분석 결과

6. b.

결측값 확인

```
> table(d_2018$SEX)

    1     2
14594 15867
> table(d_2018$RACETHX)

    1     2     3     4     5
 7306 15836  4544  1565  1210
> table(d_2018$MARRY18X)

   -7     1     2     3     4     5     6
    5 11508  1656  2891   578  7263  6560
> table(d_2018$INS18X)

   -1     1     2
  216 26803  3442
> table(d_2018$POVCAT18)

    1     2     3     4     5
 5384  1580  4548  8726 10223
> table(d_2018$REGION18)

   -1     1     2     3     4
  216  4556  6388 11795  7506
> table(d_2019$SEX)

    1     2
13659 14853
> table(d_2019$RACETHX)

    1     2     3     4     5
 6562 15281  4158  1435  1076
> table(d_2019$MARRY19X)

   -8    -7     1     2     3     4     5     6
    1     4 10955  1636  2765   527  6874  5750

   -1     1     2
  243 25072  3197
> table(d_2019$POVCAT19)

    1     2     3     4     5
 4775  1481  4104  8139 10013
> table(d_2019$REGION19)

   -1     1     2     3     4
  243  4291  6006 10722  7250
>
```

Pooled 데이터셋

```
> pooled=rbind(d_2018, d_2019)
> head(pooled)
# A tibble: 6 × 10
  OBDRV      OBTOTV      AGE SEX       RACETHX MARRY   INS POVCAT  REGION OFFVST
  <dbl+lb>   <dbl+>    <dbl> <dbl+lb>  <dbl+lb> <dbl> <dbl> <dbl+lb>  <dbl>  <dbl>
1 5          5            27 2 [2 F…   2 [2 N…      1     1 3 [3 L…      2     10
2 1          1            25 1 [1 M…   2 [2 N…      1     1 3 [3 L…      2      2
3 1          1            34 2 [2 F…   1 [1 H…      1     1 3 [3 L…      2      2
4 0 [0]      0 [0]        39 1 [1 M…   1 [1 H…      1     1 3 [3 L…      2      0
5 1          1            11 1 [1 M…   1 [1 H…      6     1 3 [3 L…      2      2
6 0 [0]      0 [0]         8 1 [1 M…   1 [1 H…      6     1 3 [3 L…      2      0
>
```

Model18_lin

```
> summary(model18_lin)

Call:
lm(formula = OFFVST ~ AGE + factor(SEX) + factor(RACETHX) + factor(MARRY) +
    factor(INS) + factor(POVCAT) + factor(REGION), data = d_2018)

Residuals:
   Min     1Q Median     3Q    Max
-24.69  -7.40  -3.59   1.54 564.30

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      -1.795225   0.622702  -2.883 0.003942 **
AGE               0.236194   0.007911  29.858  < 2e-16 ***
factor(SEX)2      2.253947   0.209536  10.757  < 2e-16 ***
factor(RACETHX)2  3.108350   0.279863  11.107  < 2e-16 ***
factor(RACETHX)3  0.217305   0.356704   0.609 0.542394
factor(RACETHX)4 -1.659626   0.510356  -3.252 0.001148 **
factor(RACETHX)5  2.352777   0.562922   4.180 2.93e-05 ***
factor(MARRY)2   -0.210562   0.516219  -0.408 0.683356
factor(MARRY)3    1.141068   0.383295   2.977 0.002913 **
factor(MARRY)4    2.103851   0.777246   2.707 0.006797 **
factor(MARRY)5    1.923768   0.323032   5.955 2.62e-09 ***
factor(MARRY)6    5.356379   0.456569  11.732  < 2e-16 ***
factor(INS)2     -5.337399   0.342419 -15.587  < 2e-16 ***
factor(POVCAT)2  -0.591229   0.518482  -1.140 0.254167
factor(POVCAT)3  -0.879282   0.366482  -2.399 0.016435 *
factor(POVCAT)4  -1.298367   0.321817  -4.034 5.49e-05 ***
factor(POVCAT)5  -0.770765   0.330682  -2.331 0.019768 *
factor(REGION)2  -1.344752   0.350038  -3.842 0.000122 ***
factor(REGION)3  -1.085289   0.319838  -3.393 0.000692 ***
factor(REGION)4  -1.144317   0.343399  -3.332 0.000862 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 17.98 on 30220 degrees of freedom
Multiple R-squared:  0.08364,   Adjusted R-squared:  0.08306
F-statistic: 145.2 on 19 and 30220 DF,  p-value: < 2.2e-16
```

Model19_lin

```
> summary(model19_lin)

Call:
lm(formula = OFFVST ~ AGE + factor(SEX) + factor(RACETHX) + factor(MARRY) +
    factor(INS) + factor(POVCAT) + factor(REGION), data = d_2019)

Residuals:
   Min     1Q Median     3Q    Max
-23.78  -7.63  -3.77   1.55 670.52

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      -0.718094   0.648515  -1.107 0.268178
AGE               0.225774   0.008135  27.752  < 2e-16 ***
factor(SEX)2      2.098525   0.217485   9.649  < 2e-16 ***
factor(RACETHX)2  2.680861   0.291180   9.207  < 2e-16 ***
factor(RACETHX)3 -0.662495   0.374341  -1.770 0.076777 .
factor(RACETHX)4 -1.279324   0.535952  -2.387 0.016992 *
factor(RACETHX)5  2.351451   0.598820   3.927 8.63e-05 ***
factor(MARRY)2    1.240895   0.522014   2.377 0.017455 *
factor(MARRY)3    1.903820   0.393632   4.837 1.33e-06 ***
factor(MARRY)4    1.651438   0.820709   2.012 0.044207 *
factor(MARRY)5    1.847403   0.335453   5.507 3.68e-08 ***
factor(MARRY)6    5.031025   0.478741  10.509  < 2e-16 ***
factor(INS)2     -5.526779   0.356081 -15.521  < 2e-16 ***
factor(POVCAT)2  -0.795876   0.541207  -1.471 0.141422
factor(POVCAT)3  -1.422520   0.389383  -3.653 0.000259 ***
factor(POVCAT)4  -1.658600   0.338862  -4.895 9.90e-07 ***
factor(POVCAT)5  -0.482464   0.344911  -1.399 0.161882
factor(REGION)2  -1.454275   0.362177  -4.015 5.95e-05 ***
factor(REGION)3  -1.327688   0.331483  -4.005 6.21e-05 ***
factor(REGION)4  -1.304008   0.351942  -3.705 0.000212 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 18.03 on 28245 degrees of freedom
Multiple R-squared:  0.08606,   Adjusted R-squared:  0.08544
F-statistic:   140 on 19 and 28245 DF,  p-value: < 2.2e-16
```

Model_lin

```
> summary(model_lin)

Call:
lm(formula = OFFVST ~ AGE + factor(SEX) + factor(RACETHX) + factor(MARRY) +
    factor(INS) + factor(POVCAT) + factor(REGION), data = pooled)

Residuals:
   Min     1Q Median     3Q    Max
-24.19  -7.50  -3.67   1.55 670.84

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      -1.269519   0.448977  -2.828 0.004692 **
AGE               0.230974   0.005669  40.746  < 2e-16 ***
factor(SEX)2      2.177464   0.150881  14.432  < 2e-16 ***
factor(RACETHX)2  2.902323   0.201689  14.390  < 2e-16 ***
factor(RACETHX)3 -0.202724   0.258167  -0.785 0.432315
factor(RACETHX)4 -1.474507   0.369516  -3.990 6.61e-05 ***
factor(RACETHX)5  2.358223   0.410068   5.751 8.93e-09 ***
factor(MARRY)2    0.510554   0.366998   1.391 0.164181
factor(MARRY)3    1.511758   0.274570   5.506 3.69e-08 ***
factor(MARRY)4    1.883654   0.564205   3.339 0.000843 ***
factor(MARRY)5    1.882351   0.232626   8.092 5.99e-16 ***
factor(MARRY)6    5.192385   0.330273  15.721  < 2e-16 ***
factor(INS)2     -5.430039   0.246722 -22.009  < 2e-16 ***
factor(POVCAT)2  -0.691208   0.374247  -1.847 0.064763 .
factor(POVCAT)3  -1.133088   0.266792  -4.247 2.17e-05 ***
factor(POVCAT)4  -1.471536   0.233275  -6.308 2.84e-10 ***
factor(POVCAT)5  -0.623354   0.238590  -2.613 0.008986 **
factor(REGION)2  -1.394978   0.251653  -5.543 2.98e-08 ***
factor(REGION)3  -1.199577   0.230122  -5.213 1.87e-07 ***
factor(REGION)4  -1.229869   0.245662  -5.006 5.56e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 18.01 on 58485 degrees of freedom
Multiple R-squared:  0.08464,   Adjusted R-squared:  0.08434
F-statistic: 284.6 on 19 and 58485 DF,  p-value: < 2.2e-16
```

6.c

Model18_p

```
> summary(model18_p)

Call:
glm(formula = OFFVST ~ AGE + factor(SEX) + factor(RACETHX) +
    factor(MARRY) + factor(INS) + factor(POVCAT) + factor(REGION),
    family = "poisson", data = d_2018)

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)       1.122929   0.011436  98.194  < 2e-16 ***
AGE               0.020926   0.000132 158.576  < 2e-16 ***
factor(SEX)2      0.242196   0.003824  63.333  < 2e-16 ***
factor(RACETHX)2  0.359627   0.005646  63.700  < 2e-16 ***
factor(RACETHX)3  0.086464   0.007251  11.925  < 2e-16 ***
factor(RACETHX)4 -0.207127   0.011488 -18.030  < 2e-16 ***
factor(RACETHX)5  0.311351   0.010670  29.181  < 2e-16 ***
factor(MARRY)2   -0.101929   0.007268 -14.025  < 2e-16 ***
factor(MARRY)3    0.096819   0.005949  16.274  < 2e-16 ***
factor(MARRY)4    0.230151   0.012525  18.375  < 2e-16 ***
factor(MARRY)5    0.114579   0.006002  19.089  < 2e-16 ***
factor(MARRY)6    0.371887   0.008955  41.528  < 2e-16 ***
factor(INS)2     -0.970927   0.010146 -95.696  < 2e-16 ***
factor(POVCAT)2  -0.078078   0.009716  -8.036 9.29e-16 ***
factor(POVCAT)3  -0.112135   0.006819 -16.445  < 2e-16 ***
factor(POVCAT)4  -0.148567   0.005949 -24.974  < 2e-16 ***
factor(POVCAT)5  -0.081298   0.005902 -13.776  < 2e-16 ***
factor(REGION)2  -0.124307   0.005919 -21.000  < 2e-16 ***
factor(REGION)3  -0.100057   0.005412 -18.489  < 2e-16 ***
factor(REGION)4  -0.101806   0.005882 -17.307  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 549776  on 30239  degrees of freedom
Residual deviance: 456543  on 30220  degrees of freedom
AIC: 542010

Number of Fisher Scoring iterations: 6
```

Model19_p

```
> summary(model19_p)

Call:
glm(formula = OFFVST ~ AGE + factor(SEX) + factor(RACETHX) +
    factor(MARRY) + factor(INS) + factor(POVCAT) + factor(REGION),
    family = "poisson", data = d_2019)

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)       1.2452970  0.0117616 105.879  < 2e-16 ***
AGE               0.0198749  0.0001355 146.632  < 2e-16 ***
factor(SEX)2      0.2198107  0.0038985  56.383  < 2e-16 ***
factor(RACETHX)2  0.2998842  0.0057684  51.987  < 2e-16 ***
factor(RACETHX)3 -0.0295442  0.0076254  -3.874 0.000107 ***
factor(RACETHX)4 -0.1210643  0.0112249 -10.785  < 2e-16 ***
factor(RACETHX)5  0.2974100  0.0110419  26.935  < 2e-16 ***
factor(MARRY)2   -0.0006087  0.0071897  -0.085 0.932531
factor(MARRY)3    0.1533198  0.0059673  25.693  < 2e-16 ***
factor(MARRY)4    0.1843712  0.0134954  13.662  < 2e-16 ***
factor(MARRY)5    0.1103981  0.0061896  17.836  < 2e-16 ***
factor(MARRY)6    0.3307543  0.0093415  35.407  < 2e-16 ***
factor(INS)2     -1.0377103  0.0107982 -96.101  < 2e-16 ***
factor(POVCAT)2  -0.1037120  0.0099781 -10.394  < 2e-16 ***
factor(POVCAT)3  -0.1625083  0.0071511 -22.725  < 2e-16 ***
factor(POVCAT)4  -0.1805642  0.0061991 -29.128  < 2e-16 ***
factor(POVCAT)5  -0.0533815  0.0060510  -8.822  < 2e-16 ***
factor(REGION)2  -0.1259392  0.0060147 -20.938  < 2e-16 ***
factor(REGION)3  -0.1133951  0.0055176 -20.552  < 2e-16 ***
factor(REGION)4  -0.1069940  0.0059089 -18.107  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 527122  on 28264  degrees of freedom
Residual deviance: 438150  on 28245  degrees of freedom
AIC: 518484

Number of Fisher Scoring iterations: 6
```

Model_p

```
> summary(model_p)

Call:
glm(formula = OFFVST ~ AGE + factor(SEX) + factor(RACETHX) +
    factor(MARRY) + factor(INS) + factor(POVCAT) + factor(REGION),
    family = "poisson", data = pooled)

Coefficients:
                   Estimate Std. Error  z value Pr(>|z|)
(Intercept)       1.1832786  0.0081953  144.384  < 2e-16 ***
AGE               0.0203939  0.0000945  215.798  < 2e-16 ***
factor(SEX)2      0.2311632  0.0027298   84.680  < 2e-16 ***
factor(RACETHX)2  0.3308413  0.0040339   82.016  < 2e-16 ***
factor(RACETHX)3  0.0314258  0.0052518    5.984 2.18e-09 ***
factor(RACETHX)4 -0.1627420  0.0080253  -20.279  < 2e-16 ***
factor(RACETHX)5  0.3054364  0.0076710   39.817  < 2e-16 ***
factor(MARRY)2   -0.0509827  0.0051103   -9.976  < 2e-16 ***
factor(MARRY)3    0.1246967  0.0042123   29.603  < 2e-16 ***
factor(MARRY)4    0.2080775  0.0091784   22.670  < 2e-16 ***
factor(MARRY)5    0.1121028  0.0043083   26.020  < 2e-16 ***
factor(MARRY)6    0.3512127  0.0064622   54.349  < 2e-16 ***
factor(INS)2     -1.0026879  0.0073927 -135.633  < 2e-16 ***
factor(POVCAT)2  -0.0905525  0.0069590  -13.012  < 2e-16 ***
factor(POVCAT)3  -0.1360523  0.0049332  -27.579  < 2e-16 ***
factor(POVCAT)4  -0.1643384  0.0042911  -38.298  < 2e-16 ***
factor(POVCAT)5  -0.0670044  0.0042228  -15.867  < 2e-16 ***
factor(REGION)2  -0.1247595  0.0042184  -29.575  < 2e-16 ***
factor(REGION)3  -0.1065634  0.0038628  -27.587  < 2e-16 ***
factor(REGION)4  -0.1050677  0.0041670  -25.214  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1077008  on 58504  degrees of freedom
Residual deviance:  895192  on 58485  degrees of freedom
AIC: 1060952

Number of Fisher Scoring iterations: 6
```

6.d.

Model18_nb

```
> summary(model18_nb)

Call:
glm.nb(formula = OFFVST ~ AGE + factor(SEX) + factor(RACETHX) +
    factor(MARRY) + factor(INS) + factor(POVCAT) + factor(REGION),
    data = d_2018, init.theta = 0.5249852654, link = log)

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)       0.9845217  0.0496357  19.835  < 2e-16 ***
AGE               0.0217413  0.0006263  34.715  < 2e-16 ***
factor(SEX)2      0.2811767  0.0167002  16.837  < 2e-16 ***
factor(RACETHX)2  0.4177117  0.0224276  18.625  < 2e-16 ***
factor(RACETHX)3  0.0945060  0.0286591   3.298 0.000975 ***
factor(RACETHX)4 -0.1073252  0.0411804  -2.606 0.009155 **
factor(RACETHX)5  0.2816591  0.0449364   6.268 3.66e-10 ***
factor(MARRY)2   -0.0215230  0.0404189  -0.532 0.594380
factor(MARRY)3    0.1420865  0.0302090   4.703 2.56e-06 ***
factor(MARRY)4    0.3297591  0.0613880   5.372 7.80e-08 ***
factor(MARRY)5    0.1036501  0.0258179   4.015 5.95e-05 ***
factor(MARRY)6    0.5235727  0.0364585  14.361  < 2e-16 ***
factor(INS)2     -0.9434253  0.0284275 -33.187  < 2e-16 ***
factor(POVCAT)2  -0.0420460  0.0415011  -1.013 0.310998
factor(POVCAT)3  -0.0707133  0.0293237  -2.411 0.015888 *
factor(POVCAT)4  -0.1280536  0.0257290  -4.977 6.46e-07 ***
factor(POVCAT)5  -0.0346476  0.0263363  -1.316 0.188314
factor(REGION)2  -0.1537522  0.0277244  -5.546 2.93e-08 ***
factor(REGION)3  -0.1468531  0.0253465  -5.794 6.88e-09 ***
factor(REGION)4  -0.1161958  0.0272262  -4.268 1.97e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Negative Binomial(0.525) family taken to be 1)

    Null deviance: 39433  on 30239  degrees of freedom
Residual deviance: 34154  on 30220  degrees of freedom
AIC: 185812

Number of Fisher Scoring iterations: 1


              Theta:  0.52499
          Std. Err.:  0.00482

 2 x log-likelihood:  -185770.01700
```

Model19_nb

```
> summary(model19_nb)

Call:
glm.nb(formula = OFFVST ~ AGE + factor(SEX) + factor(RACETHX) +
    factor(MARRY) + factor(INS) + factor(POVCAT) + factor(REGION),
    data = d_2019, init.theta = 0.5188283997, link = log)

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)       1.2116968  0.0517802  23.401  < 2e-16 ***
AGE               0.0203363  0.0006458  31.492  < 2e-16 ***
factor(SEX)2      0.2436944  0.0173674  14.032  < 2e-16 ***
factor(RACETHX)2  0.3336387  0.0233833  14.268  < 2e-16 ***
factor(RACETHX)3 -0.0448468  0.0302047  -1.485  0.13761
factor(RACETHX)4 -0.0859685  0.0431153  -1.994  0.04616 *
factor(RACETHX)5  0.3268926  0.0477882   6.840 7.89e-12 ***
factor(MARRY)2    0.0536561  0.0409666   1.310  0.19028
factor(MARRY)3    0.1771860  0.0310808   5.701 1.19e-08 ***
factor(MARRY)4    0.2653075  0.0650895   4.076 4.58e-05 ***
factor(MARRY)5    0.0810752  0.0268716   3.017  0.00255 **
factor(MARRY)6    0.4381822  0.0383208  11.435  < 2e-16 ***
factor(INS)2     -1.0417010  0.0297550 -35.009  < 2e-16 ***
factor(POVCAT)2  -0.1224965  0.0434414  -2.820  0.00481 **
factor(POVCAT)3  -0.1583665  0.0312277  -5.071 3.95e-07 ***
factor(POVCAT)4  -0.1926520  0.0271547  -7.095 1.30e-12 ***
factor(POVCAT)5  -0.0507202  0.0275239  -1.843  0.06536 .
factor(REGION)2  -0.1441862  0.0287365  -5.018 5.23e-07 ***
factor(REGION)3  -0.1863144  0.0263251  -7.077 1.47e-12 ***
factor(REGION)4  -0.1387762  0.0279476  -4.966 6.85e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Negative Binomial(0.5188) family taken to be 1)

    Null deviance: 36857  on 28264  degrees of freedom
Residual deviance: 31919  on 28245  degrees of freedom
AIC: 174831

Number of Fisher Scoring iterations: 1


              Theta:  0.51883
          Std. Err.:  0.00490

 2 x log-likelihood:  -174788.80400
```

Model_nb

```
> summary(model_nb)

Call:
glm.nb(formula = OFFVST ~ AGE + factor(SEX) + factor(RACETHX) +
    factor(MARRY) + factor(INS) + factor(POVCAT) + factor(REGION),
    data = pooled, init.theta = 0.521575464, link = log)

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)       1.0981505  0.0358293  30.650  < 2e-16 ***
AGE               0.0210130  0.0004495  46.746  < 2e-16 ***
factor(SEX)2      0.2618944  0.0120409  21.750  < 2e-16 ***
factor(RACETHX)2  0.3765982  0.0161851  23.268  < 2e-16 ***
factor(RACETHX)3  0.0286349  0.0207912   1.377  0.16843
factor(RACETHX)4 -0.0945856  0.0297820  -3.176  0.00149 **
factor(RACETHX)5  0.3043329  0.0327405   9.295  < 2e-16 ***
factor(MARRY)2    0.0162045  0.0287789   0.563  0.57339
factor(MARRY)3    0.1589947  0.0216676   7.338 2.17e-13 ***
factor(MARRY)4    0.2983168  0.0446661   6.679 2.41e-11 ***
factor(MARRY)5    0.0914151  0.0186191   4.910 9.12e-07 ***
factor(MARRY)6    0.4796815  0.0264133  18.161  < 2e-16 ***
factor(INS)2     -0.9896604  0.0205516 -48.155  < 2e-16 ***
factor(POVCAT)2  -0.0805081  0.0300069  -2.683  0.00730 **
factor(POVCAT)3  -0.1114217  0.0213779  -5.212 1.87e-07 ***
factor(POVCAT)4  -0.1599967  0.0186777  -8.566  < 2e-16 ***
factor(POVCAT)5  -0.0420430  0.0190267  -2.210  0.02713 *
factor(REGION)2  -0.1495185  0.0199566  -7.492 6.77e-14 ***
factor(REGION)3  -0.1654895  0.0182624  -9.062  < 2e-16 ***
factor(REGION)4  -0.1286416  0.0194993  -6.597 4.19e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Negative Binomial(0.5216) family taken to be 1)

    Null deviance: 76249  on 58504  degrees of freedom
Residual deviance: 66074  on 58485  degrees of freedom
AIC: 360642

Number of Fisher Scoring iterations: 1


              Theta:  0.52158
          Std. Err.:  0.00343

 2 x log-likelihood:  -360600.39400
```

6.e.

2018년 데이터셋

```
> summary(model18_lin_predict$fit)
    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -5.347   5.629   8.839   9.645  13.865  25.567
> mae(d_2018$OFFVST, model18_lin_predict$fit)
[1] 8.979293
> mse(d_2018$OFFVST, model18_lin_predict$fit)
[1] 323.1008
> rmse(d_2018$OFFVST, model18_lin_predict$fit)
[1] 17.975
>
> model18_p_predict=predict(model18_p, se.fit = TRUE)
> summary(model18_p_predict$fit)
    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.110   1.714   2.076   2.094   2.565   3.621
> mae(d_2018$OFFVST, model18_p_predict$fit)
[1] 8.577611
> mse(d_2018$OFFVST, model18_p_predict$fit)
[1] 403.566
> rmse(d_2018$OFFVST, model18_p_predict$fit)
[1] 20.08895
>
> model18_nb_predict=predict(model18_nb, se.fit = TRUE)
> summary(model18_nb_predict$fit)
    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.2167  1.6961  2.0725  2.0887  2.5554  3.7905
> mae(d_2018$OFFVST, model18_nb_predict$fit)
[1] 8.56824
> mse(d_2018$OFFVST, model18_nb_predict$fit)
[1] 403.4781
> rmse(d_2018$OFFVST, model18_nb_predict$fit)
[1] 20.08676
```

2019년 데이터셋

```
> model19_lin_predict=predict(model19_lin, se.fit = TRUE)
> summary(model19_lin_predict$fit)
    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -4.941   5.844   9.136   9.917  14.245  24.930
> mae(d_2019$OFFVST, model19_lin_predict$fit)
[1] 9.279857
> mse(d_2019$OFFVST, model19_lin_predict$fit)
[1] 324.9502
> rmse(d_2019$OFFVST, model19_lin_predict$fit)
[1] 18.02638
>
> model19_p_predict=predict(model19_p, se.fit = TRUE)
> summary(model19_p_predict$fit)
    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.2222  1.7448  2.1150  2.1197  2.6016  3.5878
> mae(d_2019$OFFVST, model19_p_predict$fit)
[1] 8.837322
> mse(d_2019$OFFVST, model19_p_predict$fit)
[1] 410.1158
> rmse(d_2019$OFFVST, model19_p_predict$fit)
[1] 20.25132
>
> model19_nb_predict=predict(model19_nb, se.fit = TRUE)
> summary(model19_nb_predict$fit)
    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.1666  1.7373  2.1137  2.1144  2.5951  3.6745
> mae(d_2019$OFFVST, model19_nb_predict$fit)
[1] 8.829565
> mse(d_2019$OFFVST, model19_nb_predict$fit)
[1] 410.0218
> rmse(d_2019$OFFVST, model19_nb_predict$fit)
[1] 20.24899
>
```

Pooled 데이터셋

```
> model_lin_predict=predict(model_lin, se.fit = TRUE)
> summary(model_lin_predict$fit)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -5.191   5.726   8.973   9.776  14.045  25.325
> mae(pooled$OFFVST, model_lin_predict$fit)
[1] 9.123341
> mse(pooled$OFFVST, model_lin_predict$fit)
[1] 324.0727
> rmse(pooled$OFFVST, model_lin_predict$fit)
[1] 18.00202
>
> model_p_predict=predict(model_p, se.fit = TRUE)
> summary(model_p_predict$fit)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.1389  1.7292  2.0953  2.1070  2.5846  3.5963
> mae(pooled$OFFVST, model_p_predict$fit)
[1] 8.702988
> mse(pooled$OFFVST, model_p_predict$fit)
[1] 406.7398
> rmse(pooled$OFFVST, model_p_predict$fit)
[1] 20.16779
>
> model_nb_predict=predict(model_nb, se.fit = TRUE)
> summary(model_nb_predict$fit)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.1806  1.7159  2.0922  2.1018  2.5762  3.7406
> mae(pooled$OFFVST, model_nb_predict$fit)
[1] 8.694595
> mse(pooled$OFFVST, model_nb_predict$fit)
[1] 406.6513
> rmse(pooled$OFFVST, model_nb_predict$fit)
[1] 20.1656
>
```

7.a.

SF6D 결과 확인

```
> head(sf6d_data)
  ADHECR42 EQU42 ADDAYA42 ADPLMT42 ADMACC42 ADMLMT42 ADPAIN42 ADPEP42
1       NA     1       NA        2        2        2       NA       NA
2       NA    NA       NA       NA       NA       -1       NA       NA
3       NA    NA       NA       NA       NA       -1       NA       NA
4       NA    NA       NA       NA       NA       -1       NA       NA
5       NA    NA       NA       NA       NA       -1       NA       NA
6       NA     1       NA        2        2        2       NA        1
  ADBLUE42 ADSOCA42 PCS42 MCS42 ADLHLP31 ADLHLP42 ADLHLP53 ADL3MO31
1       NA       NA    NA    NA        2        2        2       NA
2       NA       NA    NA    NA        2        2        2       NA
3       NA       NA    NA    NA        2        2        2       NA
4       NA       NA    NA    NA        2        2        2       NA
5       NA       NA    NA    NA        2        2        2       NA
6        1        1 60.02 38.61        2        2        2       NA
  ADL3MO42 ADL3MO53 IADLHP31 IADLHP42 IADLHP53 IADL3M31 IADL3M42
1       NA       NA        2        2        2       NA       NA
2       NA       NA        2        2        2       NA       NA
3       NA       NA        2        2        2       NA       NA
4       NA       NA        2        2        2       NA       NA
5       NA       NA        2        2        2       NA       NA
6       NA       NA        2        2        2       NA       NA
  IADL3M53 VAS EQ5D SF12_p SF12_m PF RL SF PAIN MH VT sf6d_score
1       NA  -9    1  -9.00  -9.00 NA  4 NA   NA NA NA         NA
2       NA  -1   -1  -1.00  -1.00 NA NA NA   NA NA NA         NA
3       NA  -1   -1  -1.00  -1.00 NA NA NA   NA NA NA         NA
4       NA  -1   -1  -1.00  -1.00 NA NA NA   NA NA NA         NA
5       NA  -1   -1  -1.00  -1.00 NA NA NA   NA NA NA         NA
6       NA  -9    1  60.02  38.61 NA  4  1   NA  1  1         NA
```

ADL, IADL 결과 확인

```
> head(final_data)
  ADHECR42 EQU42 ADDAYA42 ADPLMT42 ADMACC42 ADMLMT42 ADPAIN42 ADPEP42
1       NA     1       NA        2        2        2       NA       NA
2       NA    NA       NA       NA       NA       -1       NA       NA
3       NA    NA       NA       NA       NA       -1       NA       NA
4       NA    NA       NA       NA       NA       -1       NA       NA
5       NA    NA       NA       NA       NA       -1       NA       NA
6       NA     1       NA        2        2        2       NA        1
  ADBLUE42 ADSOCA42 PCS42 MCS42 ADLHLP31 ADLHLP42 ADLHLP53 ADL3MO31
1       NA       NA    NA    NA        0        0        0        0
2       NA       NA    NA    NA        0        0        0        0
3       NA       NA    NA    NA        0        0        0        0
4       NA       NA    NA    NA        0        0        0        0
5       NA       NA    NA    NA        0        0        0        0
6        1        1 60.02 38.61        0        0        0        0
  ADL3MO42 ADL3MO53 IADLHP31 IADLHP42 IADLHP53 IADL3M31 IADL3M42
1        0        0        0        0        0        0        0
2        0        0        0        0        0        0        0
3        0        0        0        0        0        0        0
4        0        0        0        0        0        0        0
5        0        0        0        0        0        0        0
6        0        0        0        0        0        0        0
  IADL3M53 VAS EQ5D SF12_p SF12_m PF RL SF PAIN MH VT sf6d_score
1        0  -9    1  -9.00  -9.00 NA  4 NA   NA NA NA         NA
2        0  -1   -1  -1.00  -1.00 NA NA NA   NA NA NA         NA
3        0  -1   -1  -1.00  -1.00 NA NA NA   NA NA NA         NA
4        0  -1   -1  -1.00  -1.00 NA NA NA   NA NA NA         NA
5        0  -1   -1  -1.00  -1.00 NA NA NA   NA NA NA         NA
6        0  -9    1  60.02  38.61 NA  4  1   NA  1  1         NA
  adl_score iadl_score
1         0          0
2         0          0
3         0          0
4         0          0
5         0          0
6         0          0
```

결측값 개수 확인

```
> # 결측값 개수 확인
> total_rows = nrow(final_data)
> missing_sf6d=sum(is.na(final_data$sf6d_score))
> missing_adl=sum(is.na(final_data$adl_score))
> missing_iadl=sum(is.na(final_data$iadl_score))
> missing_VAS=sum(is.na(final_data$VAS))
> missing_EQ5D=sum(is.na(final_data$EQ5D))
> missing_PCS=sum(is.na(final_data$PCS42))
> missing_MCS=sum(is.na(final_data$MCS42))
> total_rows
[1] 39165
> missing_sf6d
[1] 24709
> missing_adl
[1] 1569
> missing_iadl
[1] 1650
> missing_VAS
[1] 0
> missing_EQ5D
[1] 0
> missing_PCS
[1] 15061
> missing_MCS
[1] 15061
```

Final

```
> final=final_data %>%
+   dplyr::select(VAS,EQ5D,sf6d_score,SF12_p,SF12_m,adl_score,iadl_score)
> head(final)
  VAS EQ5D sf6d_score SF12_p SF12_m adl_score iadl_score
1  -9    1         NA  -9.00  -9.00         0          0
2  -1   -1         NA  -1.00  -1.00         0          0
3  -1   -1         NA  -1.00  -1.00         0          0
4  -1   -1         NA  -1.00  -1.00         0          0
5  -1   -1         NA  -1.00  -1.00         0          0
6  -9    1         NA  60.02  38.61         0          0
```

7.b.

```
> print(summary_stats)
  mean_VAS   sd_VAS mean_EQ5D  sd_EQ5D mean_sf6d    sd_sf6d mean_SF12_p
1 2.770152 4.801735 0.0108707 1.331881 0.5447209 0.06832193       29.49
  sd_SF12_p mean_SF12_m sd_SF12_m   mean_adl    sd_adl mean_iadl
1  25.54017    30.84635  26.36463 0.09110012 0.6269851  0.163428
    sd_iadl
1 0.8337113
```

7.c.

```
> print(cor_matrix)
                      VAS          EQ5D    sf6d_score       SF12_p
VAS          1.0000000000  0.007945043  0.0005186331 -0.14773538
EQ5D         0.0079450431  1.000000000 -0.1034223056  0.18249922
sf6d_score   0.0005186331 -0.103422306  1.0000000000 -0.27299799
SF12_p      -0.1477353816  0.182499220 -0.2729979929  1.00000000
SF12_m       0.0029635533  0.112429398 -0.5468195411  0.08532713
adl_score    0.0513387291 -0.077554624  0.1130054220 -0.28034170
iadl_score   0.0747060035 -0.096798327  0.1533228027 -0.37496170
                  SF12_m   adl_score  iadl_score
VAS          0.002963553  0.05133873  0.07470600
EQ5D         0.112429398 -0.07755462 -0.09679833
sf6d_score  -0.546819541  0.11300542  0.15332280
SF12_p       0.085327131 -0.28034170 -0.37496170
SF12_m       1.000000000 -0.09106091 -0.12705441
adl_score   -0.091060905  1.00000000  0.68472078
iadl_score  -0.127054413  0.68472078  1.00000000
```

7.d.

Vas_model

```
> summary(vas_model)

Call:
glm(formula = ADHECR42 ~ AGE02X + factor(SEX) + factor(RACETHNX) +
    factor(MARRY02X) + factor(REGION02), family = "poisson",
    data = data)

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)        2.0207288  0.0138257 146.158  < 2e-16 ***
AGE02X             0.0016325  0.0001923   8.489  < 2e-16 ***
factor(SEX)2       0.0129669  0.0057314   2.262  0.02367 *
factor(RACETHNX)2  0.0021309  0.0104393   0.204  0.83826
factor(RACETHNX)3 -0.0290671  0.0168809  -1.722  0.08509 .
factor(RACETHNX)4  0.0031584  0.0078708   0.401  0.68821
factor(MARRY02X)2  0.0063674  0.0107648   0.592  0.55418
factor(MARRY02X)3 -0.0278991  0.0090255  -3.091  0.00199 **
factor(MARRY02X)4 -0.0180415  0.0184925  -0.976  0.32926
factor(MARRY02X)5  0.0018461  0.0080855   0.228  0.81939
factor(REGION02)2 -0.0037536  0.0088080  -0.426  0.67000
factor(REGION02)3 -0.0219915  0.0079880  -2.753  0.00590 **
factor(REGION02)4 -0.0414187  0.0087990  -4.707 2.51e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 8563.0  on 16702  degrees of freedom
Residual deviance: 8388.8  on 16690  degrees of freedom
  (결측으로 인하여 22462개의 관측치가 삭제되었습니다.)
AIC: 73571

Number of Fisher Scoring iterations: 4
```

Eq5d_model

```
> summary(eq5d_model)

Call:
lm(formula = EQU42 ~ AGE02X + factor(SEX) + factor(RACETHNX) +
    factor(MARRY02X) + factor(POVCAT02) + factor(REGION02), data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-1.40366 -0.07978  0.05053  0.13873  0.50056

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         0.9679020  0.0082011 118.021  < 2e-16 ***
AGE02X             -0.0043755  0.0001075 -40.708  < 2e-16 ***
factor(SEX)2       -0.0232166  0.0030370  -7.645 2.17e-14 ***
factor(RACETHNX)2  -0.0286553  0.0054215  -5.285 1.26e-07 ***
factor(RACETHNX)3   0.0033073  0.0085746   0.386 0.699714
factor(RACETHNX)4  -0.0259549  0.0041224  -6.296 3.11e-10 ***
factor(MARRY02X)2   0.0012068  0.0065524   0.184 0.853876
factor(MARRY02X)3  -0.0366307  0.0049505  -7.399 1.41e-13 ***
factor(MARRY02X)4  -0.0638250  0.0096985  -6.581 4.77e-11 ***
factor(MARRY02X)5  -0.0227643  0.0041890  -5.434 5.55e-08 ***
factor(POVCAT02)2   0.0392314  0.0076691   5.116 3.15e-07 ***
factor(POVCAT02)3   0.0748098  0.0054380  13.757  < 2e-16 ***
factor(POVCAT02)4   0.1098562  0.0048438  22.680  < 2e-16 ***
factor(POVCAT02)5   0.1523360  0.0049802  30.589  < 2e-16 ***
factor(REGION02)2  -0.0025638  0.0049988  -0.513 0.608038
factor(REGION02)3  -0.0171918  0.0045069  -3.815 0.000137 ***
factor(REGION02)4   0.0016108  0.0048715   0.331 0.740912
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2281 on 23773 degrees of freedom
  (결측으로 인하여 15375개의 관측치가 삭제되었습니다.)
Multiple R-squared:  0.1436,    Adjusted R-squared:  0.143
F-statistic: 249.1 on 16 and 23773 DF,  p-value: < 2.2e-16
```

Mae, mse, rmse

```
> cat("VAS 모델 - MAE:", mae_vas, "MSE:", mse_vas, "RMSE:", rmse_vas, "\n")
VAS 모델 - MAE: 1.368771 MSE: 3.320027 RMSE: 1.822094
> cat("EQ5D 모델 - MAE:", mae_eq5d, "MSE:", mse_eq5d, "RMSE:", rmse_eq5d,
"\n")
EQ5D 모델 - MAE: 0.1578429 MSE: 0.05198 RMSE: 0.2279912
>
```