**바이탈의 Missing value에 대한 Linear Interpolation의 타당성 검증(T-test)**

David Hwang

**Introduction**:

바이탈의 Missing value에 대해 우리 연구는 단변량 기준(각각의 vital) Linear 함수를 이용하여 missing value를 보간하였다. 이는 carry-forward의 constant한 값보다 훨씬 시간의 흐름에 따라 vital이 합리적으로 보간된다고 보았기 때문이다.

다만, carry-forwarding, Mean, Linear Interpolation 모두가 갖고 있는 한계점은, 이들이 **분산 variance를 과소평가한다는 것**이다. 그러나 이러한 Missing value 극복 방안 중에서는 제일 자연스럽다고 생각된다.

우리는 여기서 이러한 한계점을 갖고 있는 Linear Interpolation을 적용한 데이터가 실제 임상 데이터에 대해서 어느정도 유효한지 그 보간한 범위의 타당성을 검토해보고자 한다.

**Method**:

1. Missing value가 없는 바이탈 데이터셋(full data)를 준비한다.
2. Full data로부터 각각의 바이탈에 대하여 20% 정도의 missing value 데이터셋을 생성한다.
3. 실제값과 예측값에 대한 T-test를 진행한다.
4. (실제값 – 예측값)의 잔차 residual에 대한 분산, 평균, 표준편차를 비교한다.

세부적인 구현은 다음과 같다.

             y0           y               y1

--------t0----------t---------t1---------->t

          |<-- s1   -> |<-  s2 ->|

1. linear interpolation을 위해서는 t-t0의 간격인 s1, t1-t의 간격인 s2를 설정

2. 데이터셋

    ----------------------------------------------

    -----------input--------|-----output-----------

    t0  t   t1  y0      y1  |      y    predict

    ----------------------------------------------

    1   2   3   100     50  |      71   75

    데이터셋: 1000개의 vital

3. 구현 절차

    1. 먼저 t 무작위로 선정

    2. t0 = t-1 ~ t-6 사이의 값으로 무작위 설정(즉, 1시간 ~ 6시간 간격)

    3. t1 = t+1 ~ t+6 사이의 값으로 무작위 설정(즉, 1시간 ~ 6시간 간격)