GPH 2353 Project

Yi Yang

2023-03-14

```
knitr::opts_chunk$set(echo = TRUE)
library(haven)
library(psych)
library(ggplot2)
```

```
##
## Attaching package: 'ggplot2'
```

```
## The following objects are masked from 'package:psych':
##
##     %+%, alpha
```

```
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
library(MASS)
library(psych)
library(leaps)
library(pander)
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:psych':
##
##     logit
```

```
library(faraway)
```

```
##
## Attaching package: 'faraway'

## The following objects are masked from 'package:car':
##
##     logit, vif

## The following object is masked from 'package:psych':
##
##     logit
```

```
library(readr)
library(lmtest)
library(corrplot)
library(r02pro)
```

«««< HEAD Data Preparation

```
my_data <- read.csv("IHDP.csv")
head(my_data)
```

```
##   momage b.marr momed work.dur prenatal cig sex   bw bwg preterm black hispanic
## 1     33      1     4        1        1   0   1 1559   0      10     0        0
## 2     22      0     1        0        1   0   1 2240   1       3     1        0
## 3     13      0     1        0        1   0   1 1900   0       6     1        0
## 4     25      1     4        1        1   0   1 1550   0       8     1        0
## 5     19      0     1        0        1   1   1 2270   1       5     1        0
## 6     19      0     2        1        1   1   0 1550   0       4     1        0
##   white lths hs ltcoll college dayskidh income treat ppvtr.36
## 1     1    0  0      0       1       31  42500     1      111
## 2     0    1  0      0       0        4   5000     1       81
## 3     0    1  0      0       0        9  12500     1       92
## 4     0    0  0      0       1       50  42500     1      103
## 5     0    1  0      0       0        4   5000     1       81
## 6     0    0  1      0       0       13  12500     1       94
```

```
dim(my_data)
```

```
## [1] 4381   21
```

```
for (i in colnames(my_data)) {
  print(sum(is.na(my_data$i)))
}
```

```
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
```

```
View(my_data)
dim(my_data)
```

```
## [1] 4381   21
```

```
### START
summary(my_data)
```

```
##      momage          b.marr           momed          work.dur
##  Min.   :13.0   Min.   :0.0000   Min.   :1.000   Min.   :0.0000
##  1st Qu.:21.0   1st Qu.:0.0000   1st Qu.:1.000   1st Qu.:0.0000
##  Median :24.0   Median :1.0000   Median :2.000   Median :1.0000
##  Mean   :23.8   Mean   :0.6699   Mean   :2.048   Mean   :0.6188
##  3rd Qu.:26.0   3rd Qu.:1.0000   3rd Qu.:3.000   3rd Qu.:1.0000
##  Max.   :41.0   Max.   :1.0000   Max.   :4.000   Max.   :1.0000
##     prenatal           cig              sex               bw
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :1503
##  1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:2892
##  Median :1.0000   Median :0.0000   Median :0.0000   Median :3289
##  Mean   :0.9852   Mean   :0.3314   Mean   :0.4962   Mean   :3247
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:3657
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :7598
##       bwg            preterm           black           hispanic
##  Min.   :0.0000   Min.   :-7.000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:1.0000   1st Qu.: 1.000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.0000   Median : 1.000   Median :0.0000   Median :0.0000
##  Mean   :0.9493   Mean   : 1.503   Mean   :0.2979   Mean   :0.2054
##  3rd Qu.:1.0000   3rd Qu.: 2.000   3rd Qu.:1.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :14.000   Max.   :1.0000   Max.   :1.0000
##      white             lths             hs             ltcoll
##  Min.   :0.0000   Min.   :0.00   Min.   :0.0000   Min.   :0.0000
```

```
##  1st Qu.:0.0000   1st Qu.:0.00   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.00   Median :0.0000   Median :0.0000
##  Mean   :0.4967   Mean   :0.31   Mean   :0.4154   Mean   :0.1911
##  3rd Qu.:1.0000   3rd Qu.:1.00   3rd Qu.:1.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.00   Max.   :1.0000   Max.   :1.0000
##     college           dayskidh          income            treat
##  Min.   :0.00000   Min.   :  0.000   Min.   : -55307   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:  2.000   1st Qu.:   7729   1st Qu.:0.00000
##  Median :0.00000   Median :  3.000   Median :  17025   Median :0.00000
##  Mean   :0.08354   Mean   :  4.864   Mean   :  28085   Mean   :0.06619
##  3rd Qu.:0.00000   3rd Qu.:  5.000   3rd Qu.:  31200   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :100.000   Max.   :1378212   Max.   :1.00000
##     ppvtr.36
##  Min.   : 33.00
##  1st Qu.: 73.00
##  Median : 88.00
##  Mean   : 86.43
##  3rd Qu.:101.00
##  Max.   :129.00
```

```
names(my_data)
```

```
##  [1] "momage"   "b.marr"   "momed"    "work.dur" "prenatal" "cig"
##  [7] "sex"      "bw"       "bwg"      "preterm"  "black"    "hispanic"
## [13] "white"    "lths"     "hs"       "ltcoll"   "college"  "dayskidh"
## [19] "income"   "treat"    "ppvtr.36"
```

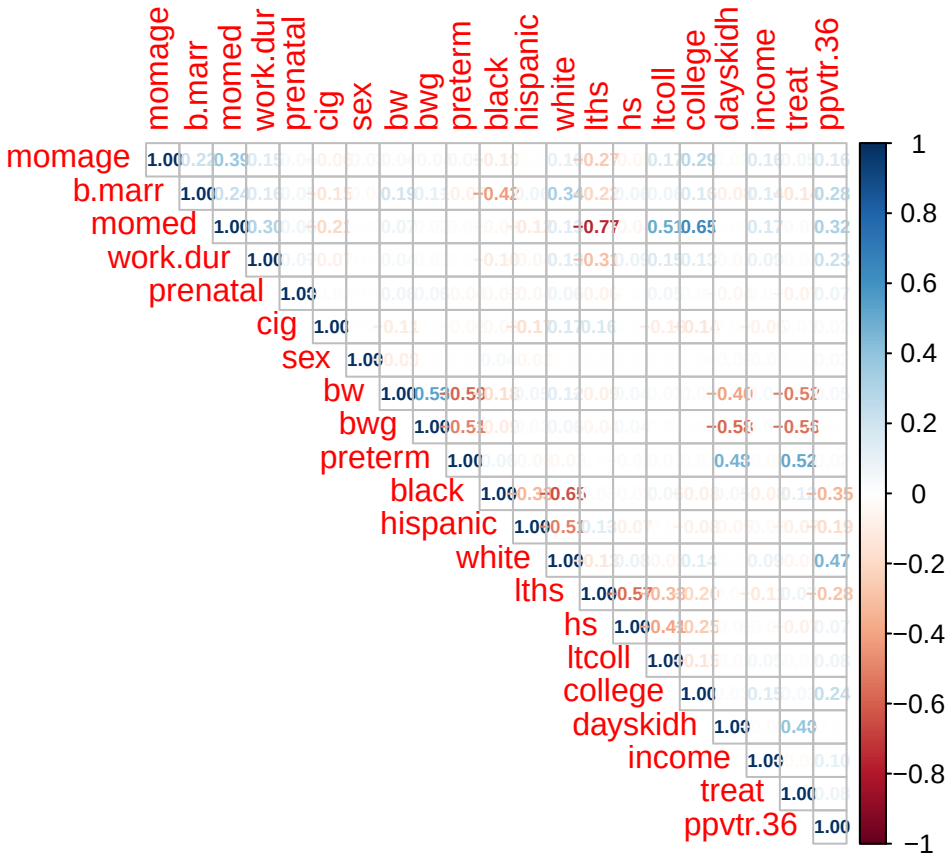
```
head(my_data)
```

```
##   momage b.marr momed work.dur prenatal cig sex   bw bwg preterm black hispanic
## 1     33      1     4        1        1   0   1 1559   0      10     0        0
## 2     22      0     1        0        1   0   1 2240   1       3     1        0
## 3     13      0     1        0        1   0   1 1900   0       6     1        0
## 4     25      1     4        1        1   0   1 1550   0       8     1        0
## 5     19      0     1        0        1   1   1 2270   1       5     1        0
## 6     19      0     2        1        1   1   0 1550   0       4     1        0
##   white lths hs ltcoll college dayskidh income treat ppvtr.36
## 1     1    0  0      0       1       31  42500     1      111
## 2     0    1  0      0       0        4   5000     1       81
## 3     0    1  0      0       0        9  12500     1       92
## 4     0    0  0      0       1       50  42500     1      103
## 5     0    1  0      0       0        4   5000     1       81
## 6     0    0  1      0       0       13  12500     1       94
```

```
my_data = na.omit(my_data)
```

Check Correlation between each variables

```
M=cor(my_data)
corrplot(M,method = "number",type="upper",number.cex = 0.6) # make correlation plot
```



Define Training and test dataset

```
set.seed(0)
tr_size <- nrow(my_data) * 0.7 # training sample size
tr_ind <-sample(nrow(my_data), tr_size)
data_tr <-my_data[tr_ind, ] # training data
data_te <-my_data[-tr_ind, ] # test data
nrow(my_data)
```

```
## [1] 4381
```

```
nrow(data_tr)
```

```
## [1] 3066
```

```
nrow(data_te)
```

```
## [1] 1315
```

Train Model

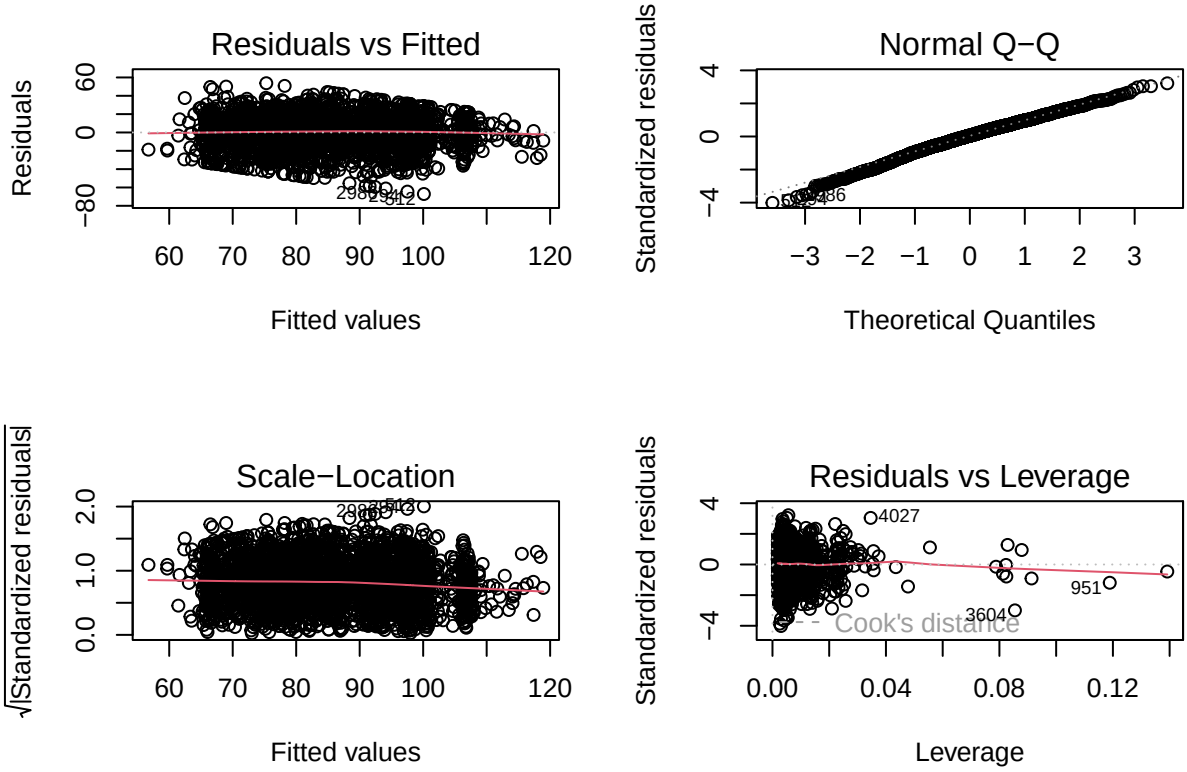
```
model <- lm(ppvtr.36~., data = data_tr)
summary(model)
```

```
##
## Call:
## lm(formula = ppvtr.36 ~ ., data = data_tr)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -67.115  -9.930   0.899  11.322  53.710
##
## Coefficients: (3 not defined because of singularities)
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.718e+01  6.026e+00  11.148  < 2e-16 ***
## momage      -6.517e-02  9.867e-02  -0.660  0.50900
## b.marr       2.225e+00  7.586e-01   2.932  0.00339 **
## momed        6.810e+00  1.288e+00   5.286 1.34e-07 ***
## work.dur     4.344e+00  6.671e-01   6.512 8.65e-11 ***
## prenatal     4.383e+00  2.368e+00   1.851  0.06424 .
## cig          9.519e-01  6.911e-01   1.377  0.16847
## sex          8.090e-02  6.123e-01   0.132  0.89491
## bw           4.431e-04  6.671e-04   0.664  0.50665
## bwg          1.604e+00  1.960e+00   0.818  0.41326
## preterm      2.819e-02  1.761e-01   0.160  0.87283
## black       -1.790e+01  8.059e-01 -22.210  < 2e-16 ***
## hispanic    -1.502e+01  8.403e-01 -17.877  < 2e-16 ***
## white               NA         NA      NA       NA
## lths         4.160e+00  3.028e+00   1.374  0.16961
## hs           4.179e+00  1.818e+00   2.299  0.02156 *
## ltcoll              NA         NA      NA       NA
## college             NA         NA      NA       NA
## dayskidh    -1.039e-01  6.340e-02  -1.638  0.10148
## income       1.121e-06  4.606e-06   0.243  0.80766
## treat        1.240e+01  1.577e+00   7.862 5.19e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.75 on 3048 degrees of freedom
## Multiple R-squared:  0.3255, Adjusted R-squared:  0.3217
## F-statistic: 86.51 on 17 and 3048 DF,  p-value: < 2.2e-16
```

```
alias(model)
```

```
## Model :
## ppvtr.36 ~ momage + b.marr + momed + work.dur + prenatal + cig +
##     sex + bw + bwg + preterm + black + hispanic + white + lths +
##     hs + ltcoll + college + dayskidh + income + treat
##
## Complete :
##         (Intercept) momage b.marr momed work.dur prenatal cig sex bw bwg
## white    1           0      0      0     0        0        0   0   0  0
## ltcoll   4           0      0     -1     0        0        0   0   0  0
## college -3           0      0      1     0        0        0   0   0  0
##         preterm black hispanic lths hs dayskidh income treat
## white    0      -1    -1        0    0  0        0      0
## ltcoll   0       0     0       -3   -2  0        0      0
## college  0       0     0        2    1  0        0      0
```

```
par (mfrow = c(2,2))
plot (model)
```




```
shapiro.test(model$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  model$residuals
## W = 0.99312, p-value = 6.716e-11
```

```
dwtest(model)
```

```
##
##  Durbin-Watson test
##
## data:  model
## DW = 2.0416, p-value = 0.7604
## alternative hypothesis: true autocorrelation is greater than 0
```

```
bptest(model)
```

```
##
```

```
##  studentized Breusch-Pagan test
##
## data:  model
## BP = 65.171, df = 17, p-value = 1.438e-07
```

```
model1<-lm(ppvtr.36~.,data_tr)
step(model1)
```

```
## Start:  AIC=17298.66
## ppvtr.36 ~ momage + b.marr + momed + work.dur + prenatal + cig +
##     sex + bw + bwg + preterm + black + hispanic + white + lths +
##     hs + ltcoll + college + dayskidh + income + treat
##
##
## Step:  AIC=17298.66
## ppvtr.36 ~ momage + b.marr + momed + work.dur + prenatal + cig +
##     sex + bw + bwg + preterm + black + hispanic + white + lths +
##     hs + ltcoll + dayskidh + income + treat
##
##
## Step:  AIC=17298.66
## ppvtr.36 ~ momage + b.marr + momed + work.dur + prenatal + cig +
##     sex + bw + bwg + preterm + black + hispanic + white + lths +
##     hs + dayskidh + income + treat
##
##
## Step:  AIC=17298.66
## ppvtr.36 ~ momage + b.marr + momed + work.dur + prenatal + cig +
##     sex + bw + bwg + preterm + black + hispanic + lths + hs +
##     dayskidh + income + treat
##
##             Df Sum of Sq    RSS   AIC
## - sex        1         5 854685 17297
## - preterm    1         7 854687 17297
## - income     1        17 854697 17297
## - momage     1       122 854803 17297
## - bw         1       124 854804 17297
## - bwg        1       188 854868 17297
## - lths       1       529 855209 17299
## - cig        1       532 855212 17299
## <none>                   854680 17299
## - dayskidh   1       753 855433 17299
## - prenatal   1       961 855641 17300
## - hs         1      1482 856163 17302
## - b.marr     1      2411 857091 17305
## - momed      1      7835 862516 17325
## - work.dur   1     11890 866571 17339
## - treat      1     17334 872014 17358
## - hispanic   1     89618 944298 17602
## - black      1    138319 992999 17757
##
## Step:  AIC=17296.68
## ppvtr.36 ~ momage + b.marr + momed + work.dur + prenatal + cig +
##     bw + bwg + preterm + black + hispanic + lths + hs + dayskidh +
```

```
##     income + treat
##
##            Df Sum of Sq    RSS   AIC
## - preterm   1         7 854692 17295
## - income    1        16 854702 17295
## - bw        1       119 854805 17295
## - momage    1       123 854808 17295
## - bwg       1       188 854873 17295
## - lths      1       529 855214 17297
## - cig       1       529 855215 17297
## <none>                  854685 17297
## - dayskidh  1       761 855446 17297
## - prenatal  1       963 855648 17298
## - hs        1      1482 856167 17300
## - b.marr    1      2411 857096 17303
## - momed     1      7837 862522 17323
## - work.dur  1     11887 866573 17337
## - treat     1     17330 872015 17356
## - hispanic  1     89841 944526 17601
## - black     1    138315 993001 17755
##
## Step:  AIC=17294.71
## ppvtr.36 ~ momage + b.marr + momed + work.dur + prenatal + cig +
##     bw + bwg + black + hispanic + lths + hs + dayskidh + income +
##     treat
##
##            Df Sum of Sq    RSS   AIC
## - income    1        16 854708 17293
## - bw        1       116 854808 17293
## - momage    1       121 854813 17293
## - bwg       1       183 854875 17293
## - cig       1       523 855215 17295
## - lths      1       525 855216 17295
## <none>                  854692 17295
## - dayskidh  1       760 855452 17295
## - prenatal  1       963 855655 17296
## - hs        1      1476 856168 17298
## - b.marr    1      2414 857106 17301
## - momed     1      7831 862523 17321
## - work.dur  1     11888 866579 17335
## - treat     1     18337 873029 17358
## - hispanic  1     89855 944547 17599
## - black     1    138812 993504 17754
##
## Step:  AIC=17292.76
## ppvtr.36 ~ momage + b.marr + momed + work.dur + prenatal + cig +
##     bw + bwg + black + hispanic + lths + hs + dayskidh + treat
##
##            Df Sum of Sq    RSS   AIC
## - bw        1       115 854823 17291
## - momage    1       115 854823 17291
## - bwg       1       181 854889 17291
## - cig       1       519 855227 17293
## - lths      1       538 855246 17293
```

```
## <none>                  854708 17293
## - dayskidh  1       765 855473 17294
## - prenatal  1       963 855671 17294
## - hs        1      1498 856205 17296
## - b.marr    1      2460 857168 17300
## - momed     1      7947 862655 17319
## - work.dur  1     11936 866643 17333
## - treat     1     18321 873029 17356
## - hispanic  1     89940 944647 17598
## - black     1    138817 993525 17752
##
## Step:  AIC=17291.17
## ppvtr.36 ~ momage + b.marr + momed + work.dur + prenatal + cig +
##     bwg + black + hispanic + lths + hs + dayskidh + treat
##
##            Df Sum of Sq    RSS   AIC
## - momage    1       105 854928 17290
## - bwg       1       278 855100 17290
## - cig       1       464 855287 17291
## - lths      1       534 855356 17291
## <none>                  854823 17291
## - dayskidh  1       869 855691 17292
## - prenatal  1       972 855795 17293
## - hs        1      1493 856316 17295
## - b.marr    1      2516 857338 17298
## - momed     1      7943 862765 17318
## - work.dur  1     11892 866715 17332
## - treat     1     19075 873897 17357
## - hispanic  1     90319 945141 17597
## - black     1    142171 996993 17761
##
## Step:  AIC=17289.55
## ppvtr.36 ~ b.marr + momed + work.dur + prenatal + cig + bwg +
##     black + hispanic + lths + hs + dayskidh + treat
##
##            Df Sum of Sq    RSS   AIC
## - bwg       1       296 855224 17289
## - cig       1       450 855378 17289
## - lths      1       519 855447 17289
## <none>                  854928 17290
## - dayskidh  1       850 855778 17291
## - prenatal  1       965 855893 17291
## - hs        1      1474 856402 17293
## - b.marr    1      2435 857363 17296
## - momed     1      7841 862769 17316
## - work.dur  1     11824 866752 17330
## - treat     1     18975 873903 17355
## - hispanic  1     90646 945574 17597
## - black     1    142082 997010 17759
##
## Step:  AIC=17288.61
## ppvtr.36 ~ b.marr + momed + work.dur + prenatal + cig + black +
##     hispanic + lths + hs + dayskidh + treat
##
```

```
##            Df Sum of Sq    RSS   AIC
## - cig       1       444 855668 17288
## - lths      1       540 855764 17289
## <none>                  855224 17289
## - prenatal  1       998 856222 17290
## - hs        1      1515 856739 17292
## - dayskidh  1      1938 857162 17294
## - b.marr    1      2449 857673 17295
## - momed     1      7935 863158 17315
## - work.dur  1     11785 867009 17329
## - treat     1     20453 875677 17359
## - hispanic  1     91222 946446 17597
## - black     1    142643 997867 17760
##
## Step:  AIC=17288.21
## ppvtr.36 ~ b.marr + momed + work.dur + prenatal + black + hispanic +
##     lths + hs + dayskidh + treat
##
##            Df Sum of Sq     RSS   AIC
## - lths      1       500  856168 17288
## <none>                   855668 17288
## - prenatal  1       979  856647 17290
## - hs        1      1433  857101 17291
## - dayskidh  1      1978  857646 17293
## - b.marr    1      2210  857879 17294
## - momed     1      7652  863320 17314
## - work.dur  1     11663  867331 17328
## - treat     1     20400  876068 17358
## - hispanic  1    100492  956160 17627
## - black     1    149904 1005572 17781
##
## Step:  AIC=17288
## ppvtr.36 ~ b.marr + momed + work.dur + prenatal + black + hispanic +
##     hs + dayskidh + treat
##
##            Df Sum of Sq     RSS   AIC
## <none>                   856168 17288
## - prenatal  1       938  857106 17289
## - dayskidh  1      1898  858066 17293
## - b.marr    1      2228  858396 17294
## - hs        1      2374  858542 17295
## - work.dur  1     11421  867589 17327
## - treat     1     20569  876737 17359
## - momed     1     54391  910559 17475
## - hispanic  1    101613  957781 17630
## - black     1    153676 1009844 17792


##
## Call:
## lm(formula = ppvtr.36 ~ b.marr + momed + work.dur + prenatal +
##     black + hispanic + hs + dayskidh + treat, data = data_tr)
##
## Coefficients:
## (Intercept)       b.marr        momed     work.dur     prenatal        black
```

```
##     75.6378       2.0968       4.9870       4.2410       4.3253     -18.2679
##    hispanic           hs     dayskidh        treat
##    -15.4274       1.8241      -0.1378      11.5228
```

```
model2 <- regsubsets(ppvtr.36~., data = data_tr)
```

```
## Warning in leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax, force.in =
## force.in, : 3 linear dependencies found
```

```
## Reordering variables and trying again:
```

```
rs <- summary(model2)
rs
```

```
## Subset selection object
## Call: regsubsets.formula(ppvtr.36 ~ ., data = data_tr)
## 20 Variables  (and intercept)
##          Forced in Forced out
## momage       FALSE      FALSE
## b.marr       FALSE      FALSE
## momed        FALSE      FALSE
## work.dur     FALSE      FALSE
## prenatal     FALSE      FALSE
## cig          FALSE      FALSE
## sex          FALSE      FALSE
## bw           FALSE      FALSE
## bwg          FALSE      FALSE
## preterm      FALSE      FALSE
## black        FALSE      FALSE
## hispanic     FALSE      FALSE
## lths         FALSE      FALSE
## hs           FALSE      FALSE
## dayskidh     FALSE      FALSE
## income       FALSE      FALSE
## treat        FALSE      FALSE
## white        FALSE      FALSE
## ltcoll       FALSE      FALSE
## college      FALSE      FALSE
## 1 subsets of each size up to 9
## Selection Algorithm: exhaustive
##          momage b.marr momed work.dur prenatal cig sex bw  bwg preterm black
## 1  ( 1 ) " "    " "    " "   " "      " "      " " " " " " " " " "     " "
## 2  ( 1 ) " "    " "    "*"   " "      " "      " " " " " " " " " "     " "
## 3  ( 1 ) " "    " "    "*"   "*"      " "      " " " " " " " " " "     " "
## 4  ( 1 ) " "    " "    "*"   "*"      " "      " " " " " " " " " "     " "
## 5  ( 1 ) " "    "*"    "*"   "*"      " "      " " " " " " " " " "     " "
## 6  ( 1 ) " "    "*"    "*"   "*"      " "      " " " " " " " " " "     " "
## 7  ( 1 ) " "    "*"    "*"   "*"      " "      " " " " " " " " " "     " "
## 8  ( 1 ) " "    "*"    "*"   "*"      " "      " " " " " " " " " "     "*"
## 9  ( 1 ) " "    "*"    "*"   "*"      "*"      " " " " " " " " " "     " "
##          hispanic white lths hs  ltcoll college dayskidh income treat
## 1  ( 1 ) " "      "*"   " "  " " " "    " "     " "      " "    " "
```

```
## 2  ( 1 ) " "      "*"   " "  " " " " " "    " "     " "      " "    " "
## 3  ( 1 ) " "      "*"   " "  " " " " " "    " "     " "      " "    " "
## 4  ( 1 ) " "      "*"   " "  " " " " " "    " "     " "      " "    "*"
## 5  ( 1 ) " "      "*"   " "  " " " " " "    " "     " "      " "    "*"
## 6  ( 1 ) "*"      "*"   " "  " " " " " "    " "     " "      " "    "*"
## 7  ( 1 ) "*"      "*"   " "  "*" " " " "    " "     " "      " "    "*"
## 8  ( 1 ) " "      "*"   " "  "*" " " " "    " "     "*"      " "    "*"
## 9  ( 1 ) "*"      "*"   " "  "*" " " " "    " "     "*"      " "    "*"
```

```
rs$which
```

```
##   (Intercept) momage b.marr momed work.dur prenatal   cig   sex    bw   bwg
## 1        TRUE  FALSE  FALSE FALSE    FALSE    FALSE FALSE FALSE FALSE FALSE
## 2        TRUE  FALSE  FALSE  TRUE    FALSE    FALSE FALSE FALSE FALSE FALSE
## 3        TRUE  FALSE  FALSE  TRUE     TRUE    FALSE FALSE FALSE FALSE FALSE
## 4        TRUE  FALSE  FALSE  TRUE     TRUE    FALSE FALSE FALSE FALSE FALSE
## 5        TRUE  FALSE   TRUE  TRUE     TRUE    FALSE FALSE FALSE FALSE FALSE
## 6        TRUE  FALSE   TRUE  TRUE     TRUE    FALSE FALSE FALSE FALSE FALSE
## 7        TRUE  FALSE   TRUE  TRUE     TRUE    FALSE FALSE FALSE FALSE FALSE
## 8        TRUE  FALSE   TRUE  TRUE     TRUE    FALSE FALSE FALSE FALSE FALSE
## 9        TRUE  FALSE   TRUE  TRUE     TRUE     TRUE FALSE FALSE FALSE FALSE
##   preterm black hispanic white  lths    hs ltcoll college dayskidh income treat
## 1   FALSE FALSE    FALSE  TRUE FALSE FALSE  FALSE   FALSE    FALSE  FALSE FALSE
## 2   FALSE FALSE    FALSE  TRUE FALSE FALSE  FALSE   FALSE    FALSE  FALSE FALSE
## 3   FALSE FALSE    FALSE  TRUE FALSE FALSE  FALSE   FALSE    FALSE  FALSE FALSE
## 4   FALSE FALSE    FALSE  TRUE FALSE FALSE  FALSE   FALSE    FALSE  FALSE  TRUE
## 5   FALSE FALSE    FALSE  TRUE FALSE FALSE  FALSE   FALSE    FALSE  FALSE  TRUE
## 6   FALSE FALSE     TRUE  TRUE FALSE FALSE  FALSE   FALSE    FALSE  FALSE  TRUE
## 7   FALSE FALSE     TRUE  TRUE FALSE  TRUE  FALSE   FALSE    FALSE  FALSE  TRUE
## 8   FALSE  TRUE    FALSE  TRUE FALSE  TRUE  FALSE   FALSE     TRUE  FALSE  TRUE
## 9   FALSE FALSE     TRUE  TRUE FALSE  TRUE  FALSE   FALSE     TRUE  FALSE  TRUE
```

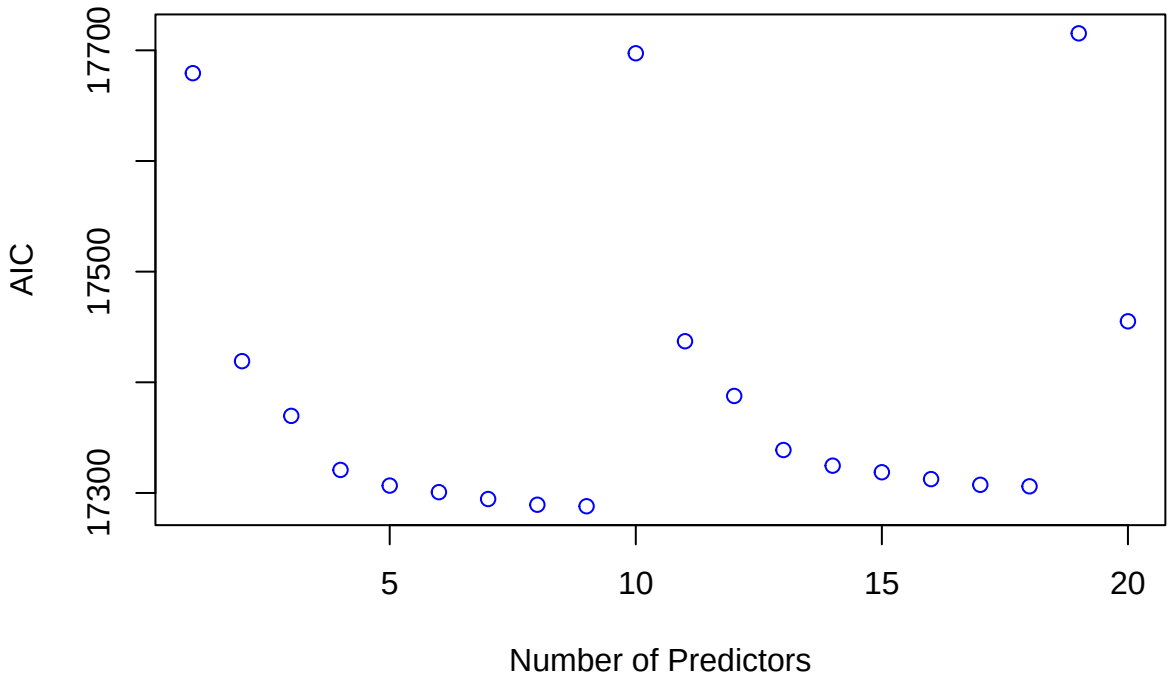
```
n <- nrow(data_tr)
p <- 2:ncol(data_tr)
AIC <- n*log(rs$rss / n) + 2 * p
```

```
## Warning in n * log(rs$rss/n) + 2 * p: longer object length is not a multiple of
## shorter object length
```

```
AIC
```

```
##  [1] 17679.34 17419.10 17369.66 17320.80 17306.66 17300.71 17294.51 17289.35
##  [9] 17288.00 17697.34 17437.10 17387.66 17338.80 17324.66 17318.71 17312.51
## [17] 17307.35 17306.00 17715.34 17455.10
```

```
plot(AIC ~ I(p - 1), ylab = "AIC", xlab = "Number of Predictors", col = "blue")
```




```
which.min(AIC)
```

```
## [1] 9
```

```
model2_AIC <- lm(ppvtr.36 ~ momage + b.marr + momed  + work.dur
                + prenatal + cig + sex + bw + bwg, data = my_data)
summary(model2_AIC)
```

```
##
## Call:
## lm(formula = ppvtr.36 ~ momage + b.marr + momed + work.dur +
##     prenatal + cig + sex + bw + bwg, data = my_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -62.889 -11.369   1.129  12.761  50.884
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 59.4820343  3.2955267  18.049  < 2e-16 ***
## momage      -0.0184390  0.0892195  -0.207  0.83628
## b.marr       9.1587709  0.6261202  14.628  < 2e-16 ***
## momed        5.6748151  0.3493378  16.244  < 2e-16 ***
## work.dur     5.0802517  0.5996652   8.472  < 2e-16 ***
## prenatal     6.5325527  2.2973662   2.843  0.00448 **
```

```
## cig          5.1149020  0.6063371   8.436  < 2e-16 ***
## sex          1.1781469  0.5556539   2.120  0.03404 *
## bw           0.0006011  0.0005223   1.151  0.24983
## bwg         -4.4056528  1.4945746  -2.948  0.00322 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.28 on 4371 degrees of freedom
## Multiple R-squared:  0.1765, Adjusted R-squared:  0.1748
## F-statistic: 104.1 on 9 and 4371 DF,  p-value: < 2.2e-16
```