Stat 331 Project

2022-12-05

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
library(glmnet)
```

```
## Loading required package: Matrix
```

```
## Loaded glmnet 4.1-6
```

```
set.seed(1000)
exp <- load("/Users/pivaldhingra/Downloads/exposome_NA.RData")
exposome <- exposomeNA[order(exposomeNA$ID),]
covariates <- covariatesNA[order(covariatesNA$ID),]
phenotype <- phenotypeNA[order(phenotypeNA$ID),]

## read in data
a <- phenotype[5]
b <- covariates[12]
c <- exposome[,86:105]
expm <- data.frame(a,b,c)
expm <- expm[sample(nrow(expm)),]

## bounds for model selection
M0 <- lm(hs_correct_raven ~ 1,data=expm)  # minimal model
Mfull <- lm(hs_correct_raven ~.,data=expm) # full model

length(coef(Mfull))
```

```
## [1] 22
```

```
anyNA(coef(Mfull))
```

```
## [1] FALSE
```

```
df.penalty <- 2

## forward
system.time({
  Mfwd <- step(object = M0, # base model
               scope = list(lower = M0, upper = Mfull), # smallest and largest model
               trace = 1, # trace prints out information
               direction = "forward" )
})
```

```
## Start:  AIC=4810.55
## hs_correct_raven ~ 1
##
##                       Df Sum of Sq   RSS    AIC
## + hs_child_age_None    1   25099.7 28422 3995.4
## + hs_cs_c_Log2         1   11300.4 42221 4506.4
## + hs_hg_c_Log2         1    3283.3 50238 4730.8
## + hs_cu_m_Log2         1    2678.2 50844 4746.3
## + hs_as_c_Log2         1    2447.8 51074 4752.1
## + hs_cd_m_Log2         1    1619.3 51902 4772.9
## + hs_cs_m_Log2         1    1608.0 51914 4773.2
## + hs_co_c_Log2         1    1268.1 52254 4781.6
## + hs_as_m_Log2         1    1231.0 52291 4782.5
## + hs_pb_c_Log2         1    1171.9 52350 4784.0
## + hs_mo_c_Log2         1     927.7 52594 4790.0
## + hs_pb_m_Log2         1     738.1 52784 4794.6
## + hs_cu_c_Log2         1     723.3 52798 4795.0
## + hs_hg_m_Log2         1     472.4 53049 4801.1
## + hs_co_m_Log2         1     253.5 53268 4806.4
## + hs_mn_m_Log2         1     207.8 53314 4807.5
## + hs_tl_mdich_None     1     110.2 53412 4809.9
## <none>                             53522 4810.5
## + hs_mn_c_Log2         1      16.3 53505 4812.2
## + hs_tl_cdich_None     1       4.2 53518 4812.4
## + hs_cd_c_Log2         1       2.7 53519 4812.5
## + hs_mo_m_Log2         1       0.3 53521 4812.5
##
## Step:  AIC=3995.45
## hs_correct_raven ~ hs_child_age_None
##
##                       Df Sum of Sq   RSS    AIC
## + hs_cs_m_Log2         1    863.04 27559 3957.6
## + hs_pb_c_Log2         1    667.70 27754 3966.8
## + hs_co_m_Log2         1    482.96 27939 3975.3
## + hs_cu_c_Log2         1    454.41 27968 3976.6
## + hs_cd_m_Log2         1    347.56 28074 3981.6
## + hs_as_c_Log2         1    307.82 28114 3983.4
## + hs_co_c_Log2         1    186.83 28235 3988.9
## + hs_cs_c_Log2         1    169.42 28253 3989.7
## + hs_hg_m_Log2         1    141.74 28280 3991.0
## + hs_mn_m_Log2         1     49.68 28372 3995.2
## <none>                             28422 3995.4
## + hs_hg_c_Log2         1     41.41 28381 3995.6
## + hs_cu_m_Log2         1     39.19 28383 3995.7
## + hs_mo_c_Log2         1     39.01 28383 3995.7
## + hs_tl_cdich_None     1     38.72 28383 3995.7
## + hs_pb_m_Log2         1     28.33 28394 3996.2
## + hs_tl_mdich_None     1     23.91 28398 3996.4
## + hs_cd_c_Log2         1     11.34 28411 3996.9
## + hs_mn_c_Log2         1     11.03 28411 3996.9
## + hs_mo_m_Log2         1      3.68 28418 3997.3
## + hs_as_m_Log2         1      0.82 28421 3997.4
##
## Step:  AIC=3957.64
```

```
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2
##
##                     Df Sum of Sq   RSS    AIC
## + hs_pb_c_Log2       1    470.09 27089 3937.4
## + hs_cu_c_Log2       1    340.56 27218 3943.6
## + hs_cd_m_Log2       1    309.07 27250 3945.1
## + hs_as_c_Log2       1    262.27 27297 3947.3
## + hs_co_c_Log2       1    177.15 27382 3951.3
## + hs_co_m_Log2       1    116.51 27442 3954.2
## + hs_mo_c_Log2       1     76.20 27483 3956.1
## + hs_as_m_Log2       1     43.12 27516 3957.6
## <none>                           27559 3957.6
## + hs_pb_m_Log2       1     42.64 27516 3957.6
## + hs_cd_c_Log2       1     42.36 27517 3957.7
## + hs_cu_m_Log2       1     41.56 27517 3957.7
## + hs_hg_c_Log2       1     37.15 27522 3957.9
## + hs_cs_c_Log2       1     30.26 27529 3958.2
## + hs_tl_cdich_None   1     30.18 27529 3958.2
## + hs_tl_mdich_None   1     24.34 27535 3958.5
## + hs_mo_m_Log2       1     22.59 27536 3958.6
## + hs_hg_m_Log2       1     19.24 27540 3958.7
## + hs_mn_c_Log2       1     16.09 27543 3958.9
## + hs_mn_m_Log2       1      1.11 27558 3959.6
##
## Step:  AIC=3937.43
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2
##
##                     Df Sum of Sq   RSS    AIC
## + hs_as_c_Log2       1   258.432 26830 3927.1
## + hs_cd_m_Log2       1   232.218 26857 3928.3
## + hs_cu_c_Log2       1   201.989 26887 3929.8
## + hs_co_c_Log2       1   113.730 26975 3934.0
## + hs_co_m_Log2       1    85.577 27003 3935.3
## + hs_mo_c_Log2       1    81.627 27007 3935.5
## + hs_as_m_Log2       1    73.397 27016 3935.9
## <none>                           27089 3937.4
## + hs_hg_c_Log2       1    41.358 27048 3937.5
## + hs_cd_c_Log2       1    37.353 27052 3937.6
## + hs_mo_m_Log2       1    34.640 27054 3937.8
## + hs_cs_c_Log2       1    32.425 27056 3937.9
## + hs_tl_cdich_None   1    28.901 27060 3938.0
## + hs_tl_mdich_None   1    20.124 27069 3938.5
## + hs_cu_m_Log2       1    18.294 27071 3938.6
## + hs_hg_m_Log2       1    13.159 27076 3938.8
## + hs_mn_m_Log2       1    10.558 27078 3938.9
## + hs_mn_c_Log2       1     4.907 27084 3939.2
## + hs_pb_m_Log2       1     1.377 27088 3939.4
##
## Step:  AIC=3927.05
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2 +
##     hs_as_c_Log2
##
##                     Df Sum of Sq   RSS    AIC
## + hs_cd_m_Log2       1   273.163 26557 3915.8
```

```
## + hs_cu_c_Log2      1   258.938 26572 3916.5
## + hs_hg_c_Log2      1   200.305 26630 3919.4
## + hs_as_m_Log2      1   122.840 26708 3923.1
## + hs_co_c_Log2      1   114.806 26716 3923.5
## + hs_co_m_Log2      1   106.849 26724 3923.9
## + hs_mo_c_Log2      1    99.889 26731 3924.2
## + hs_mo_m_Log2      1    50.706 26780 3926.6
## <none>                          26830 3927.1
## + hs_cd_c_Log2      1    34.696 26796 3927.4
## + hs_cu_m_Log2      1    23.315 26807 3927.9
## + hs_tl_cdich_None  1    19.821 26811 3928.1
## + hs_mn_m_Log2      1    16.281 26814 3928.3
## + hs_tl_mdich_None  1    14.887 26816 3928.3
## + hs_mn_c_Log2      1     8.432 26822 3928.6
## + hs_cs_c_Log2      1     6.514 26824 3928.7
## + hs_pb_m_Log2      1     1.032 26830 3929.0
## + hs_hg_m_Log2      1     0.490 26830 3929.0
##
## Step:  AIC=3915.84
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2 +
##     hs_as_c_Log2 + hs_cd_m_Log2
##
##                    Df Sum of Sq   RSS    AIC
## + hs_cu_c_Log2      1   224.593 26333 3906.9
## + hs_hg_c_Log2      1   195.040 26362 3908.3
## + hs_mo_c_Log2      1   109.006 26448 3912.5
## + hs_as_m_Log2      1   105.065 26452 3912.7
## + hs_co_c_Log2      1    90.429 26467 3913.4
## + hs_co_m_Log2      1    88.029 26469 3913.6
## <none>                          26557 3915.8
## + hs_cd_c_Log2      1    38.065 26519 3916.0
## + hs_mo_m_Log2      1    33.249 26524 3916.2
## + hs_mn_m_Log2      1    27.454 26530 3916.5
## + hs_tl_mdich_None  1    20.114 26537 3916.9
## + hs_tl_cdich_None  1    19.122 26538 3916.9
## + hs_mn_c_Log2      1    10.257 26547 3917.3
## + hs_cs_c_Log2      1     7.226 26550 3917.5
## + hs_cu_m_Log2      1     6.620 26551 3917.5
## + hs_hg_m_Log2      1     4.273 26553 3917.6
## + hs_pb_m_Log2      1     2.924 26554 3917.7
##
## Step:  AIC=3906.88
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2 +
##     hs_as_c_Log2 + hs_cd_m_Log2 + hs_cu_c_Log2
##
##                    Df Sum of Sq   RSS    AIC
## + hs_hg_c_Log2      1   164.441 26168 3900.8
## + hs_as_m_Log2      1    96.109 26237 3904.2
## + hs_mo_c_Log2      1    93.284 26239 3904.3
## + hs_co_m_Log2      1    81.516 26251 3904.9
## + hs_co_c_Log2      1    76.894 26256 3905.1
## + hs_cd_c_Log2      1    47.679 26285 3906.5
## <none>                          26333 3906.9
## + hs_mo_m_Log2      1    28.442 26304 3907.5
```

```
## + hs_mn_m_Log2      1    25.932 26307 3907.6
## + hs_cs_c_Log2      1    23.971 26309 3907.7
## + hs_tl_mdich_None  1    19.482 26313 3907.9
## + hs_tl_cdich_None  1    12.508 26320 3908.3
## + hs_hg_m_Log2      1     9.016 26324 3908.4
## + hs_pb_m_Log2      1     3.668 26329 3908.7
## + hs_mn_c_Log2      1     3.241 26330 3908.7
## + hs_cu_m_Log2      1     1.574 26331 3908.8
##
## Step:  AIC=3900.79
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2 +
##     hs_as_c_Log2 + hs_cd_m_Log2 + hs_cu_c_Log2 + hs_hg_c_Log2
##
##                    Df Sum of Sq   RSS    AIC
## + hs_mo_c_Log2      1   100.055 26068 3897.8
## + hs_co_c_Log2      1    76.176 26092 3899.0
## + hs_as_m_Log2      1    64.243 26104 3899.6
## + hs_co_m_Log2      1    52.783 26116 3900.2
## + hs_cd_c_Log2      1    51.690 26117 3900.2
## + hs_cs_c_Log2      1    44.149 26124 3900.6
## <none>                          26168 3900.8
## + hs_hg_m_Log2      1    36.625 26132 3901.0
## + hs_mn_m_Log2      1    28.229 26140 3901.4
## + hs_mo_m_Log2      1    21.923 26146 3901.7
## + hs_tl_mdich_None  1    16.781 26152 3902.0
## + hs_tl_cdich_None  1    10.282 26158 3902.3
## + hs_pb_m_Log2      1     4.478 26164 3902.6
## + hs_cu_m_Log2      1     3.710 26165 3902.6
## + hs_mn_c_Log2      1     3.637 26165 3902.6
##
## Step:  AIC=3897.84
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2 +
##     hs_as_c_Log2 + hs_cd_m_Log2 + hs_cu_c_Log2 + hs_hg_c_Log2 +
##     hs_mo_c_Log2
##
##                    Df Sum of Sq   RSS    AIC
## + hs_co_c_Log2      1    62.830 26005 3896.7
## + hs_as_m_Log2      1    58.313 26010 3897.0
## + hs_co_m_Log2      1    57.816 26010 3897.0
## + hs_cd_c_Log2      1    49.182 26019 3897.4
## + hs_cs_c_Log2      1    45.934 26022 3897.6
## <none>                          26068 3897.8
## + hs_hg_m_Log2      1    33.626 26035 3898.2
## + hs_mo_m_Log2      1    21.491 26047 3898.8
## + hs_mn_m_Log2      1    21.449 26047 3898.8
## + hs_tl_mdich_None  1    17.247 26051 3899.0
## + hs_tl_cdich_None  1     7.117 26061 3899.5
## + hs_cu_m_Log2      1     3.783 26064 3899.7
## + hs_pb_m_Log2      1     3.312 26065 3899.7
## + hs_mn_c_Log2      1     2.012 26066 3899.7
##
## Step:  AIC=3896.73
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2 +
##     hs_as_c_Log2 + hs_cd_m_Log2 + hs_cu_c_Log2 + hs_hg_c_Log2 +
```

```
##     hs_mo_c_Log2 + hs_co_c_Log2
##
##                     Df Sum of Sq   RSS    AIC
## + hs_as_m_Log2       1    59.369 25946 3895.8
## + hs_co_m_Log2       1    51.368 25954 3896.2
## + hs_cs_c_Log2       1    51.165 25954 3896.2
## + hs_cd_c_Log2       1    41.108 25964 3896.7
## <none>                           26005 3896.7
## + hs_hg_m_Log2       1    34.529 25971 3897.0
## + hs_mn_m_Log2       1    21.204 25984 3897.7
## + hs_mo_m_Log2       1    19.298 25986 3897.8
## + hs_tl_mdich_None   1    17.194 25988 3897.9
## + hs_tl_cdich_None   1     6.059 25999 3898.4
## + hs_pb_m_Log2       1     4.441 26001 3898.5
## + hs_cu_m_Log2       1     3.337 26002 3898.6
## + hs_mn_c_Log2       1     0.378 26005 3898.7
##
## Step:  AIC=3895.78
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2 +
##     hs_as_c_Log2 + hs_cd_m_Log2 + hs_cu_c_Log2 + hs_hg_c_Log2 +
##     hs_mo_c_Log2 + hs_co_c_Log2 + hs_as_m_Log2
##
##                     Df Sum of Sq   RSS    AIC
## + hs_hg_m_Log2       1    64.946 25881 3894.5
## + hs_cs_c_Log2       1    59.485 25886 3894.8
## + hs_co_m_Log2       1    52.787 25893 3895.1
## <none>                           25946 3895.8
## + hs_cd_c_Log2       1    39.467 25907 3895.8
## + hs_tl_mdich_None   1    16.578 25930 3897.0
## + hs_mn_m_Log2       1    14.660 25931 3897.0
## + hs_mo_m_Log2       1    11.823 25934 3897.2
## + hs_tl_cdich_None   1     6.921 25939 3897.4
## + hs_cu_m_Log2       1     3.766 25942 3897.6
## + hs_pb_m_Log2       1     2.444 25944 3897.7
## + hs_mn_c_Log2       1     0.368 25946 3897.8
##
## Step:  AIC=3894.54
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2 +
##     hs_as_c_Log2 + hs_cd_m_Log2 + hs_cu_c_Log2 + hs_hg_c_Log2 +
##     hs_mo_c_Log2 + hs_co_c_Log2 + hs_as_m_Log2 + hs_hg_m_Log2
##
##                     Df Sum of Sq   RSS    AIC
## + hs_co_m_Log2       1    53.586 25828 3893.9
## + hs_cs_c_Log2       1    48.774 25832 3894.1
## <none>                           25881 3894.5
## + hs_cd_c_Log2       1    39.184 25842 3894.6
## + hs_mn_m_Log2       1    18.486 25863 3895.6
## + hs_tl_mdich_None   1    17.628 25864 3895.7
## + hs_mo_m_Log2       1    14.426 25867 3895.8
## + hs_tl_cdich_None   1     5.596 25876 3896.3
## + hs_cu_m_Log2       1     1.571 25880 3896.5
## + hs_pb_m_Log2       1     1.387 25880 3896.5
## + hs_mn_c_Log2       1     0.020 25881 3896.5
##
```

```
## Step:  AIC=3893.87
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2 +
##     hs_as_c_Log2 + hs_cd_m_Log2 + hs_cu_c_Log2 + hs_hg_c_Log2 +
##     hs_mo_c_Log2 + hs_co_c_Log2 + hs_as_m_Log2 + hs_hg_m_Log2 +
##     hs_co_m_Log2
##
##                     Df Sum of Sq   RSS    AIC
## + hs_cs_c_Log2       1    50.741 25777 3893.3
## + hs_cd_c_Log2       1    43.570 25784 3893.7
## <none>                           25828 3893.9
## + hs_mn_m_Log2       1    34.699 25793 3894.1
## + hs_tl_mdich_None   1    15.244 25812 3895.1
## + hs_mo_m_Log2       1    12.836 25815 3895.2
## + hs_cu_m_Log2       1     3.754 25824 3895.7
## + hs_tl_cdich_None   1     3.672 25824 3895.7
## + hs_pb_m_Log2       1     0.472 25827 3895.8
## + hs_mn_c_Log2       1     0.023 25828 3895.9
##
## Step:  AIC=3893.33
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2 +
##     hs_as_c_Log2 + hs_cd_m_Log2 + hs_cu_c_Log2 + hs_hg_c_Log2 +
##     hs_mo_c_Log2 + hs_co_c_Log2 + hs_as_m_Log2 + hs_hg_m_Log2 +
##     hs_co_m_Log2 + hs_cs_c_Log2
##
##                     Df Sum of Sq   RSS    AIC
## + hs_cd_c_Log2       1    46.538 25730 3893.0
## + hs_mn_m_Log2       1    46.078 25731 3893.0
## <none>                           25777 3893.3
## + hs_mo_m_Log2       1    14.874 25762 3894.6
## + hs_tl_mdich_None   1    12.847 25764 3894.7
## + hs_tl_cdich_None   1     4.652 25772 3895.1
## + hs_cu_m_Log2       1     2.776 25774 3895.2
## + hs_pb_m_Log2       1     1.041 25776 3895.3
## + hs_mn_c_Log2       1     0.622 25776 3895.3
##
## Step:  AIC=3892.99
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2 +
##     hs_as_c_Log2 + hs_cd_m_Log2 + hs_cu_c_Log2 + hs_hg_c_Log2 +
##     hs_mo_c_Log2 + hs_co_c_Log2 + hs_as_m_Log2 + hs_hg_m_Log2 +
##     hs_co_m_Log2 + hs_cs_c_Log2 + hs_cd_c_Log2
##
##                     Df Sum of Sq   RSS    AIC
## + hs_mn_m_Log2       1    40.878 25689 3892.9
## <none>                           25730 3893.0
## + hs_mo_m_Log2       1    12.770 25717 3894.4
## + hs_tl_mdich_None   1    10.444 25720 3894.5
## + hs_tl_cdich_None   1     6.254 25724 3894.7
## + hs_cu_m_Log2       1     2.904 25727 3894.8
## + hs_pb_m_Log2       1     0.888 25729 3894.9
## + hs_mn_c_Log2       1     0.156 25730 3895.0
##
## Step:  AIC=3892.94
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2 +
##     hs_as_c_Log2 + hs_cd_m_Log2 + hs_cu_c_Log2 + hs_hg_c_Log2 +
```

```
##     hs_mo_c_Log2 + hs_co_c_Log2 + hs_as_m_Log2 + hs_hg_m_Log2 +
##     hs_co_m_Log2 + hs_cs_c_Log2 + hs_cd_c_Log2 + hs_mn_m_Log2
##
##                     Df Sum of Sq   RSS    AIC
## <none>                           25689 3892.9
## + hs_mo_m_Log2       1   13.7055 25676 3894.3
## + hs_tl_mdich_None   1   11.3186 25678 3894.4
## + hs_tl_cdich_None   1    5.9937 25683 3894.6
## + hs_cu_m_Log2       1    2.7974 25686 3894.8
## + hs_mn_c_Log2       1    1.7825 25688 3894.9
## + hs_pb_m_Log2       1    0.1704 25689 3894.9
```

```
##    user  system elapsed
##   0.115   0.008   0.124
```

```
## backward
system.time({
  Mback <- step(object = Mfull, # base model
                scope = list(lower = M0, upper = Mfull),
                direction = "backward", trace = 1)
})
```

```
## Start:  AIC=3903.35
## hs_correct_raven ~ hs_child_age_None + hs_as_c_Log2 + hs_as_m_Log2 +
##     hs_cd_c_Log2 + hs_cd_m_Log2 + hs_co_c_Log2 + hs_co_m_Log2 +
##     hs_cs_c_Log2 + hs_cs_m_Log2 + hs_cu_c_Log2 + hs_cu_m_Log2 +
##     hs_hg_c_Log2 + hs_hg_m_Log2 + hs_mn_c_Log2 + hs_mn_m_Log2 +
##     hs_mo_c_Log2 + hs_mo_m_Log2 + hs_pb_c_Log2 + hs_pb_m_Log2 +
##     hs_tl_cdich_None + hs_tl_mdich_None
##
##                      Df Sum of Sq   RSS    AIC
## - hs_pb_m_Log2        1       0.3 25658 3901.4
## - hs_cu_m_Log2        1       1.4 25659 3901.4
## - hs_mn_c_Log2        1       1.5 25659 3901.4
## - hs_tl_cdich_None    1       4.6 25662 3901.6
## - hs_tl_mdich_None    1       9.1 25667 3901.8
## - hs_mo_m_Log2        1      10.8 25669 3901.9
## - hs_cd_c_Log2        1      37.5 25695 3903.2
## <none>                           25658 3903.4
## - hs_mn_m_Log2        1      42.7 25700 3903.5
## - hs_co_c_Log2        1      50.1 25708 3903.9
## - hs_hg_m_Log2        1      55.7 25713 3904.2
## - hs_cs_c_Log2        1      65.9 25724 3904.7
## - hs_mo_c_Log2        1      70.5 25728 3904.9
## - hs_co_m_Log2        1      73.8 25732 3905.1
## - hs_as_m_Log2        1      75.7 25733 3905.2
## - hs_hg_c_Log2        1     162.3 25820 3909.5
## - hs_cu_c_Log2        1     189.7 25847 3910.9
## - hs_cd_m_Log2        1     213.7 25871 3912.1
## - hs_pb_c_Log2        1     243.8 25902 3913.6
## - hs_cs_m_Log2        1     297.4 25955 3916.2
## - hs_as_c_Log2        1     476.2 26134 3925.1
## - hs_child_age_None   1   11576.6 37234 4382.1
##
## Step:  AIC=3901.37
```

```
## hs_correct_raven ~ hs_child_age_None + hs_as_c_Log2 + hs_as_m_Log2 +
##     hs_cd_c_Log2 + hs_cd_m_Log2 + hs_co_c_Log2 + hs_co_m_Log2 +
##     hs_cs_c_Log2 + hs_cs_m_Log2 + hs_cu_c_Log2 + hs_cu_m_Log2 +
##     hs_hg_c_Log2 + hs_hg_m_Log2 + hs_mn_c_Log2 + hs_mn_m_Log2 +
##     hs_mo_c_Log2 + hs_mo_m_Log2 + hs_pb_c_Log2 + hs_tl_cdich_None +
##     hs_tl_mdich_None
##
##                      Df Sum of Sq   RSS    AIC
## - hs_cu_m_Log2        1       1.4 25659 3899.4
## - hs_mn_c_Log2        1       1.5 25660 3899.4
## - hs_tl_cdich_None    1       4.6 25663 3899.6
## - hs_tl_mdich_None    1       9.0 25667 3899.8
## - hs_mo_m_Log2        1      10.8 25669 3899.9
## - hs_cd_c_Log2        1      37.5 25696 3901.3
## <none>                             25658 3901.4
## - hs_mn_m_Log2        1      43.6 25702 3901.6
## - hs_co_c_Log2        1      49.9 25708 3901.9
## - hs_hg_m_Log2        1      56.3 25714 3902.2
## - hs_cs_c_Log2        1      65.7 25724 3902.7
## - hs_mo_c_Log2        1      70.8 25729 3902.9
## - hs_co_m_Log2        1      75.1 25733 3903.1
## - hs_as_m_Log2        1      76.9 25735 3903.2
## - hs_hg_c_Log2        1     162.0 25820 3907.5
## - hs_cu_c_Log2        1     189.5 25848 3908.9
## - hs_cd_m_Log2        1     216.2 25874 3910.2
## - hs_pb_c_Log2        1     251.2 25909 3911.9
## - hs_cs_m_Log2        1     300.5 25959 3914.4
## - hs_as_c_Log2        1     475.9 26134 3923.1
## - hs_child_age_None   1   11586.3 37244 4380.4
##
## Step:  AIC=3899.44
## hs_correct_raven ~ hs_child_age_None + hs_as_c_Log2 + hs_as_m_Log2 +
##     hs_cd_c_Log2 + hs_cd_m_Log2 + hs_co_c_Log2 + hs_co_m_Log2 +
##     hs_cs_c_Log2 + hs_cs_m_Log2 + hs_cu_c_Log2 + hs_hg_c_Log2 +
##     hs_hg_m_Log2 + hs_mn_c_Log2 + hs_mn_m_Log2 + hs_mo_c_Log2 +
##     hs_mo_m_Log2 + hs_pb_c_Log2 + hs_tl_cdich_None + hs_tl_mdich_None
##
##                      Df Sum of Sq   RSS    AIC
## - hs_mn_c_Log2        1       1.6 25661 3897.5
## - hs_tl_cdich_None    1       5.2 25665 3897.7
## - hs_tl_mdich_None    1       9.0 25668 3897.9
## - hs_mo_m_Log2        1      11.3 25671 3898.0
## - hs_cd_c_Log2        1      37.4 25697 3899.3
## <none>                             25659 3899.4
## - hs_mn_m_Log2        1      43.8 25703 3899.6
## - hs_co_c_Log2        1      50.1 25710 3900.0
## - hs_hg_m_Log2        1      58.1 25718 3900.4
## - hs_cs_c_Log2        1      66.8 25726 3900.8
## - hs_mo_c_Log2        1      70.5 25730 3901.0
## - hs_co_m_Log2        1      73.9 25733 3901.2
## - hs_as_m_Log2        1      76.9 25736 3901.3
## - hs_hg_c_Log2        1     161.5 25821 3905.5
## - hs_cu_c_Log2        1     194.2 25854 3907.2
## - hs_cd_m_Log2        1     225.5 25885 3908.7
```

```
## - hs_pb_c_Log2       1     254.9 25914 3910.2
## - hs_cs_m_Log2       1     300.0 25960 3912.4
## - hs_as_c_Log2       1     474.5 26134 3921.1
## - hs_child_age_None  1   11934.0 37593 4390.5
##
## Step:  AIC=3897.52
## hs_correct_raven ~ hs_child_age_None + hs_as_c_Log2 + hs_as_m_Log2 +
##     hs_cd_c_Log2 + hs_cd_m_Log2 + hs_co_c_Log2 + hs_co_m_Log2 +
##     hs_cs_c_Log2 + hs_cs_m_Log2 + hs_cu_c_Log2 + hs_hg_c_Log2 +
##     hs_hg_m_Log2 + hs_mn_m_Log2 + hs_mo_c_Log2 + hs_mo_m_Log2 +
##     hs_pb_c_Log2 + hs_tl_cdich_None + hs_tl_mdich_None
##
##                      Df Sum of Sq   RSS    AIC
## - hs_tl_cdich_None   1       5.0 25666 3895.8
## - hs_tl_mdich_None   1       9.1 25670 3896.0
## - hs_mo_m_Log2       1      11.5 25673 3896.1
## - hs_cd_c_Log2       1      38.6 25700 3897.5
## <none>                           25661 3897.5
## - hs_mn_m_Log2       1      42.3 25703 3897.6
## - hs_co_c_Log2       1      52.4 25714 3898.2
## - hs_hg_m_Log2       1      60.5 25722 3898.6
## - hs_cs_c_Log2       1      65.4 25726 3898.8
## - hs_mo_c_Log2       1      71.6 25733 3899.1
## - hs_co_m_Log2       1      72.7 25734 3899.2
## - hs_as_m_Log2       1      77.6 25739 3899.4
## - hs_hg_c_Log2       1     161.6 25823 3903.6
## - hs_cu_c_Log2       1     198.6 25860 3905.5
## - hs_cd_m_Log2       1     224.5 25886 3906.8
## - hs_pb_c_Log2       1     257.3 25918 3908.4
## - hs_cs_m_Log2       1     299.2 25960 3910.5
## - hs_as_c_Log2       1     473.3 26134 3919.1
## - hs_child_age_None  1   11934.5 37596 4388.6
##
## Step:  AIC=3895.77
## hs_correct_raven ~ hs_child_age_None + hs_as_c_Log2 + hs_as_m_Log2 +
##     hs_cd_c_Log2 + hs_cd_m_Log2 + hs_co_c_Log2 + hs_co_m_Log2 +
##     hs_cs_c_Log2 + hs_cs_m_Log2 + hs_cu_c_Log2 + hs_hg_c_Log2 +
##     hs_hg_m_Log2 + hs_mn_m_Log2 + hs_mo_c_Log2 + hs_mo_m_Log2 +
##     hs_pb_c_Log2 + hs_tl_mdich_None
##
##                      Df Sum of Sq   RSS    AIC
## - hs_tl_mdich_None   1       9.5 25676 3894.3
## - hs_mo_m_Log2       1      11.9 25678 3894.4
## - hs_cd_c_Log2       1      37.2 25703 3895.6
## <none>                           25666 3895.8
## - hs_mn_m_Log2       1      42.6 25709 3895.9
## - hs_co_c_Log2       1      53.3 25719 3896.5
## - hs_hg_m_Log2       1      62.0 25728 3896.9
## - hs_cs_c_Log2       1      64.2 25730 3897.0
## - hs_mo_c_Log2       1      73.9 25740 3897.5
## - hs_co_m_Log2       1      75.2 25741 3897.6
## - hs_as_m_Log2       1      76.9 25743 3897.6
## - hs_hg_c_Log2       1     163.0 25829 3901.9
## - hs_cu_c_Log2       1     201.7 25868 3903.9
```

```
## - hs_cd_m_Log2       1     224.6 25891 3905.0
## - hs_pb_c_Log2       1     256.5 25923 3906.6
## - hs_cs_m_Log2       1     298.5 25965 3908.7
## - hs_as_c_Log2       1     482.7 26149 3917.8
## - hs_child_age_None  1   11932.1 37598 4386.7
##
## Step:  AIC=3894.25
## hs_correct_raven ~ hs_child_age_None + hs_as_c_Log2 + hs_as_m_Log2 +
##     hs_cd_c_Log2 + hs_cd_m_Log2 + hs_co_c_Log2 + hs_co_m_Log2 +
##     hs_cs_c_Log2 + hs_cs_m_Log2 + hs_cu_c_Log2 + hs_hg_c_Log2 +
##     hs_hg_m_Log2 + hs_mn_m_Log2 + hs_mo_c_Log2 + hs_mo_m_Log2 +
##     hs_pb_c_Log2
##
##                      Df Sum of Sq   RSS    AIC
## - hs_mo_m_Log2        1      13.7 25689 3892.9
## - hs_cd_c_Log2        1      39.2 25715 3894.2
## <none>                             25676 3894.3
## - hs_mn_m_Log2        1      41.8 25717 3894.4
## - hs_co_c_Log2        1      53.0 25729 3894.9
## - hs_hg_m_Log2        1      61.1 25737 3895.3
## - hs_cs_c_Log2        1      66.7 25742 3895.6
## - hs_mo_c_Log2        1      73.7 25749 3896.0
## - hs_as_m_Log2        1      76.9 25753 3896.1
## - hs_co_m_Log2        1      77.3 25753 3896.1
## - hs_hg_c_Log2        1     164.4 25840 3900.5
## - hs_cu_c_Log2        1     202.5 25878 3902.4
## - hs_cd_m_Log2        1     220.6 25896 3903.3
## - hs_pb_c_Log2        1     259.0 25935 3905.2
## - hs_cs_m_Log2        1     296.4 25972 3907.1
## - hs_as_c_Log2        1     489.1 26165 3916.6
## - hs_child_age_None   1   11941.2 37617 4385.3
##
## Step:  AIC=3892.94
## hs_correct_raven ~ hs_child_age_None + hs_as_c_Log2 + hs_as_m_Log2 +
##     hs_cd_c_Log2 + hs_cd_m_Log2 + hs_co_c_Log2 + hs_co_m_Log2 +
##     hs_cs_c_Log2 + hs_cs_m_Log2 + hs_cu_c_Log2 + hs_hg_c_Log2 +
##     hs_hg_m_Log2 + hs_mn_m_Log2 + hs_mo_c_Log2 + hs_pb_c_Log2
##
##                      Df Sum of Sq   RSS    AIC
## <none>                             25689 3892.9
## - hs_mn_m_Log2        1      40.9 25730 3893.0
## - hs_cd_c_Log2        1      41.3 25731 3893.0
## - hs_co_c_Log2        1      54.4 25744 3893.7
## - hs_hg_m_Log2        1      58.9 25748 3893.9
## - hs_cs_c_Log2        1      64.5 25754 3894.2
## - hs_mo_c_Log2        1      73.7 25763 3894.6
## - hs_co_m_Log2        1      79.1 25768 3894.9
## - hs_as_m_Log2        1      85.5 25775 3895.2
## - hs_hg_c_Log2        1     165.9 25855 3899.3
## - hs_cu_c_Log2        1     204.3 25894 3901.2
## - hs_cd_m_Log2        1     229.2 25919 3902.4
## - hs_pb_c_Log2        1     253.1 25942 3903.6
## - hs_cs_m_Log2        1     291.4 25981 3905.5
## - hs_as_c_Log2        1     485.2 26175 3915.1
```

```
## - hs_child_age_None  1   12009.2 37699 4386.1
```

```
##    user  system elapsed
##   0.099   0.005   0.105
```

```
## stepwise (both directions)
system.time({
  Mstep <- step(object = M0,
                scope = list(lower = M0, upper = Mfull),
                direction = "both", trace = 1)
})
```

```
## Start:  AIC=4810.55
## hs_correct_raven ~ 1
##
##                       Df Sum of Sq   RSS    AIC
## + hs_child_age_None    1   25099.7 28422 3995.4
## + hs_cs_c_Log2         1   11300.4 42221 4506.4
## + hs_hg_c_Log2         1    3283.3 50238 4730.8
## + hs_cu_m_Log2         1    2678.2 50844 4746.3
## + hs_as_c_Log2         1    2447.8 51074 4752.1
## + hs_cd_m_Log2         1    1619.3 51902 4772.9
## + hs_cs_m_Log2         1    1608.0 51914 4773.2
## + hs_co_c_Log2         1    1268.1 52254 4781.6
## + hs_as_m_Log2         1    1231.0 52291 4782.5
## + hs_pb_c_Log2         1    1171.9 52350 4784.0
## + hs_mo_c_Log2         1     927.7 52594 4790.0
## + hs_pb_m_Log2         1     738.1 52784 4794.6
## + hs_cu_c_Log2         1     723.3 52798 4795.0
## + hs_hg_m_Log2         1     472.4 53049 4801.1
## + hs_co_m_Log2         1     253.5 53268 4806.4
## + hs_mn_m_Log2         1     207.8 53314 4807.5
## + hs_tl_mdich_None     1     110.2 53412 4809.9
## <none>                             53522 4810.5
## + hs_mn_c_Log2         1      16.3 53505 4812.2
## + hs_tl_cdich_None     1       4.2 53518 4812.4
## + hs_cd_c_Log2         1       2.7 53519 4812.5
## + hs_mo_m_Log2         1       0.3 53521 4812.5
##
## Step:  AIC=3995.45
## hs_correct_raven ~ hs_child_age_None
##
##                       Df Sum of Sq   RSS    AIC
## + hs_cs_m_Log2         1     863.0 27559 3957.6
## + hs_pb_c_Log2         1     667.7 27754 3966.8
## + hs_co_m_Log2         1     483.0 27939 3975.3
## + hs_cu_c_Log2         1     454.4 27968 3976.6
## + hs_cd_m_Log2         1     347.6 28074 3981.6
## + hs_as_c_Log2         1     307.8 28114 3983.4
## + hs_co_c_Log2         1     186.8 28235 3988.9
## + hs_cs_c_Log2         1     169.4 28253 3989.7
## + hs_hg_m_Log2         1     141.7 28280 3991.0
## + hs_mn_m_Log2         1      49.7 28372 3995.2
## <none>                             28422 3995.4
## + hs_hg_c_Log2         1      41.4 28381 3995.6
```

```
## + hs_cu_m_Log2       1      39.2 28383 3995.7
## + hs_mo_c_Log2       1      39.0 28383 3995.7
## + hs_tl_cdich_None   1      38.7 28383 3995.7
## + hs_pb_m_Log2       1      28.3 28394 3996.2
## + hs_tl_mdich_None   1      23.9 28398 3996.4
## + hs_cd_c_Log2       1      11.3 28411 3996.9
## + hs_mn_c_Log2       1      11.0 28411 3996.9
## + hs_mo_m_Log2       1       3.7 28418 3997.3
## + hs_as_m_Log2       1       0.8 28421 3997.4
## - hs_child_age_None  1   25099.7 53522 4810.5
##
## Step:  AIC=3957.64
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2
##
##                     Df Sum of Sq   RSS    AIC
## + hs_pb_c_Log2       1     470.1 27089 3937.4
## + hs_cu_c_Log2       1     340.6 27218 3943.6
## + hs_cd_m_Log2       1     309.1 27250 3945.1
## + hs_as_c_Log2       1     262.3 27297 3947.3
## + hs_co_c_Log2       1     177.1 27382 3951.3
## + hs_co_m_Log2       1     116.5 27442 3954.2
## + hs_mo_c_Log2       1      76.2 27483 3956.1
## + hs_as_m_Log2       1      43.1 27516 3957.6
## <none>                           27559 3957.6
## + hs_pb_m_Log2       1      42.6 27516 3957.6
## + hs_cd_c_Log2       1      42.4 27517 3957.7
## + hs_cu_m_Log2       1      41.6 27517 3957.7
## + hs_hg_c_Log2       1      37.2 27522 3957.9
## + hs_cs_c_Log2       1      30.3 27529 3958.2
## + hs_tl_cdich_None   1      30.2 27529 3958.2
## + hs_tl_mdich_None   1      24.3 27535 3958.5
## + hs_mo_m_Log2       1      22.6 27536 3958.6
## + hs_hg_m_Log2       1      19.2 27540 3958.7
## + hs_mn_c_Log2       1      16.1 27543 3958.9
## + hs_mn_m_Log2       1       1.1 27558 3959.6
## - hs_cs_m_Log2       1     863.0 28422 3995.4
## - hs_child_age_None  1   24354.7 51914 4773.2
##
## Step:  AIC=3937.43
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2
##
##                     Df Sum of Sq   RSS    AIC
## + hs_as_c_Log2       1     258.4 26830 3927.1
## + hs_cd_m_Log2       1     232.2 26857 3928.3
## + hs_cu_c_Log2       1     202.0 26887 3929.8
## + hs_co_c_Log2       1     113.7 26975 3934.0
## + hs_co_m_Log2       1      85.6 27003 3935.3
## + hs_mo_c_Log2       1      81.6 27007 3935.5
## + hs_as_m_Log2       1      73.4 27016 3935.9
## <none>                           27089 3937.4
## + hs_hg_c_Log2       1      41.4 27048 3937.5
## + hs_cd_c_Log2       1      37.4 27052 3937.6
## + hs_mo_m_Log2       1      34.6 27054 3937.8
## + hs_cs_c_Log2       1      32.4 27056 3937.9
```

```
## + hs_tl_cdich_None   1      28.9 27060 3938.0
## + hs_tl_mdich_None   1      20.1 27069 3938.5
## + hs_cu_m_Log2       1      18.3 27071 3938.6
## + hs_hg_m_Log2       1      13.2 27076 3938.8
## + hs_mn_m_Log2       1      10.6 27078 3938.9
## + hs_mn_c_Log2       1       4.9 27084 3939.2
## + hs_pb_m_Log2       1       1.4 27088 3939.4
## - hs_pb_c_Log2       1     470.1 27559 3957.6
## - hs_cs_m_Log2       1     665.4 27754 3966.8
## - hs_child_age_None  1   24016.3 51105 4754.9
##
## Step:  AIC=3927.05
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2 +
##     hs_as_c_Log2
##
##                     Df Sum of Sq   RSS    AIC
## + hs_cd_m_Log2       1     273.2 26557 3915.8
## + hs_cu_c_Log2       1     258.9 26572 3916.5
## + hs_hg_c_Log2       1     200.3 26630 3919.4
## + hs_as_m_Log2       1     122.8 26708 3923.1
## + hs_co_c_Log2       1     114.8 26716 3923.5
## + hs_co_m_Log2       1     106.8 26724 3923.9
## + hs_mo_c_Log2       1      99.9 26731 3924.2
## + hs_mo_m_Log2       1      50.7 26780 3926.6
## <none>                           26830 3927.1
## + hs_cd_c_Log2       1      34.7 26796 3927.4
## + hs_cu_m_Log2       1      23.3 26807 3927.9
## + hs_tl_cdich_None   1      19.8 26811 3928.1
## + hs_mn_m_Log2       1      16.3 26814 3928.3
## + hs_tl_mdich_None   1      14.9 26816 3928.3
## + hs_mn_c_Log2       1       8.4 26822 3928.6
## + hs_cs_c_Log2       1       6.5 26824 3928.7
## + hs_pb_m_Log2       1       1.0 26829 3929.0
## + hs_hg_m_Log2       1       0.5 26830 3929.0
## - hs_as_c_Log2       1     258.4 27089 3937.4
## - hs_pb_c_Log2       1     466.3 27297 3947.3
## - hs_cs_m_Log2       1     627.1 27458 3954.9
## - hs_child_age_None  1   22086.1 48917 4700.4
##
## Step:  AIC=3915.84
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2 +
##     hs_as_c_Log2 + hs_cd_m_Log2
##
##                     Df Sum of Sq   RSS    AIC
## + hs_cu_c_Log2       1     224.6 26333 3906.9
## + hs_hg_c_Log2       1     195.0 26362 3908.3
## + hs_mo_c_Log2       1     109.0 26448 3912.5
## + hs_as_m_Log2       1     105.1 26452 3912.7
## + hs_co_c_Log2       1      90.4 26467 3913.4
## + hs_co_m_Log2       1      88.0 26469 3913.6
## <none>                           26557 3915.8
## + hs_cd_c_Log2       1      38.1 26519 3916.0
## + hs_mo_m_Log2       1      33.2 26524 3916.2
## + hs_mn_m_Log2       1      27.5 26530 3916.5
```

```
## + hs_tl_mdich_None   1      20.1 26537 3916.9
## + hs_tl_cdich_None   1      19.1 26538 3916.9
## + hs_mn_c_Log2       1      10.3 26547 3917.3
## + hs_cs_c_Log2       1       7.2 26550 3917.5
## + hs_cu_m_Log2       1       6.6 26551 3917.5
## + hs_hg_m_Log2       1       4.3 26553 3917.6
## + hs_pb_m_Log2       1       2.9 26554 3917.7
## - hs_cd_m_Log2       1     273.2 26830 3927.1
## - hs_as_c_Log2       1     299.4 26857 3928.3
## - hs_pb_c_Log2       1     383.5 26941 3932.3
## - hs_cs_m_Log2       1     607.9 27165 3943.1
## - hs_child_age_None  1   20938.6 47496 4664.3
##
## Step:  AIC=3906.88
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2 +
##     hs_as_c_Log2 + hs_cd_m_Log2 + hs_cu_c_Log2
##
##                     Df Sum of Sq   RSS    AIC
## + hs_hg_c_Log2       1     164.4 26168 3900.8
## + hs_as_m_Log2       1      96.1 26237 3904.2
## + hs_mo_c_Log2       1      93.3 26239 3904.3
## + hs_co_m_Log2       1      81.5 26251 3904.9
## + hs_co_c_Log2       1      76.9 26256 3905.1
## + hs_cd_c_Log2       1      47.7 26285 3906.5
## <none>                           26333 3906.9
## + hs_mo_m_Log2       1      28.4 26304 3907.5
## + hs_mn_m_Log2       1      25.9 26307 3907.6
## + hs_cs_c_Log2       1      24.0 26309 3907.7
## + hs_tl_mdich_None   1      19.5 26313 3907.9
## + hs_tl_cdich_None   1      12.5 26320 3908.3
## + hs_hg_m_Log2       1       9.0 26324 3908.4
## + hs_pb_m_Log2       1       3.7 26329 3908.7
## + hs_mn_c_Log2       1       3.2 26329 3908.7
## + hs_cu_m_Log2       1       1.6 26331 3908.8
## - hs_cu_c_Log2       1     224.6 26557 3915.8
## - hs_cd_m_Log2       1     238.8 26572 3916.5
## - hs_pb_c_Log2       1     259.7 26592 3917.5
## - hs_as_c_Log2       1     352.9 26686 3922.1
## - hs_cs_m_Log2       1     551.4 26884 3931.6
## - hs_child_age_None  1   20781.2 47114 4655.9
##
## Step:  AIC=3900.79
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2 +
##     hs_as_c_Log2 + hs_cd_m_Log2 + hs_cu_c_Log2 + hs_hg_c_Log2
##
##                     Df Sum of Sq   RSS    AIC
## + hs_mo_c_Log2       1     100.1 26068 3897.8
## + hs_co_c_Log2       1      76.2 26092 3899.0
## + hs_as_m_Log2       1      64.2 26104 3899.6
## + hs_co_m_Log2       1      52.8 26116 3900.2
## + hs_cd_c_Log2       1      51.7 26117 3900.2
## + hs_cs_c_Log2       1      44.1 26124 3900.6
## <none>                           26168 3900.8
## + hs_hg_m_Log2       1      36.6 26132 3901.0
```

```
## + hs_mn_m_Log2       1      28.2 26140 3901.4
## + hs_mo_m_Log2       1      21.9 26146 3901.7
## + hs_tl_mdich_None   1      16.8 26152 3902.0
## + hs_tl_cdich_None   1      10.3 26158 3902.3
## + hs_pb_m_Log2       1       4.5 26164 3902.6
## + hs_cu_m_Log2       1       3.7 26165 3902.6
## + hs_mn_c_Log2       1       3.6 26165 3902.6
## - hs_hg_c_Log2       1     164.4 26333 3906.9
## - hs_cu_c_Log2       1     194.0 26362 3908.3
## - hs_cd_m_Log2       1     236.5 26405 3910.4
## - hs_pb_c_Log2       1     272.6 26441 3912.2
## - hs_as_c_Log2       1     498.6 26667 3923.2
## - hs_cs_m_Log2       1     533.7 26702 3924.9
## - hs_child_age_None  1   19432.1 45600 4615.8
##
## Step:  AIC=3897.84
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2 +
##     hs_as_c_Log2 + hs_cd_m_Log2 + hs_cu_c_Log2 + hs_hg_c_Log2 +
##     hs_mo_c_Log2
##
##                      Df Sum of Sq   RSS    AIC
## + hs_co_c_Log2        1      62.8 26005 3896.7
## + hs_as_m_Log2        1      58.3 26010 3897.0
## + hs_co_m_Log2        1      57.8 26010 3897.0
## + hs_cd_c_Log2        1      49.2 26019 3897.4
## + hs_cs_c_Log2        1      45.9 26022 3897.6
## <none>                            26068 3897.8
## + hs_hg_m_Log2        1      33.6 26035 3898.2
## + hs_mo_m_Log2        1      21.5 26047 3898.8
## + hs_mn_m_Log2        1      21.4 26047 3898.8
## + hs_tl_mdich_None    1      17.2 26051 3899.0
## + hs_tl_cdich_None    1       7.1 26061 3899.5
## + hs_cu_m_Log2        1       3.8 26064 3899.7
## + hs_pb_m_Log2        1       3.3 26065 3899.7
## + hs_mn_c_Log2        1       2.0 26066 3899.7
## - hs_mo_c_Log2        1     100.1 26168 3900.8
## - hs_hg_c_Log2        1     171.2 26239 3904.3
## - hs_cu_c_Log2        1     178.4 26247 3904.6
## - hs_cd_m_Log2        1     245.6 26314 3908.0
## - hs_pb_c_Log2        1     279.8 26348 3909.6
## - hs_as_c_Log2        1     524.2 26592 3921.5
## - hs_cs_m_Log2        1     567.9 26636 3923.7
## - hs_child_age_None   1   18593.2 44661 4590.9
##
## Step:  AIC=3896.73
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2 +
##     hs_as_c_Log2 + hs_cd_m_Log2 + hs_cu_c_Log2 + hs_hg_c_Log2 +
##     hs_mo_c_Log2 + hs_co_c_Log2
##
##                      Df Sum of Sq   RSS    AIC
## + hs_as_m_Log2        1      59.4 25946 3895.8
## + hs_co_m_Log2        1      51.4 25954 3896.2
## + hs_cs_c_Log2        1      51.2 25954 3896.2
## + hs_cd_c_Log2        1      41.1 25964 3896.7
```

```
## <none>                           26005 3896.7
## + hs_hg_m_Log2       1      34.5 25971 3897.0
## + hs_mn_m_Log2       1      21.2 25984 3897.7
## + hs_mo_m_Log2       1      19.3 25986 3897.8
## - hs_co_c_Log2       1      62.8 26068 3897.8
## + hs_tl_mdich_None   1      17.2 25988 3897.9
## + hs_tl_cdich_None   1       6.1 25999 3898.4
## + hs_pb_m_Log2       1       4.4 26001 3898.5
## + hs_cu_m_Log2       1       3.3 26002 3898.6
## + hs_mn_c_Log2       1       0.4 26005 3898.7
## - hs_mo_c_Log2       1      86.7 26092 3899.0
## - hs_cu_c_Log2       1     168.6 26174 3903.1
## - hs_hg_c_Log2       1     170.1 26175 3903.1
## - hs_cd_m_Log2       1     226.4 26232 3905.9
## - hs_pb_c_Log2       1     249.1 26254 3907.0
## - hs_as_c_Log2       1     519.2 26525 3920.2
## - hs_cs_m_Log2       1     570.0 26575 3922.7
## - hs_child_age_None  1   18196.5 44202 4579.6
##
## Step:  AIC=3895.78
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2 +
##     hs_as_c_Log2 + hs_cd_m_Log2 + hs_cu_c_Log2 + hs_hg_c_Log2 +
##     hs_mo_c_Log2 + hs_co_c_Log2 + hs_as_m_Log2
##
##                     Df Sum of Sq   RSS    AIC
## + hs_hg_m_Log2       1      64.9 25881 3894.5
## + hs_cs_c_Log2       1      59.5 25887 3894.8
## + hs_co_m_Log2       1      52.8 25893 3895.1
## <none>                           25946 3895.8
## + hs_cd_c_Log2       1      39.5 25907 3895.8
## - hs_as_m_Log2       1      59.4 26005 3896.7
## + hs_tl_mdich_None   1      16.6 25929 3897.0
## - hs_co_c_Log2       1      63.9 26010 3897.0
## + hs_mn_m_Log2       1      14.7 25931 3897.0
## + hs_mo_m_Log2       1      11.8 25934 3897.2
## + hs_tl_cdich_None   1       6.9 25939 3897.4
## + hs_cu_m_Log2       1       3.8 25942 3897.6
## + hs_pb_m_Log2       1       2.4 25944 3897.7
## + hs_mn_c_Log2       1       0.4 25946 3897.8
## - hs_mo_c_Log2       1      81.1 26027 3897.8
## - hs_hg_c_Log2       1     138.4 26084 3900.6
## - hs_cu_c_Log2       1     165.2 26111 3902.0
## - hs_cd_m_Log2       1     214.3 26160 3904.4
## - hs_pb_c_Log2       1     270.3 26216 3907.2
## - hs_as_c_Log2       1     541.7 26488 3920.5
## - hs_cs_m_Log2       1     625.8 26572 3924.5
## - hs_child_age_None  1   18178.9 44125 4579.3
##
## Step:  AIC=3894.54
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2 +
##     hs_as_c_Log2 + hs_cd_m_Log2 + hs_cu_c_Log2 + hs_hg_c_Log2 +
##     hs_mo_c_Log2 + hs_co_c_Log2 + hs_as_m_Log2 + hs_hg_m_Log2
##
##                     Df Sum of Sq   RSS    AIC
```

```
## + hs_co_m_Log2       1      53.6 25827 3893.9
## + hs_cs_c_Log2       1      48.8 25832 3894.1
## <none>                           25881 3894.5
## + hs_cd_c_Log2       1      39.2 25842 3894.6
## + hs_mn_m_Log2       1      18.5 25863 3895.6
## + hs_tl_mdich_None   1      17.6 25863 3895.7
## - hs_hg_m_Log2       1      64.9 25946 3895.8
## - hs_co_c_Log2       1      65.5 25947 3895.8
## + hs_mo_m_Log2       1      14.4 25867 3895.8
## + hs_tl_cdich_None   1       5.6 25875 3896.3
## - hs_mo_c_Log2       1      75.7 25957 3896.3
## + hs_cu_m_Log2       1       1.6 25880 3896.5
## + hs_pb_m_Log2       1       1.4 25880 3896.5
## + hs_mn_c_Log2       1       0.0 25881 3896.5
## - hs_as_m_Log2       1      89.8 25971 3897.0
## - hs_hg_c_Log2       1     171.7 26053 3901.1
## - hs_cu_c_Log2       1     173.9 26055 3901.2
## - hs_cd_m_Log2       1     229.3 26110 3903.9
## - hs_pb_c_Log2       1     262.6 26144 3905.6
## - hs_cs_m_Log2       1     515.0 26396 3918.0
## - hs_as_c_Log2       1     519.1 26400 3918.2
## - hs_child_age_None  1   18219.4 44100 4580.6
##
## Step:  AIC=3893.87
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2 +
##     hs_as_c_Log2 + hs_cd_m_Log2 + hs_cu_c_Log2 + hs_hg_c_Log2 +
##     hs_mo_c_Log2 + hs_co_c_Log2 + hs_as_m_Log2 + hs_hg_m_Log2 +
##     hs_co_m_Log2
##
##                     Df Sum of Sq   RSS    AIC
## + hs_cs_c_Log2       1      50.7 25777 3893.3
## + hs_cd_c_Log2       1      43.6 25784 3893.7
## <none>                           25827 3893.9
## + hs_mn_m_Log2       1      34.7 25793 3894.1
## - hs_co_m_Log2       1      53.6 25881 3894.5
## - hs_co_c_Log2       1      58.8 25886 3894.8
## + hs_tl_mdich_None   1      15.2 25812 3895.1
## - hs_hg_m_Log2       1      65.7 25893 3895.1
## + hs_mo_m_Log2       1      12.8 25815 3895.2
## + hs_cu_m_Log2       1       3.8 25824 3895.7
## + hs_tl_cdich_None   1       3.7 25824 3895.7
## + hs_pb_m_Log2       1       0.5 25827 3895.8
## + hs_mn_c_Log2       1       0.0 25827 3895.9
## - hs_mo_c_Log2       1      80.4 25908 3895.9
## - hs_as_m_Log2       1      91.7 25919 3896.4
## - hs_hg_c_Log2       1     143.1 25971 3899.0
## - hs_cu_c_Log2       1     171.7 25999 3900.4
## - hs_cd_m_Log2       1     217.4 26045 3902.7
## - hs_pb_c_Log2       1     248.5 26076 3904.2
## - hs_cs_m_Log2       1     330.5 26158 3908.3
## - hs_as_c_Log2       1     519.3 26347 3917.6
## - hs_child_age_None  1   17805.6 43633 4568.8
##
## Step:  AIC=3893.33
```

```
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2 +
##     hs_as_c_Log2 + hs_cd_m_Log2 + hs_cu_c_Log2 + hs_hg_c_Log2 +
##     hs_mo_c_Log2 + hs_co_c_Log2 + hs_as_m_Log2 + hs_hg_m_Log2 +
##     hs_co_m_Log2 + hs_cs_c_Log2
##
##                      Df Sum of Sq   RSS    AIC
## + hs_cd_c_Log2        1      46.5 25730 3893.0
## + hs_mn_m_Log2        1      46.1 25731 3893.0
## <none>                             25777 3893.3
## - hs_cs_c_Log2        1      50.7 25827 3893.9
## - hs_hg_m_Log2        1      54.8 25832 3894.1
## - hs_co_m_Log2        1      55.6 25832 3894.1
## - hs_co_c_Log2        1      63.7 25840 3894.5
## + hs_mo_m_Log2        1      14.9 25762 3894.6
## + hs_tl_mdich_None    1      12.8 25764 3894.7
## + hs_tl_cdich_None    1       4.7 25772 3895.1
## + hs_cu_m_Log2        1       2.8 25774 3895.2
## + hs_pb_m_Log2        1       1.0 25776 3895.3
## + hs_mn_c_Log2        1       0.6 25776 3895.3
## - hs_mo_c_Log2        1      81.8 25859 3895.4
## - hs_as_m_Log2        1      97.8 25875 3896.2
## - hs_hg_c_Log2        1     158.7 25935 3899.3
## - hs_cu_c_Log2        1     193.2 25970 3901.0
## - hs_cd_m_Log2        1     213.9 25991 3902.0
## - hs_pb_c_Log2        1     245.6 26022 3903.6
## - hs_cs_m_Log2        1     262.0 26039 3904.4
## - hs_as_c_Log2        1     478.3 26255 3915.1
## - hs_child_age_None   1   12022.4 37799 4385.5
##
## Step:  AIC=3892.99
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2 +
##     hs_as_c_Log2 + hs_cd_m_Log2 + hs_cu_c_Log2 + hs_hg_c_Log2 +
##     hs_mo_c_Log2 + hs_co_c_Log2 + hs_as_m_Log2 + hs_hg_m_Log2 +
##     hs_co_m_Log2 + hs_cs_c_Log2 + hs_cd_c_Log2
##
##                      Df Sum of Sq   RSS    AIC
## + hs_mn_m_Log2        1      40.9 25689 3892.9
## <none>                             25730 3893.0
## - hs_cd_c_Log2        1      46.5 25777 3893.3
## - hs_cs_c_Log2        1      53.7 25784 3893.7
## - hs_hg_m_Log2        1      54.2 25784 3893.7
## - hs_co_c_Log2        1      54.9 25785 3893.7
## - hs_co_m_Log2        1      60.2 25790 3894.0
## + hs_mo_m_Log2        1      12.8 25717 3894.4
## + hs_tl_mdich_None    1      10.4 25720 3894.5
## + hs_tl_cdich_None    1       6.3 25724 3894.7
## + hs_cu_m_Log2        1       2.9 25727 3894.8
## + hs_pb_m_Log2        1       0.9 25729 3894.9
## + hs_mn_c_Log2        1       0.2 25730 3895.0
## - hs_mo_c_Log2        1      80.7 25811 3895.0
## - hs_as_m_Log2        1      95.7 25826 3895.8
## - hs_hg_c_Log2        1     162.1 25892 3899.1
## - hs_cu_c_Log2        1     203.1 25933 3901.1
## - hs_cd_m_Log2        1     217.2 25947 3901.8
```

```
## - hs_pb_c_Log2       1     240.1 25970 3903.0
## - hs_cs_m_Log2       1     274.5 26005 3904.7
## - hs_as_c_Log2       1     477.6 26208 3914.7
## - hs_child_age_None  1   12032.8 37763 4386.3
##
## Step:  AIC=3892.94
## hs_correct_raven ~ hs_child_age_None + hs_cs_m_Log2 + hs_pb_c_Log2 +
##     hs_as_c_Log2 + hs_cd_m_Log2 + hs_cu_c_Log2 + hs_hg_c_Log2 +
##     hs_mo_c_Log2 + hs_co_c_Log2 + hs_as_m_Log2 + hs_hg_m_Log2 +
##     hs_co_m_Log2 + hs_cs_c_Log2 + hs_cd_c_Log2 + hs_mn_m_Log2
##
##                     Df Sum of Sq   RSS    AIC
## <none>                           25689 3892.9
## - hs_mn_m_Log2       1      40.9 25730 3893.0
## - hs_cd_c_Log2       1      41.3 25731 3893.0
## - hs_co_c_Log2       1      54.4 25744 3893.7
## - hs_hg_m_Log2       1      58.9 25748 3893.9
## - hs_cs_c_Log2       1      64.5 25754 3894.2
## + hs_mo_m_Log2       1      13.7 25676 3894.3
## + hs_tl_mdich_None   1      11.3 25678 3894.4
## + hs_tl_cdich_None   1       6.0 25683 3894.6
## - hs_mo_c_Log2       1      73.7 25763 3894.6
## + hs_cu_m_Log2       1       2.8 25687 3894.8
## + hs_mn_c_Log2       1       1.8 25688 3894.9
## - hs_co_m_Log2       1      79.1 25768 3894.9
## + hs_pb_m_Log2       1       0.2 25689 3894.9
## - hs_as_m_Log2       1      85.5 25775 3895.2
## - hs_hg_c_Log2       1     165.9 25855 3899.3
## - hs_cu_c_Log2       1     204.3 25894 3901.2
## - hs_cd_m_Log2       1     229.2 25919 3902.4
## - hs_pb_c_Log2       1     253.1 25942 3903.6
## - hs_cs_m_Log2       1     291.4 25981 3905.5
## - hs_as_c_Log2       1     485.2 26175 3915.1
## - hs_child_age_None  1   12009.2 37699 4386.1
```

```
##    user  system elapsed
##   0.156   0.008   0.168
```

```
## compare the three different models
beta.fwd <- coef(Mfwd)
beta.back <- coef(Mback)
beta.step <- coef(Mstep)
c(fwd = length(beta.fwd), back = length(beta.back),
  step = length(beta.step)) # number of coefficients in each
```

```
##  fwd back step
##   16   16   16
```

```
# check if models are nested
names(beta.fwd)[!names(beta.fwd) %in% names(beta.back)]
```

```
## character(0)
```

```
names(beta.back)[!names(beta.back) %in% names(beta.fwd)]
```

```
## character(0)
```

```
## AIC
n <- nrow(expm)
ll_fwd <- -n/2 * (1 + log(sum(resid(Mfwd)^2)/n) + log(2*pi))
aic_fwd <- -2*ll_fwd + 2*(n - Mfwd$df + 1) # total number of parameters includes sigma
aic_fwd - AIC(Mfwd)
```

```
## [1] 68.24673
```

```
aic_step <- AIC(Mstep)
aic_back <- AIC(Mback)

aic_all <- round(c(aic_fwd,aic_step,aic_back),2)
names(aic_all) <- c("FWD","Step","Back")
aic_all
```

```
##     FWD    Step    Back
## 7626.89 7558.64 7558.64
```

```
## BIC
bic_fwd <- -2*ll_fwd + log(n)*(n - Mfwd$df + 1) # total number of parameters includes sigma
bic_fwd - BIC(Mfwd)
```

```
## [1] 120.0868
```

```
bic_step <- BIC(Mstep)
bic_back <- BIC(Mback)

bic_all <- round(c(bic_fwd,bic_step,bic_back),2)
names(bic_all) <- c("FWD","Step","Back")
bic_all
```

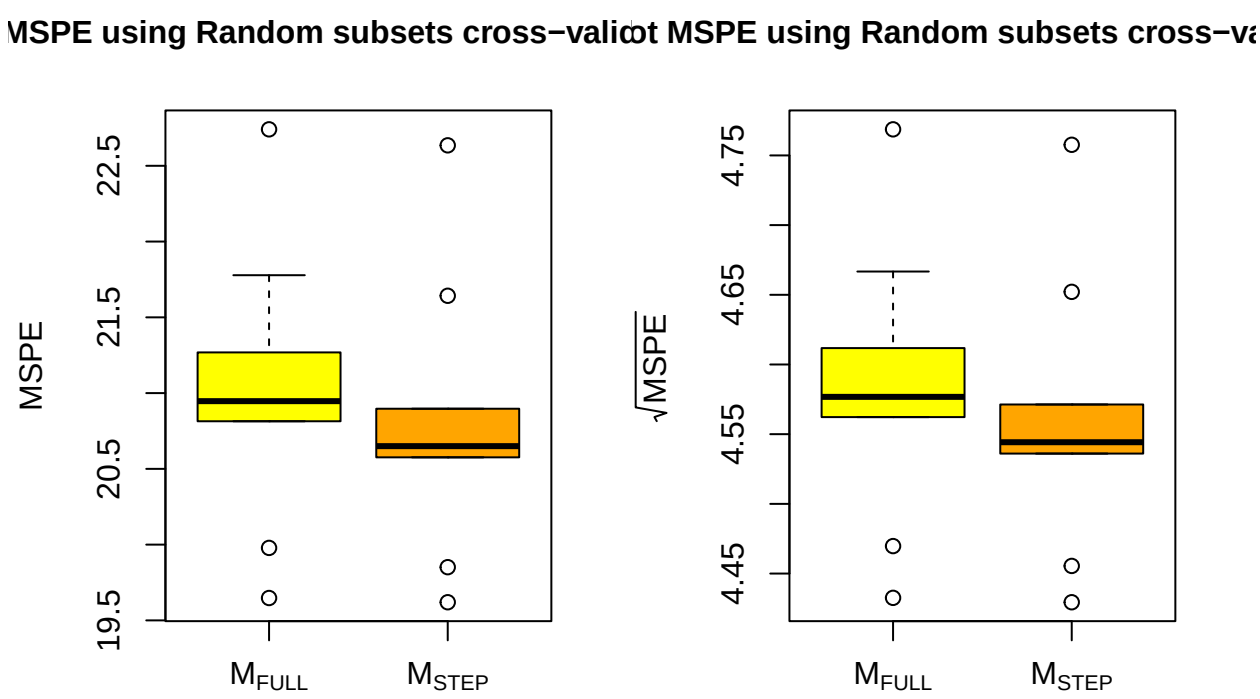
```
##     FWD    Step    Back
## 7766.50 7646.41 7646.41
```

```
## Random subsets cross-validation
# compare Mfull to Mstep
M1 <- Mfull
M2 <- Mstep
Mnames <- expression(M[FULL], M[STEP])

# number of cross-validation replications
nreps <- 1e3

ntot <- nrow(expm) # total number of observations
ntrain <- 800 # for fitting MLE's
ntest <- ntot-ntrain # for out-of-sample prediction

# storage space
mspe1 <- rep(NA, nreps) # mspe for M1
mspe2 <- rep(NA, nreps) # mspe for M2

system.time({
  for(ii in 1:nreps) {
    if(ii%%100 == 0) message("ii = ", ii)

    train.ind <- sample(ntot, ntrain) # training observations
    # long-form cross-validation
```

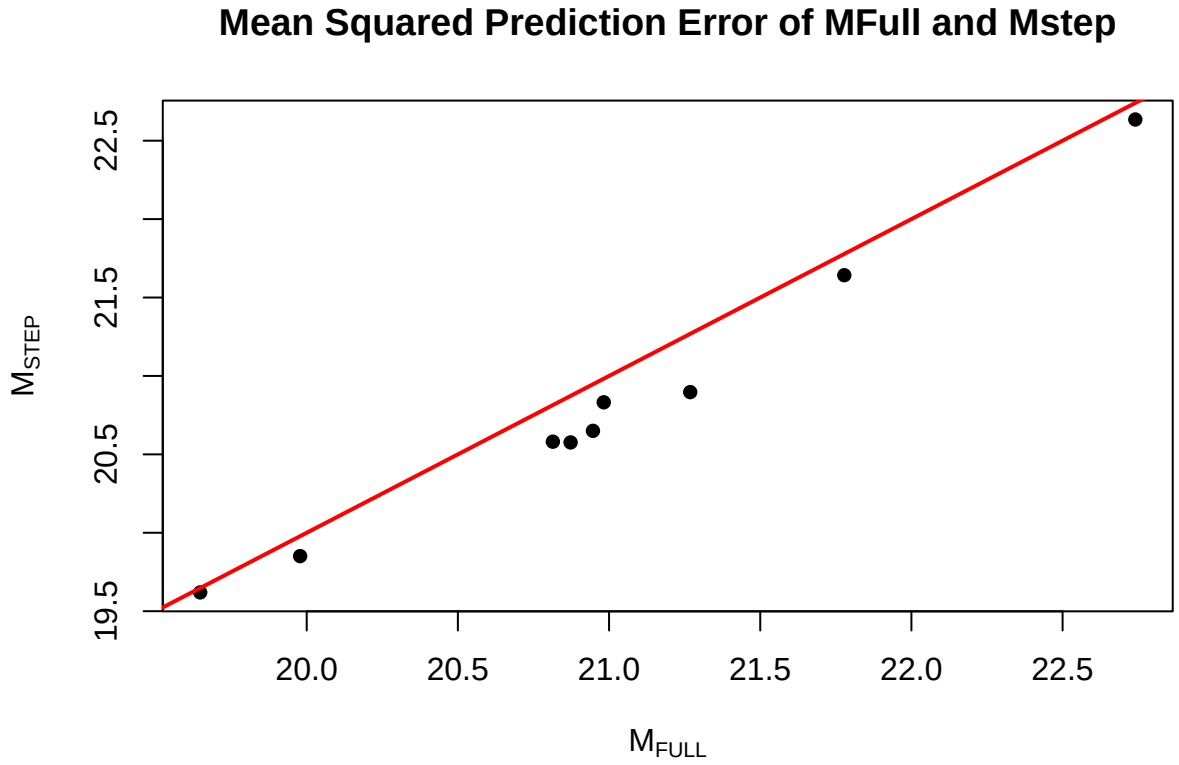
```
    M1.cv <- update(M1, subset = train.ind)
    M2.cv <- update(M2, subset = train.ind)
    # cross-validation residuals
    M1.res <- expm$hs_correct_raven[-train.ind] - # test observations
      predict(M1.cv, newdata = expm[-train.ind,]) # prediction with training data
    M2.res <- expm$hs_correct_rave[-train.ind] -predict(M2.cv, newdata = expm[-train.ind,])
    # mspe for each model
    mspe1[ii] <- mean(M1.res^2)
    mspe2[ii] <- mean(M2.res^2)

  }
})
```

```
## ii = 100
## ii = 200
## ii = 300
## ii = 400
## ii = 500
## ii = 600
## ii = 700
## ii = 800
## ii = 900
## ii = 1000
##    user  system elapsed
##   4.752   0.135   4.998
```

```
# compare
par(mfrow = c(1,2))
cex <- 1
boxplot(x = list(mspe1, mspe2), names = Mnames,
        main = "MSPE using Random subsets cross-validation",
        #ylab = expression(sqrt(bar(SSE)[CV])),
        ylab = expression(MSPE),
        col = c("yellow", "orange"),
        cex = cex, cex.lab = cex, cex.axis = cex, cex.main = cex)
boxplot(x = list(sqrt(mspe1), sqrt(mspe2)), names = Mnames,
        main = "Root MSPE using Random subsets cross-validation",
        ylab = expression(sqrt(MSPE)),
        ## ylab = expression(SSE[CV]),
        col = c("yellow", "orange"),
        cex = cex, cex.lab = cex, cex.axis = cex, cex.main = cex)
```




```
# compare predictions by training set
par(mfrow=c(1,1))
plot(mspe1, mspe2, pch = 16,
     xlab = Mnames[1], ylab = Mnames[2],
     main = "Mean Squared Prediction Error of MFull and Mstep")
abline(a = 0, b = 1, col= "red", lwd = 2)
```

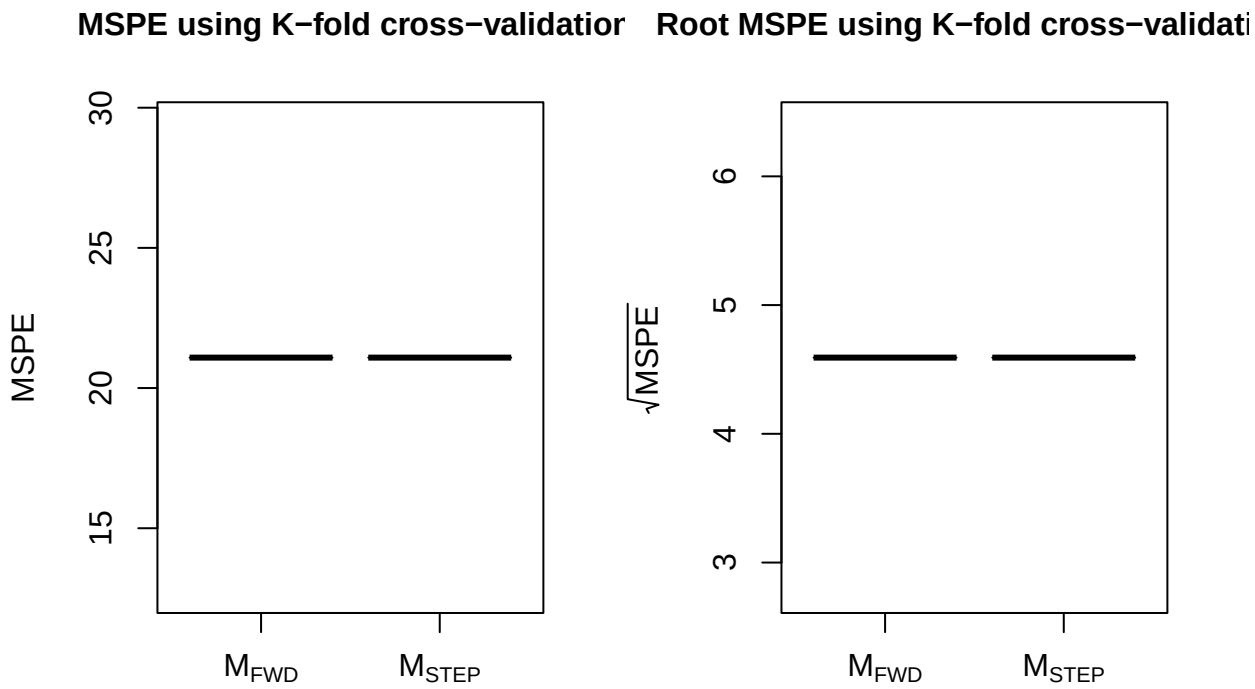


```r
## K-fold cross-validation

# compare Mfwd to Mstep
set.seed(1000)
M1 <- Mfwd
M2 <- Mstep
Mnames <- expression(M[FWD], M[STEP])
expm <- expm[-1300,]
ntot <- nrow(expm) # total number of observations

# number of cross-validation replications
Kfolds <- 5
expm <- expm[sample(ntot),] # permute rows
expm$index <- rep(1:Kfolds,each=ntot/Kfolds)

# storage space
mspe1 <- rep(NA, Kfolds) # mspe for M1
mspe2 <- rep(NA, Kfolds) # mspe for M2

system.time({
  for(ii in 1:Kfolds) {

    train.ind <- which(expm$index!=ii) # training observations


    # using R functions
    M1.cv <- update(M1, subset = train.ind)
    M2.cv <- update(M2, subset = train.ind)
    # cross-validation residuals
    M1.res <- expm$hs_correct_raven[-train.ind] - # test observations
      predict(M1.cv, newdata = expm[-train.ind,]) # prediction with training data
    M2.res <- expm$hs_correct_raven[-train.ind] -predict(M2.cv, newdata = expm[-train.ind,])
    # mspe for each model
    mspe1[ii] <- mean(M1.res^2)
    mspe2[ii] <- mean(M2.res^2)

  }
})
```

```
##    user  system elapsed
##   0.021   0.000   0.021
```

```r
# compare
par(mfrow = c(1,2))
cex <- 1
boxplot(x = list(mspe1, mspe2), names = Mnames,
        main = " MSPE using K-fold cross-validation",
        #ylab = expression(sqrt(bar(SSE)[CV])),
        ylab = expression(MSPE),
        col = c("yellow", "orange"),
        cex = cex, cex.lab = cex, cex.axis = cex, cex.main = cex)
boxplot(x = list(sqrt(mspe1), sqrt(mspe2)), names = Mnames,
        main = "Root MSPE using K-fold cross-validation",
        ylab = expression(sqrt(MSPE)),
```

```
        ## ylab = expression(SSE[CV]),
        col = c("yellow", "orange"),
        cex = cex, cex.lab = cex, cex.axis = cex, cex.main = cex)
```




```
mean(mspe1)
```

```
## [1] NA
```

```
mean(mspe2)
```

```
## [1] NA
```

```
PRESS1 <- resid(M1)/(1-hatvalues(M1))
PRESS2 <- resid(M2)/(1-hatvalues(M2))
# should match if doing LOO-CV
mean(PRESS1^2)
```

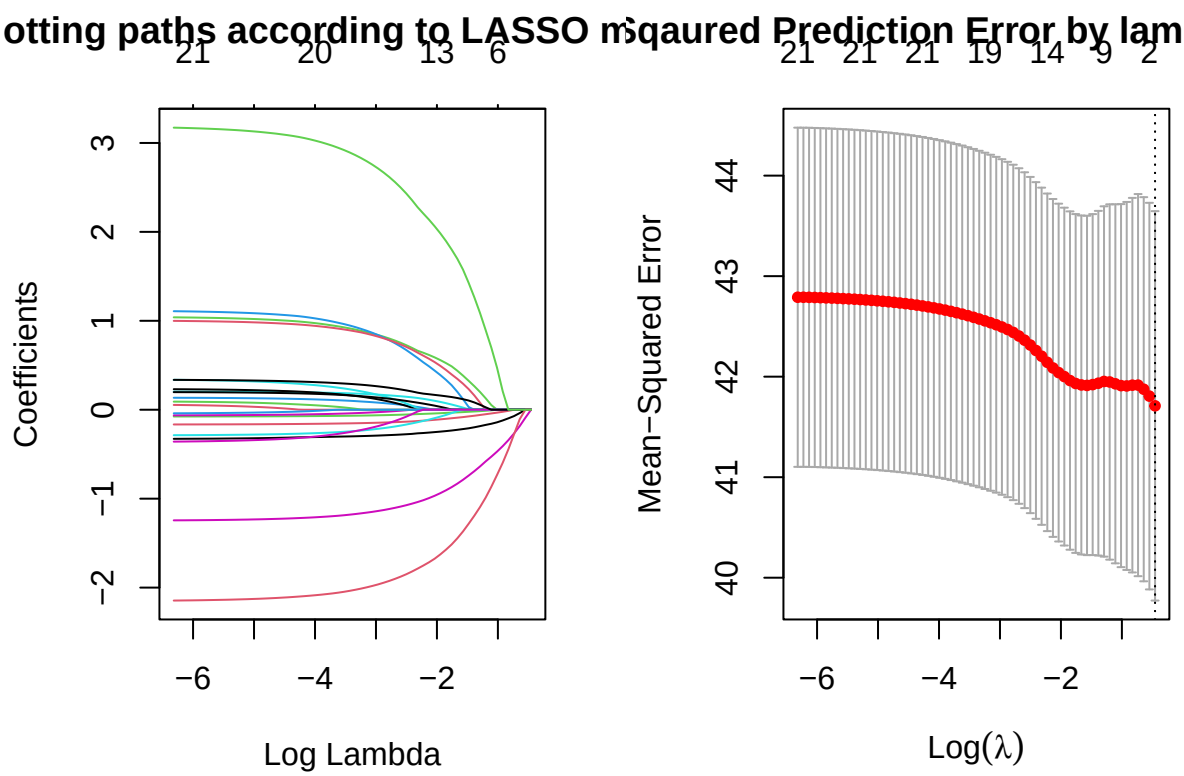
```
## [1] 20.43423
```

```
mean(PRESS2^2)
```

```
## [1] 20.43423
```

```
## LASSO method
set.seed(1000)
## add 20 useless predictors
expm <- data.frame(cbind(expm,
                         matrix(rnorm(20*nrow(expm),0,1),ncol=20)))

X <- model.matrix(Mfull)[,-1] ## get covariates
y <- expm$hs_correct_raven  ## get outcome
Y <- na.omit(y)

## split into test and train
expm <- expm[sample(nrow(expm)),]
```

```r
ntrain <- 500
train_id <- 1:ntrain
X_train <- X[train_id,]
X_test <- X[-train_id,]
Y_train <- Y[train_id]
Y_test <- Y[-train_id]

## fit models
M_lasso <- glmnet(x=X_train,y=Y_train,alpha = 1)
## plot paths
plot(M_lasso,xvar = "lambda",label=TRUE, main= "plotting paths according to LASSO method")
## fit with crossval
cvfit_lasso <-  cv.glmnet(x=X_train,y=Y_train,alpha = 1)
## plot MSPEs by lambda
plot(cvfit_lasso, main="plotting Mean Sqaured Prediction Error by lambda for LASSO method")
```




```r
## estimated betas for minimum lambda
coef(cvfit_lasso, s = "lambda.min")## alternatively could use "lambda.1se"
```

```
## 22 x 1 sparse Matrix of class "dgCMatrix"
##                                 s1
## (Intercept)                 26.036
## hs_child_age_None            .
## hs_as_c_Log2                 .
## hs_as_m_Log2                 .
## hs_cd_c_Log2                 .
## hs_cd_m_Log2                 .
## hs_co_c_Log2                 .
## hs_co_m_Log2                 .
## hs_cs_c_Log2                 .
```

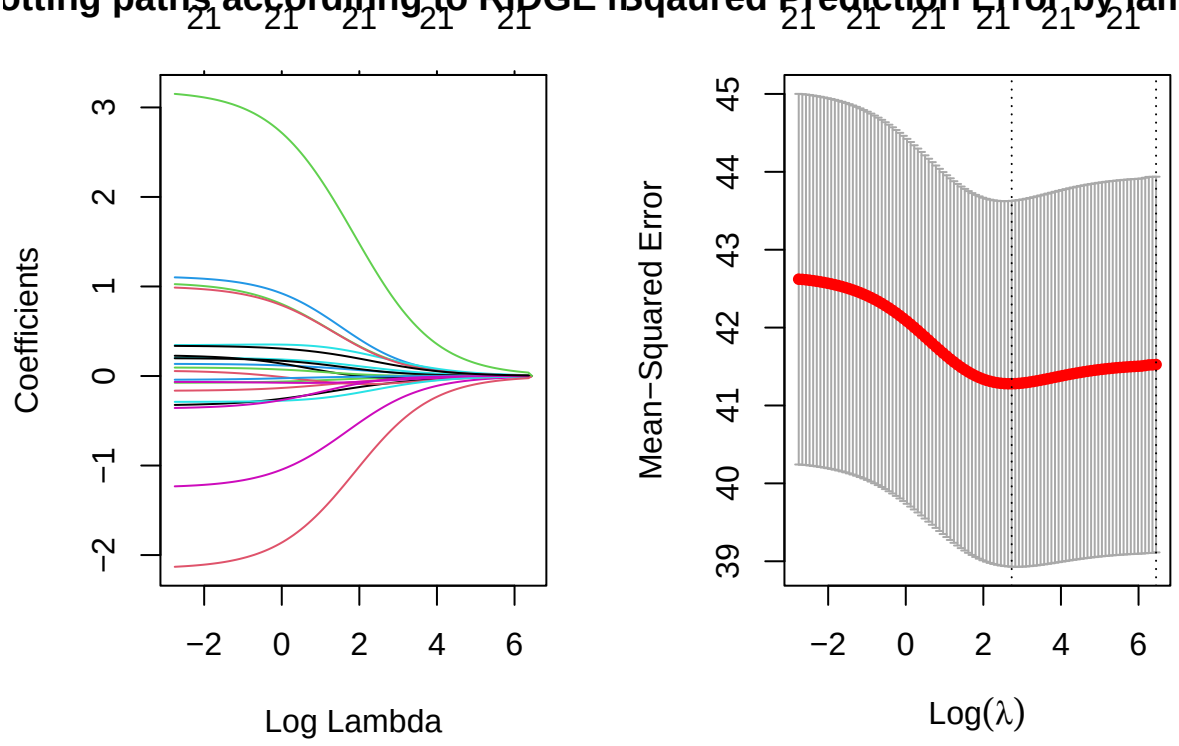
```
## hs_cs_m_Log2                 .
## hs_cu_c_Log2                 .
## hs_cu_m_Log2                 .
## hs_hg_c_Log2                 .
## hs_hg_m_Log2                 .
## hs_mn_c_Log2                 .
## hs_mn_m_Log2                 .
## hs_mo_c_Log2                 .
## hs_mo_m_Log2                 .
## hs_pb_c_Log2                 .
## hs_pb_m_Log2                 .
## hs_tl_cdich_NoneUndetected   .
## hs_tl_mdich_NoneUndetected   .
```

```
## predictions
pred_lasso <- predict(cvfit_lasso,newx=X_test,  s="lambda.min")
## MSPE in test set
MSPE_lasso <- mean((pred_lasso-Y_test)^2)
```

```
## Warning in pred_lasso - Y_test: longer object length is not a multiple of
## shorter object length
```

```
## RIDGE
## fit models
M_ridge <- glmnet(x=X_train,y=Y_train,alpha = 0)
## plot paths
plot(M_ridge,xvar = "lambda",label=TRUE, main= "plotting paths accordinng to RIDGE method")
## fit with crossval
cvfit_ridge <-  cv.glmnet(x=X_train,y=Y_train,alpha = 0)
## plot MSPEs by lambda
plot(cvfit_ridge, main ="plotting Mean Sqaured Prediction Error by lambda for RIDGE method")
```

otting paths accordinng to RIDGE nSqaured Prediction Error by lambda



```r
## estimated betas for minimum lambda
coef(cvfit_ridge, s = "lambda.min")## alternatively could use "lambda.1se"
```

```
## 22 x 1 sparse Matrix of class "dgCMatrix"
##                                       s1
## (Intercept)                 20.402658745
## hs_child_age_None           -0.081897174
## hs_as_c_Log2                -0.039663496
## hs_as_m_Log2                -0.018929286
## hs_cd_c_Log2                 0.042914075
## hs_cd_m_Log2                 0.072899580
## hs_co_c_Log2                -0.320351836
## hs_co_m_Log2                -0.016381096
## hs_cs_c_Log2                -0.083620480
## hs_cs_m_Log2                 0.194306520
## hs_cu_c_Log2                 0.234928523
## hs_cu_m_Log2                 0.188430322
## hs_hg_c_Log2                -0.049447118
## hs_hg_m_Log2                 0.043888199
## hs_mn_c_Log2                 0.186373982
## hs_mn_m_Log2                 0.007028066
## hs_mo_c_Log2                -0.003814262
## hs_mo_m_Log2                -0.126631143
## hs_pb_c_Log2                -0.039207153
## hs_pb_m_Log2                 0.134761994
## hs_tl_cdich_NoneUndetected -0.646247409
## hs_tl_mdich_NoneUndetected  0.961720045
```

```r
## predictions
pred_ridge <- predict(cvfit_ridge,newx=X_test,  s="lambda.min")
## MSPE in test set
MSPE_ridge <- mean((pred_ridge-Y_test)^2)
```

```
## Warning in pred_ridge - Y_test: longer object length is not a multiple of
## shorter object length
```

```r
## compare prediction error for lasso and ridge
MSPE_lasso
```

```
## [1] 41.57002
```

```r
MSPE_ridge
```

```
## [1] 41.7584
```