{\sum_i^n (X_iY_i - X_i\bar{Y} - \bar{X}Y_i + \bar{X}\bar{Y})}{n}$$

$$= \frac{\sum_i^n (X_iY_i) - \bar{Y}\sum_i^n X_i - \bar{X}\sum_i^n Y_i + n \cdot \bar{X} \cdot \bar{Y}}{n}$$

이때, $\bar{X}$와 $\bar{Y}$는 각각 ~~표본~~평균으로 $\bar{X} = \sum_i^n X_i$, $\bar{Y} = \sum_i^n Y_i$이다.

$$\therefore \frac{\sum_i^n (X_iY_i) - \frac{1}{n}\sum_i^n X_i \sum_i^n Y_i - \frac{1}{n}\sum_i^n X_i \sum_i^n Y_i + \frac{1}{n}\sum_i^n X_i \cdot \sum_i^n Y_i}{n}$$

$$= \frac{\sum_i^n X_iY_i - \frac{1}{n}\sum_i^n X_i \sum_i^n Y_i}{n}$$

//

② 분모의 경우

$$\sqrt{\frac{\sum_i^n (X_i-\overline{X})^2}{n}} = \sqrt{\frac{\sum_i^n (X_i^2 - 2X_i\cdot\overline{X} + \overline{X}^2)}{n}} = \sqrt{\frac{\sum_i^n (X_i)^2 - 2\overline{X}\cdot\sum_i^n X_i + n\cdot\overline{X}^2}{n}}$$

$$= \sqrt{\frac{\sum_i^n (X_i)^2 - \frac{2}{n}(\sum_i^n X_i)^2 + \frac{1}{n}(\sum_i^n X_i)^2}{n}}$$

$$= \sqrt{\frac{\sum_i^n (X_i)^2 - \frac{1}{n}(\sum_i^n X_i)^2}{n}}$$

$\sqrt{\frac{\sum_i^n (Y_i-\overline{Y})^2}{n}}$ 도 위와 마찬가지로 하면 $\sqrt{\frac{\sum_i^n (Y_i)^2 - \frac{1}{n}(\sum_i^n Y_i)^2}{n}}$

$\therefore$ 분모: $\sqrt{\frac{\sum_i^n (X_i)^2 - \frac{1}{n}(\sum_i^n X_i)^2}{n}} \sqrt{\frac{\sum_i^n (Y_i)^2 - \frac{1}{n}(\sum_i^n Y_i)^2}{n}}$

$\sum_i^n$은 편의상 $\sum$로 두면

$$= \sqrt{\frac{\sum X^2 \sum Y^2 - \frac{1}{n}\sum X^2 (\sum Y)^2 - \frac{1}{n}(\sum X)^2 \sum Y^2 + \frac{1}{n}(\sum X)^2 (\sum Y)^2}{n^2}}$$

$$= \frac{1}{n^2}\sqrt{n^2 \sum X^2 \sum Y^2 - n\sum X^2 (\sum Y)^2 - n(\sum X)^2 \sum Y^2 + (\sum X)^2(\sum Y)^2}$$

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

//

$$\therefore \quad r_{XY} = \frac{\dfrac{\sum XY - \frac{1}{n}\sum X \sum Y}{n}}{\frac{1}{n^2}\sqrt{\left(n\sum X^2 - (\sum X)^2\right)\left(n\sum Y^2 - (\sum Y)^2\right)}}$$

$$= \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{\left(n\sum X^2 - (\sum X)^2\right)\left(n\sum Y^2 - (\sum Y)^2\right)}} \quad //$$

따라서 Pearson Correlation coefficient 함수와 위의 수식은 동일하다.