

## # 7월 15일 산보연 보고사항 #

1) 중분류별 Spline 모형 적합해 forecasting 시행 후 결과 시각화!

; train Data는 2000년 ~ 2014년, validation data는 2015년 ~ 2018년으로 Split

; 반응변수는 "통합 누적 발생률", 설명변수는 "YEAR"로 linear Regression model 적합!

; Performance Criteria는  $MAPE = \frac{1}{4} \sum_{t=2015}^{2018} \left| \frac{y_{it} - \hat{y}_{it}}{y_{it}} \right|$  (i는 UP2, t는 시점) 사용 / 나중에 비교할 수 있도록 따로 저장!

; 시각화 시, 실제 통합 누적 발생률과 fitted line 구분되도록, 2014년 기준으로 trend line 구분되도록 시각화!

/ MAPE를 graph 안 text로 표시! → "Nan" 값이 나오는 경우 "-99"로 값 변경해 놓기

; Spline 모형 2가지 종류 적합해보기 / 1차 Spline 모형 (knots: 2000년 ~ 2005년, 2005년 ~ 2010년,

/ 3차 Spline 모형 (knots 제시할 필요 없음)

2010년 ~ 2014년)