**C07: Model Diagnostics 1**

배균섭

**Single run based NONMEM extended output**

NONMEM을 한 번 실행하였을 때 어떠한 결과(output)들이 있을 것인가는 NONMEM이 제공하는 표준적인 OUTPUT file을 보면 알 수 있다. 실은 이것들 외에도 사용자가 많은 diagnostics를 생성해 낼 수 있는데, 다음과 같은 것들이 있다.

Extended output은 다음과 같이 5 범주로 구분하여 보았다.

1) Objective function value를 포함한 minimization 관련 수치들

2) Theta, Omega, Sigma와 그 관련 수치들

3) Prediction value와 관련 수치들

4) Residual value와 관련 수치들

5) Empirical Bayes Estimate와 관련 수치들

이외에도 다음과 같은 3가지의 부수적인 관련 file을 검토하면 좋다.

A) Input dataset과 관련 수치들 – 입력자료의 무결성을 점검하기 위함

B) NONMEM standard output을 beautify한 것 – 보고서 포함용

C) Individual parameters와 관련 수치들 – 구조 모형 파라미터의 현실성 검증

여기에서는 앞의 다섯 범주에 대해서 좀 더 살펴 보기로 한다.

1. 최적화 과정 관련 지표들
2. Objective Function Value (OFV): NONMEM Objective Function의 최소화 된 값. Minimum value of objective function (MVOF)라고도 한다.
3. OFV per DV: OFV divided by DV record count, DV record당의 OFV로서 작을수록 적합(fitting)이 잘 되었다는 것을 알 수 있다. CWRES의 평균과 유사한 개념이다.
4. Corrected AIC (AICc): AIC 값이 DV records count를 반영하지 않은 것에 대해 보정해준 것으로 Peter Bonate책에 설명이 되어 있다. 계산식은 다음과 같다.  
   AICc = OFV + 2 \* nPara + 2\*nPara\*(nPara+1)/(nDV-nPara-1)  
   nPara: number of parameters  
   nDV: number of DV records (DV records count)
5. Schwartz Criterion (SBC, BIC): 계산식은 다음과 같다.  
   SBC = OFV + nPara \* log(nDV)
6. Count of Gradients absolute values are larger than 1: UCP(STP)에서 마지막 gradient가 1보다 큰 파라미터의 개수. Final gradient는 0에 가까울수록 좋은 것이므로, 절대치가 1보다 큰 값이 많을수록 최적화 과정에 문제가 있음을 시사한다. Original scale에서의 값은 계산할 수 있으나, 출력되지 않는다. UCP에서의 값은 final estimate뿐 아니라, 초기에 준 bound에도 영향을 받는다. NONMEM OUTPUT에서의 순서는 GRADIENT: 부분에서 순서대로 theta vector, vech(omega matrix), vech(sigma matrix) 이다.

|  |
| --- |
| GRADIENT: 0.5327E-02 -0.7969E-02 0.9797E-03 0.1949E-02 0.2570E-02 … |

R에서 Vech연산자는 아래와 같이 구현할 수 있다.

c(aSquareMatrix[upper.tri(aSquareMatrix, diag=T)])

1. 개인별 OFV per DV의 분포: 여기에 대해서는 아직 거의 알려져 있지 않지만, 따로 떨어진 피험자는 outlier를 의심해볼 수 있다. 각종 descriptive statistics를 볼 수도 있고, Shapiro-Wilks test같은 정규분포 검정을 할 수도 있지만, 그 의미는 명확하지 않다. OFV per DV값이 작은 피험자 일수록 자료가 모형에 잘 맞으며, 큰 피험자일수록 자료와 모형이 잘 맞지 않음을 의미한다. NONMEM이 이를 구해주지는 않기 때문에 이것을 구하기 위해서는 NONMEM에서 지원하는 individual OFV를 파일로 저장한 뒤에 개인별 DV record count로 나누어 주어야 한다. NONMEM control code에서는 다음과 같은 부분이다. $INFN, $ERROR, $PRED 등에 쓸 수 있다.

|  |
| --- |
| IF (ICALL.EQ.3) THEN  WRITE (55,\*) IIDX, CNTID  ENDIF |

IIDX: index of individual,

CNTID: contribution of that subject

1. 추정된 파라미터 관련 지표들
2. Number of parameters: unfixed theta, omega, sigma
3. Theta vector  
   Point estimate, Standard error, Lower Limit (Bound), Upper Limit (Bound)  
   Test on (H0: parameter is 0) and (H0: parameter is 1)
4. Omega matrix  
   - Point estimate  
   - Square root of diagonal elements: Exponential eta model을 사용하는 경우에 이것은 구조모형의 개인별 parameter들의 CV값에 대한 근사치이다.  
   - Standard error  
   - Relative Standard Error (RSE) in %: Minimization에 어려움이 있는 경우 이것이 50% 또는 100%가 넘는 omega element는 estimation하지 않을 것도 고려해본다.
5. Sigma matrix: 다음과 같은 theta를 이용한 combined error 모형을 사용하는 경우에는 sigma를 1개만 쓰고, 표준정규분포로 fix하기 때문에 따로 diagnostic를 볼 필요는 없다. 다만, 이 경우에 Epsilon residual의 표본표준편차가 1인지를 확인하면 이것이 곧 epsilon shrinkage를 계산한 것이 된다.
6. 예측치(엄밀히는 적합치) 관련 지표들
7. DV vs. PRED (Typical prediction, prediction using typical values of structural model parameters)
8. DV vs. IPRE (Individual prediction using EBE)
9. All subjects’ superposed [DV, PRED, IPRE] vs. Time after dose (TAD) in original scale
10. All subjects’ superposed [DV, PRED, IPRE] vs. Time after dose (TAD) in log scale of Y
11. Individual plot of 3), 4)
12. 잔차(residuals)와 관련 지표들
13. All subjects’ superposed WRES vs. [PRED, TAD]
14. All subjects’ superposed CWRES vs. [IPRE, TAD]
15. All subjects’ superposed IWRES vs. [IPRE, TAD]
16. Absolute y values of 1), 2), 3)
17. Distribution statistics of [WRES, CWRES, IWRES] : histogram, density plot, Q-Q plot, Shapiro-Wilk test, SD of residual (epsilon shrinkage)
18. Test and CI of residual counts using binomial distribution
19. Ratio of [WRES, CWRES, IWRE] count to expected count
20. Residuals with extreme z values
21. Individual residual plot of 1), 2), 3)
22. Run test results of 9)
23. EBE와 관련 지표들
24. Descriptive statistics of Etas: minimum, 1Q, median, mean, 3Q, maximum, SD
25. T-test with H0: mean is zero.
26. Shapiro-Wilk test on normality
27. Etas’ population shrinkage
28. SE of EBE (individual eta shrinkage)
29. Histogram, density plot, Q-Q plot
30. Plot of etas vs. continuous covariates: histogram, trend plot
31. Plot of etas vs. categorical covariates
32. Omega estimation vs. [variance-covariance/correlation/SD] of EBEs  
    Ratio of Omega to Sample Variance-Covariance Matrix of EBEs : Replacement of eta population shrinkage  
    Those of correlation matrices
33. Multiple linear regression (MLR) of EBEs using covariates as independent variables

부록 1. Extended output을 생성하기 위한 control file 예시: NONMEM에 포함된 THEOPP file을 dataset으로 한 경우

|  |
| --- |
| $PROB THEOPHYLLINE ORAL P:ROOT F:BASE  $INPUT ID AMT TIME DV WT  $DATA THEOPP IGNORE=@  $SUBROUTINES ADVAN2 TRANS2  **$INFN**  **IF (ICALL.EQ.3) THEN**  **CLOSE (56)**  **WRITE (51,\*) OBJECT**  **WRITE (52,\*) THETA**  **WRITE (52,\*) SETHET**  **WRITE (53,\*) OMEGA(BLOCK)**  **WRITE (53,\*) SEOMEG(BLOCK)**  **WRITE (54,\*) SIGMA(BLOCK)**  **WRITE (54,\*) SESIGM(BLOCK)**  **WRITE (55,\*) IIDX, CNTID**  **CLOSE (55)**  **ENDIF**  $PK  TVCL = THETA(1)  TVV = THETA(2)  TVKA = THETA(3)  CL = TVCL \* EXP(ETA(1))  V = TVV \* EXP(ETA(2))  KA = TVKA \* EXP(ETA(3))  S2 = V  K = CL / V  $ERROR  **"FIRST**  **" INCLUDE 'C:/NMVI/SIZES'**  **" COMMON /ROCM36/ POSTV(LVR,LVR)**  **" DOUBLE PRECISION POSTV, PPSTV1**  **" INTEGER PREVID, CVRSTEP**  IPRE = F  W = SQRT(THETA(4)\*\*2 + THETA(5)\*\*2 \* IPRE\*\*2) ; W=0 경우 주의  IRES = DV - IPRE  IWRE = IRES / W  Y = IPRE + W \* EPS(1)  **"IF (PPSTV1.NE.POSTV(1,1)) THEN**  **" IF (PREVID.EQ.NIREC) THEN**  **" CVRSTEP = 1**  **" ELSEIF (CVRSTEP.EQ.1) THEN**  **" WRITE (56,\*) PREVID, (SQRT(POSTV(J\_,J\_)), J\_=1,NETAS\_)**  **" ENDIF**  **" PPSTV1 = POSTV(1,1)**  **" PREVID = NIREC**  **"ENDIF**  $THETA  (0, 0.1)  (0, 1)  (0, 10)  (0, 1)  (0, 0.5)  $OMEGA BLOCK(3)  0.2  0.1 0.2  0.1 0.1 0.2  $SIGMA 1 FIX  $EST MAXEVAL=9999 PRINT=1 METHOD=COND INTER MSFO=THIS.MSF NOABORT  **NOTBT NOOBT NOSBT**  $COV PRINT=ERS  $TAB ID TIME MDV IPRE IWRE  **G11 G21 G31 H11**  FILE=sdtab NOPRINT ONEHEADER  $TAB ID V CL KA K ETA(1) ETA(2) ETA(3)  FILE=patab NOPRINT ONEHEADER NOAPPEND  $TAB ID WT  FILE=cotab NOPRINT ONEHEADER NOAPPEND  $TAB ID  FILE=catab NOPRINT ONEHEADER NOAPPEND  $TAB ID WT  ETA(1) ETA(2) ETA(3)  FILE=COVARETA NOPRINT ONEHEADER NOAPPEND FIRSTONLY  $TAB ID CL K V KA  FILE=INDIPARA NOPRINT ONEHEADER NOAPPEND FIRSTONLY |