|  |  |
| --- | --- |
| 교육 제목 | **Statistics : Basics** |
| 교육 일시 | 27 OCT 2021 |
| 교육 장소 | YGL-C1 |
| **교육 내용** | |
| 오전 | Numerical data (Regression) – of our interest in terms of AI  Interval : Absolute value of zero X Ratio : Absolute value of zero O ex)  Frequency table Histogram(Probability density function) Boxplot(i.e., Five number statistics - Composed of Whisker, Staple etc)      Violin plot    Categorical data (Classification)  Nominal : Rank X Ordinal : Rank O  ex) Contingency table, Barplot, Pie chart    Bigdata : mother group to sample group?  Population : Entire group to be analysed  Sample : Selected part of the group : Estimator only comes out  Pros and cons of summarising data through graph  Intuitive Subjective – can be distorted somehow  Summary of numerical data – valid assuming data follows normal distribution  Sample mean : Preferred : Influence of Outlier O : Trimmed mean Median : Preferred : Influence of Outlier X Mode : Rarely employed |
| 오후 | Between population and sample  Variance, Var : not absolute value : only for comparsion  Standard deviation, SD  Range = Max value– Min value  Inter Quartile Range, IQR =  Upper outlier Lower outlier    Correlation : Search especially for linear correlation()  Regression through ML  Pearson’s r  0.7 ~ 0.9 : High corelation  0.4 ~ 0.7 : Intermediate  Spearman’s r  In case of irregular distribution |