Econ570-G5-Final

2023-05-06

```
#load the dataset
setwd("/Users/yidanchen/Documents/570 data")
data= read.csv("birpanel.csv")

#see the summary of data
summary(data)
```

```
##      momid3            idx          stateres         dmage            dmeduc
##  Min.   :     1   Min.   :1.0   Min.   : 1.00   Min.   :13.00   Min.   : 0
.00
##  1st Qu.: 35483   1st Qu.:1.0   1st Qu.:14.00   1st Qu.:24.00   1st Qu.:12
.00
##  Median : 70965   Median :1.5   Median :26.00   Median :28.00   Median :14
.00
##  Mean   : 70965   Mean   :1.5   Mean   :26.61   Mean   :28.37   Mean   :13
.88
##  3rd Qu.:106447   3rd Qu.:2.0   3rd Qu.:39.00   3rd Qu.:32.00   3rd Qu.:16
.00
##  Max.   :141929   Max.   :2.0   Max.   :51.00   Max.   :50.00   Max.   :17
.00
##      mplbir          nlbnl            gestat          dbirwt
##  Min.   : 1.00   Min.   : 0.000   Min.   :17.00   Min.   : 227
##  1st Qu.:15.00   1st Qu.: 0.000   1st Qu.:38.00   1st Qu.:3147
##  Median :26.00   Median : 1.000   Median :39.00   Median :3459
##  Mean   :26.14   Mean   : 1.188   Mean   :39.25   Mean   :3454
##  3rd Qu.:38.00   3rd Qu.: 2.000   3rd Qu.:40.00   3rd Qu.:3799
##  Max.   :51.00   Max.   :15.000   Max.   :47.00   Max.   :8020
##      cigar            smoke             male             year
##  Min.   : 0.000   Min.   :0.0000   Min.   :0.0000   Min.   :0.000
##  1st Qu.: 0.000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:2.000
##  Median : 0.000   Median :0.0000   Median :1.0000   Median :4.000
##  Mean   : 2.131   Mean   :0.1301   Mean   :0.5133   Mean   :3.855
##  3rd Qu.: 0.000   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:6.000
##  Max.   :99.000   Max.   :1.0000   Max.   :1.0000   Max.   :8.000
##     married           hsgrad          somecoll         collgrad
##  Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:1.0000   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.0000   Median :0.000   Median :0.0000   Median :0.0000
##  Mean   :0.8689   Mean   :0.295   Mean   :0.2343   Mean   :0.3778
##  3rd Qu.:1.0000   3rd Qu.:1.000   3rd Qu.:0.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.000   Max.   :1.0000   Max.   :1.0000
##      agesq            black           adeqcode2        adeqcode3
##  Min.   : 169.0   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.: 576.0   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000
```

```
##  Median : 784.0   Median :0.00000   Median :0.0000   Median :0.00000
##  Mean   : 834.7   Mean   :0.07426   Mean   :0.1675   Mean   :0.03884
##  3rd Qu.:1024.0   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :2500.0   Max.   :1.00000   Max.   :1.0000   Max.   :1.00000
##     novisit           pretri2          pretri3
##  Min.   :0.000000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.000000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.000000   Median :0.0000   Median :0.00000
##  Mean   :0.007218   Mean   :0.1119   Mean   :0.01957
##  3rd Qu.:0.000000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :1.000000   Max.   :1.0000   Max.   :1.00000
```

```
#clean the data, convert 99(unknown) in cigar into average value.
data$cigar[data$cigar == 99] <- mean(data$cigar[data$cigar != 99 & data$cigar
!= 0], na.rm = TRUE)

# generate and plot correlation matrix
library(ggcorrplot)
```

```
## Loading required package: ggplot2
```

```
cor_matrix <- cor(data)
cor_matrix
```

```
##                  momid3           idx      stateres         dmage       dm
educ
## momid3     1.0000000000  0.0000000000  0.1685379755 -4.143780e-02  0.06219
6329
## idx        0.0000000000  1.0000000000  0.0000000000  2.266642e-01  0.00000
0000
## stateres   0.1685379755  0.0000000000  1.0000000000  7.858944e-03  0.01106
8053
## dmage     -0.0414378026  0.2266642361  0.0078589444  1.000000e+00  0.51003
6103
## dmeduc     0.0621963288  0.0000000000  0.0110680530  5.100361e-01  1.00000
0000
## mplbir    -0.0057765780  0.0000000000 -0.0312839587  2.745268e-02  0.02284
4062
## nlbnl     -0.0771635516  0.4155796786  0.0248736245  3.031937e-01 -0.13908
4778
## gestat    -0.0085947931 -0.0413269171  0.0024638620 -1.643448e-02  0.02070
3013
## dbirwt    -0.0058699817  0.0523189205  0.0116913429  1.198694e-01  0.13217
8618
## cigar     -0.0452688474  0.0145988992 -0.0035374984 -1.249437e-01 -0.29252
2457
## smoke     -0.0404895893  0.0077279008  0.0004397186 -1.630107e-01 -0.32826
4906
## male      -0.0005732736 -0.0058571113  0.0020687410  8.751667e-05  0.00535
6821
## year       0.7232944217  0.5459217813  0.0134302529  1.515114e-01  0.05830
```

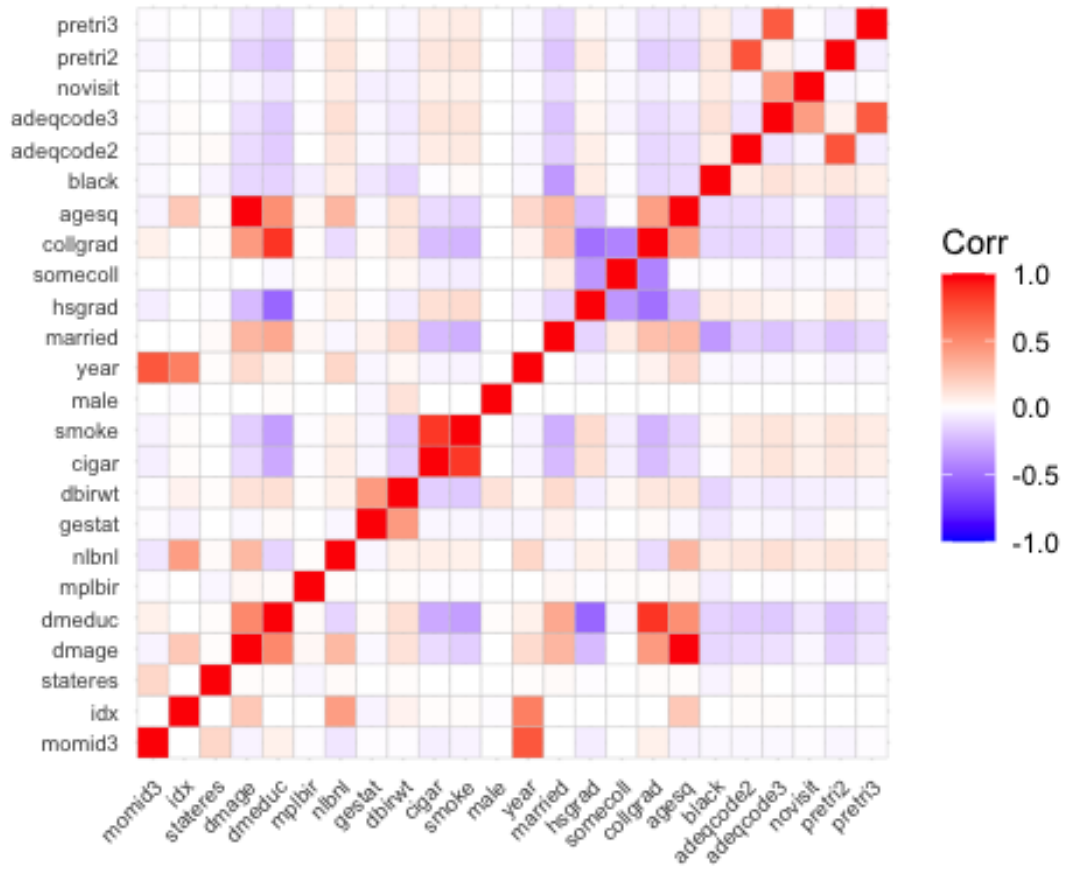
```
0137
## married   -0.0009017620  0.0000000000  0.0170557501  3.141455e-01  0.37391
9298
## hsgrad    -0.0552795729  0.0000000000 -0.0089662932 -2.343755e-01 -0.53619
1478
## somecoll   0.0010193538  0.0000000000  0.0025426828  2.376301e-03 -0.01540
7334
## collgrad   0.0598671413  0.0000000000  0.0087139600  4.262056e-01  0.85259
6130
## agesq     -0.0387767153  0.2274137574  0.0066425065  9.933940e-01  0.48413
2279
## black     -0.0191359018  0.0000000000 -0.0398433731 -1.264517e-01 -0.14604
3646
## adeqcode2 -0.0192848552  0.0096124519  0.0171217998 -1.209812e-01 -0.16830
2338
## adeqcode3 -0.0187623559  0.0063089524 -0.0037078111 -9.835245e-02 -0.17668
7855
## novisit   -0.0127175949  0.0049522043 -0.0060105785 -2.840090e-02 -0.07636
4816
## pretri2   -0.0252105865  0.0011621138 -0.0015105672 -1.491352e-01 -0.20084
7709
## pretri3   -0.0121761082  0.0008901579 -0.0037370358 -8.319470e-02 -0.12804
5545
##                  mplbir        nlbnl       gestat       dbirwt        ciga
r
## momid3    -5.776578e-03 -0.077163552 -0.008594793 -0.005869982 -0.04526884
7
## idx        0.000000e+00  0.415579679 -0.041326917  0.052318921  0.01459889
9
## stateres  -3.128396e-02  0.024873624  0.002463862  0.011691343 -0.00353749
8
## dmage      2.745268e-02  0.303193699 -0.016434483  0.119869413 -0.12494369
3
## dmeduc     2.284406e-02 -0.139084778  0.020703013  0.132178618 -0.29252245
7
## mplbir     1.000000e+00  0.008768059  0.002811841  0.010126925 -0.01128804
3
## nlbnl      8.768059e-03  1.000000000 -0.028942393  0.062091647  0.07240474
3
## gestat     2.811841e-03 -0.028942393  1.000000000  0.427045989 -0.02511205
0
## dbirwt     1.012692e-02  0.062091647  0.427045989  1.000000000 -0.16392917
2
## cigar     -1.128804e-02  0.072404743 -0.025112050 -0.163929172  1.00000000
0
## smoke     -1.257265e-02  0.058589786 -0.027570182 -0.181663032  0.83663139
8
## male       1.983705e-03 -0.004188590 -0.028911895  0.117612637 -0.00278361
8
## year       7.417111e-05  0.170847369 -0.031402706  0.029767506 -0.03548299
```

```
5
## married    3.107602e-02 -0.033284269  0.052970258  0.154827915 -0.22686486
4
## hsgrad    -7.986018e-03  0.056934258 -0.007241539 -0.055973587  0.12714257
1
## somecoll   1.272835e-02  0.025440965  0.002405393  0.028945038 -0.05421336
8
## collgrad   1.227459e-02 -0.123353560  0.017526925  0.098076553 -0.21790507
2
## agesq      2.625091e-02  0.307821029 -0.019754368  0.112874738 -0.11896045
1
## black     -5.742496e-02  0.075843677 -0.081427076 -0.144466859 -0.00710622
5
## adeqcode2 -3.982230e-03  0.099769992 -0.021690596 -0.063519870  0.08314181
2
## adeqcode3 -7.379060e-03  0.126207935 -0.034378453 -0.071392592  0.10671126
5
## novisit   -3.876281e-03  0.083610061 -0.046303712 -0.052495261  0.06056960
2
## pretri2   -7.428303e-03  0.107261453  0.009179432 -0.045883338  0.10082275
8
## pretri3   -3.197212e-03  0.075660891  0.001876181 -0.031202684  0.06917405
8
##                   smoke          male          year      married        hs
grad
## momid3    -0.0404895893 -5.732736e-04  7.232944e-01 -0.000901762 -0.055279
5729
## idx        0.0077279008 -5.857111e-03  5.459218e-01  0.000000000  0.000000
0000
## stateres   0.0004397186  2.068741e-03  1.343025e-02  0.017055750 -0.008966
2932
## dmage     -0.1630107138  8.751667e-05  1.515114e-01  0.314145545 -0.234375
5310
## dmeduc    -0.3282649056  5.356821e-03  5.830014e-02  0.373919298 -0.536191
4783
## mplbir    -0.0125726476  1.983705e-03  7.417111e-05  0.031076021 -0.007986
0176
## nlbnl      0.0585897859 -4.188590e-03  1.708474e-01 -0.033284269  0.056934
2578
## gestat    -0.0275701820 -2.891189e-02 -3.140271e-02  0.052970258 -0.007241
5393
## dbirwt    -0.1816630321  1.176126e-01  2.976751e-02  0.154827915 -0.055973
5867
## cigar      0.8366313980 -2.783618e-03 -3.548300e-02 -0.226864864  0.127142
5712
## smoke      1.0000000000 -3.814022e-03 -3.561600e-02 -0.272291149  0.149274
7294
## male      -0.0038140219  1.000000e+00 -4.009013e-03  0.002271041 -0.000529
7428
## year      -0.0356159959 -4.009013e-03  1.000000e+00  0.002148954 -0.041387
```

```
7561
## married   -0.2722911489  2.271041e-03  2.148954e-03  1.000000000 -0.141704
2276
## hsgrad     0.1492747294 -5.297428e-04 -4.138776e-02 -0.141704228  1.000000
0000
## somecoll  -0.0572526271  8.256246e-04  4.465203e-03  0.075309466 -0.357850
0578
## collgrad  -0.2503581198  3.333226e-03  5.122755e-02  0.274837657 -0.504014
1471
## agesq     -0.1544476278  1.881786e-04  1.553037e-01  0.287337972 -0.230277
1913
## black      0.0175000148 -4.792077e-03 -1.724104e-02 -0.349315107  0.083134
7258
## adeqcode2  0.0942441551  4.439176e-03 -2.546052e-02 -0.155278631  0.072757
7272
## adeqcode3  0.1149344419  7.088857e-04 -2.395042e-02 -0.199887060  0.043105
4933
## novisit    0.0614024794  3.545006e-04 -1.242435e-02 -0.108308131  0.022135
9247
## pretri2    0.1129365826  4.842492e-03 -3.541739e-02 -0.191239043  0.084199
2599
## pretri3    0.0778812853 -2.725272e-04 -1.931181e-02 -0.134569988  0.030358
6333
##                somecoll     collgrad         agesq        black    adeqcod
e2
## momid3     0.0010193538  0.059867141 -0.0387767153 -0.019135902 -0.0192848
55
## idx        0.0000000000  0.000000000  0.2274137574  0.000000000  0.0096124
52
## stateres   0.0025426828  0.008713960  0.0066425065 -0.039843373  0.0171218
00
## dmage      0.0023763009  0.426205582  0.9933940054 -0.126451746 -0.1209812
39
## dmeduc    -0.0154073336  0.852596130  0.4841322787 -0.146043646 -0.1683023
38
## mplbir     0.0127283481  0.012274590  0.0262509144 -0.057424959 -0.0039822
30
## nlbnl      0.0254409646 -0.123353560  0.3078210295  0.075843677  0.0997699
92
## gestat     0.0024053934  0.017526925 -0.0197543684 -0.081427076 -0.0216905
96
## dbirwt     0.0289450383  0.098076553  0.1128747382 -0.144466859 -0.0635198
70
## cigar     -0.0542133683 -0.217905072 -0.1189604510 -0.007106225  0.0831418
12
## smoke     -0.0572526271 -0.250358120 -0.1544476278  0.017500015  0.0942441
55
## male       0.0008256246  0.003333226  0.0001881786 -0.004792077  0.0044391
76
## year       0.0044652027  0.051227546  0.1553037382 -0.017241040 -0.0254605
```

```
19
## married    0.0753094656  0.274837657  0.2873379719 -0.349315107 -0.1552786
31
## hsgrad    -0.3578500578 -0.504014147 -0.2302771913  0.083134726  0.0727577
27
## somecoll   1.0000000000 -0.431063405 -0.0099987828 -0.010524407 -0.0132572
05
## collgrad  -0.4310634050  1.000000000  0.4125010670 -0.131066610 -0.1345583
08
## agesq     -0.0099987828  0.412501067  1.0000000000 -0.117540133 -0.1098522
84
## black     -0.0105244073 -0.131066610 -0.1175401332  1.000000000  0.0826871
21
## adeqcode2 -0.0132572053 -0.134558308 -0.1098522844  0.082687121  1.0000000
00
## adeqcode3 -0.0379379350 -0.115593879 -0.0886338252  0.124782849 -0.0901767
96
## novisit   -0.0198606905 -0.050904779 -0.0245480425  0.078221385 -0.0382532
20
## pretri2   -0.0203348141 -0.157273477 -0.1361269054  0.102907348  0.7308005
81
## pretri3   -0.0241211980 -0.084379810 -0.0753827342  0.068042559 -0.0633807
73
##               adeqcode3       novisit      pretri2       pretri3
## momid3    -0.0187623559 -0.0127175949 -0.025210587 -0.0121761082
## idx        0.0063089524  0.0049522043  0.001162114  0.0008901579
## stateres  -0.0037078111 -0.0060105785 -0.001510567 -0.0037370358
## dmage     -0.0983524504 -0.0284008964 -0.149135203 -0.0831946960
## dmeduc    -0.1766878545 -0.0763648162 -0.200847709 -0.1280455455
## mplbir    -0.0073790601 -0.0038762814 -0.007428303 -0.0031972117
## nlbnl      0.1262079346  0.0836100615  0.107261453  0.0756608906
## gestat    -0.0343784530 -0.0463037118  0.009179432  0.0018761813
## dbirwt    -0.0713925917 -0.0524952612 -0.045883338 -0.0312026845
## cigar      0.1067112653  0.0605696025  0.100822758  0.0691740577
## smoke      0.1149344419  0.0614024794  0.112936583  0.0778812853
## male       0.0007088857  0.0003545006  0.004842492 -0.0002725272
## year      -0.0239504233 -0.0124243461 -0.035417388 -0.0193118141
## married   -0.1998870598 -0.1083081311 -0.191239043 -0.1345699879
## hsgrad     0.0431054933  0.0221359247  0.084199260  0.0303586333
## somecoll  -0.0379379350 -0.0198606905 -0.020334814 -0.0241211980
## collgrad  -0.1155938792 -0.0509047789 -0.157273477 -0.0843798097
## agesq     -0.0886338252 -0.0245480425 -0.136126905 -0.0753827342
## black      0.1247828491  0.0782213848  0.102907348  0.0680425592
## adeqcode2 -0.0901767964 -0.0382532197  0.730800581 -0.0633807732
## adeqcode3  1.0000000000  0.4242024697  0.045642375  0.7028501317
## novisit    0.4242024697  1.0000000000 -0.030270951 -0.0120469370
## pretri2    0.0456423750 -0.0302709514  1.000000000 -0.0501551587
## pretri3    0.7028501317 -0.0120469370 -0.050155159  1.0000000000

ggcorrplot(cor_matrix,type=c("full"), tl.cex=7)
```




```
#choose control variables based on correlation matrix
control_vars <- c("dmage", "nlbnl", "gestat", "male", "married", "hsgrad", "s
omecoll", "collgrad", "black", "adeqcode2", "adeqcode3", "novisit", "pretri2"
, "pretri3")
# dmeduc is not chosen. Cause hsgrad, somecoll, collgrad are generated by dem
educ. agesq is also not used, since it has abnormal distribution and high var
iance, given we already had age as control varible.
```

#for each method, run twice using different treatment(smoke or cigar), except PSM

#Method1:Baseline regression(multiple linear regression)

```
#Baseline model using smoke as treatment
baseline_smoke <- lm(dbirwt ~ smoke + dmage + nlbnl + gestat + male + married
+ hsgrad + somecoll + collgrad + black + adeqcode2 + adeqcode3 + novisit + pr
etri2 + pretri3, data = data)
summary(baseline_smoke)

##
## Call:
## lm(formula = dbirwt ~ smoke + dmage + nlbnl + gestat + male +
##     married + hsgrad + somecoll + collgrad + black + adeqcode2 +
##     adeqcode3 + novisit + pretri2 + pretri3, data = data)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3022.2  -298.5    -6.4   293.3  4832.4
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -879.0976    17.1008 -51.407  < 2e-16 ***
## smoke       -222.2578     2.8178 -78.875  < 2e-16 ***
## dmage          2.5258     0.2110  11.973  < 2e-16 ***
## nlbnl         42.5421     0.8280  51.379  < 2e-16 ***
## gestat       103.9328     0.4059 256.063  < 2e-16 ***
## male         139.0326     1.7466  79.602  < 2e-16 ***
## married       45.7706     3.1045  14.743  < 2e-16 ***
## hsgrad        58.1993     3.5236  16.517  < 2e-16 ***
## somecoll      85.3493     3.8609  22.106  < 2e-16 ***
## collgrad      94.0063     4.0819  23.030  < 2e-16 ***
## black       -181.0568     3.5996 -50.299  < 2e-16 ***
## adeqcode2    -64.4294     3.5632 -18.082  < 2e-16 ***
## adeqcode3   -112.8621     8.4666 -13.330  < 2e-16 ***
## novisit       -7.1802    13.2504  -0.542    0.588
## pretri2       32.9926     4.2887   7.693 1.44e-14 ***
## pretri3       67.8833    10.4283   6.510 7.55e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 464.9 on 283842 degrees of freedom
## Multiple R-squared:  0.2566, Adjusted R-squared:  0.2565
## F-statistic:  6531 on 15 and 283842 DF,  p-value: < 2.2e-16

coef_smoke <- summary(baseline_smoke)$coefficients["smoke", ]
tvalue_smoke <- coef_smoke["t value"]
pvalue_smoke <- coef_smoke["Pr(>|t|)"]

cat("Coefficient for smoke:", coef_smoke[1], "\n")

## Coefficient for smoke: -222.2578

cat("t-value for smoke:", tvalue_smoke, "\n")

## t-value for smoke: -78.87533

cat("p-value for smoke:", pvalue_smoke, "\n")

## p-value for smoke: 0

#Baseline model using cigar as treatment
baseline_cigar <- lm(dbirwt ~ cigar + dmage + nlbnl + gestat + male + married
+ hsgrad + somecoll + collgrad + black + adeqcode2 + adeqcode3 + novisit + pr
etri2 + pretri3, data = data)
summary(baseline_cigar)
```

```
##
## Call:
## lm(formula = dbirwt ~ cigar + dmage + nlbnl + gestat + male +
##     married + hsgrad + somecoll + collgrad + black + adeqcode2 +
##     adeqcode3 + novisit + pretri2 + pretri3, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3022.8  -298.8    -6.0   293.8  4852.6
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -906.8060    17.1042 -53.017  < 2e-16 ***
## cigar        -13.5767     0.1852 -73.299  < 2e-16 ***
## dmage          2.6344     0.2113  12.465  < 2e-16 ***
## nlbnl         43.3540     0.8293  52.276  < 2e-16 ***
## gestat       103.9917     0.4065 255.836  < 2e-16 ***
## male         139.1904     1.7492  79.576  < 2e-16 ***
## married       55.5075     3.0948  17.936  < 2e-16 ***
## hsgrad        61.5105     3.5272  17.439  < 2e-16 ***
## somecoll      92.0869     3.8596  23.859  < 2e-16 ***
## collgrad     103.6931     4.0741  25.452  < 2e-16 ***
## black       -181.3536     3.6080 -50.264  < 2e-16 ***
## adeqcode2    -65.0862     3.5683 -18.240  < 2e-16 ***
## adeqcode3   -112.4470     8.4793 -13.261  < 2e-16 ***
## novisit       -6.3933    13.2701  -0.482     0.63
## pretri2       32.6769     4.2950   7.608 2.79e-14 ***
## pretri3       65.5276    10.4435   6.274 3.51e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 465.6 on 283842 degrees of freedom
## Multiple R-squared:  0.2544, Adjusted R-squared:  0.2544
## F-statistic:  6457 on 15 and 283842 DF,  p-value: < 2.2e-16

coef_cigar <- summary(baseline_cigar)$coefficients["cigar", ]
tvalue_cigar <- coef_cigar["t value"]
pvalue_cigar <- coef_cigar["Pr(>|t|)"]

# Print the results
cat("Coefficient for cigar:", coef_cigar[1], "\n")

## Coefficient for cigar: -13.57675

cat("t-value for cigar:", tvalue_cigar, "\n")

## t-value for cigar: -73.29927

cat("p-value for cigar:", pvalue_cigar, "\n")

## p-value for cigar: 0
```

#Method2:Fixed effects model: regression on difference

```
# generate difference between two pregnancy for variables
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

diff = data %>%
  group_by(momid3) %>%
  mutate(idx = idx - lag(idx)) %>%
  mutate(stateres = stateres - lag(stateres)) %>%
  mutate(dmage = dmage - lag(dmage)) %>%
  mutate(dmeduc = dmeduc - lag(dmeduc)) %>%
  mutate(mplbir = mplbir - lag(mplbir)) %>%
  mutate(nlbnl = nlbnl - lag(nlbnl)) %>%
  mutate(gestat = gestat - lag(gestat)) %>%
  mutate(dbirwt = dbirwt - lag(dbirwt)) %>%
  mutate(cigar = cigar - lag(cigar)) %>%
  mutate(smoke = smoke - lag(smoke)) %>%
  mutate(male = male - lag(male)) %>%
  mutate(year = year - lag(year)) %>%
  mutate(married = married - lag(married)) %>%
  mutate(hsgrad = hsgrad - lag(hsgrad)) %>%
  mutate(somecoll = somecoll - lag(somecoll)) %>%
  mutate(collgrad = collgrad - lag(collgrad)) %>%
  mutate(agesq = agesq - lag(agesq)) %>%
  mutate(black = black - lag(black)) %>%
  mutate(adeqcode2 = adeqcode2 - lag(adeqcode2)) %>%
  mutate(adeqcode3 = adeqcode3 - lag(adeqcode3)) %>%
  mutate(novisit = novisit - lag(novisit)) %>%
  mutate(pretri2 = pretri2 - lag(pretri2)) %>%
  mutate(pretri3 = pretri3 - lag(pretri3)) %>%
  slice_tail(n = 1)

summary(diff)

##      momid3            idx       stateres     dmage           dmeduc
##  Min.   :     1   Min.   :1   Min.   :0   Min.   :0.000   Min.   :0
##  1st Qu.: 35483   1st Qu.:1   1st Qu.:0   1st Qu.:2.000   1st Qu.:0
##  Median : 70965   Median :1   Median :0   Median :2.000   Median :0
##  Mean   : 70965   Mean   :1   Mean   :0   Mean   :2.464   Mean   :0
##  3rd Qu.:106447   3rd Qu.:1   3rd Qu.:0   3rd Qu.:3.000   3rd Qu.:0
```

```
##  Max.   :141929   Max.   :1   Max.   :0   Max.   :9.000   Max.   :0
##      mplbir      nlbnl       gestat            dbirwt
##  Min.   :0   Min.   :1   Min.   :-24.000   Min.   :-4628.00
##  1st Qu.:0   1st Qu.:1   1st Qu.: -2.000   1st Qu.: -284.00
##  Median :0   Median :1   Median :  0.000   Median :   57.00
##  Mean   :0   Mean   :1   Mean   : -0.179   Mean   :   56.42
##  3rd Qu.:0   3rd Qu.:1   3rd Qu.:  1.000   3rd Qu.:  397.00
##  Max.   :0   Max.   :1   Max.   : 22.000   Max.   : 5327.00
##      cigar              smoke              male                year
##  Min.   :-98.0000   Min.   :-1.0000   Min.   :-1.000000   Min.   :0.000
##  1st Qu.:  0.0000   1st Qu.: 0.0000   1st Qu.:-1.000000   1st Qu.:2.000
##  Median :  0.0000   Median : 0.0000   Median : 0.000000   Median :2.000
##  Mean   :  0.1471   Mean   : 0.0052   Mean   :-0.005855   Mean   :2.467
##  3rd Qu.:  0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.000000   3rd Qu.:3.000
##  Max.   : 93.0000   Max.   : 1.0000   Max.   : 1.000000   Max.   :8.000
##     married      hsgrad     somecoll    collgrad     agesq           black
##  Min.   :0   Min.   :0   Min.   :0   Min.   :0   Min.   :  0.0   Min.   :0
##  1st Qu.:0   1st Qu.:0   1st Qu.:0   1st Qu.:0   1st Qu.: 84.0   1st Qu.:0
##  Median :0   Median :0   Median :0   Median :0   Median :124.0   Median :0
##  Mean   :0   Mean   :0   Mean   :0   Mean   :0   Mean   :141.5   Mean   :0
##  3rd Qu.:0   3rd Qu.:0   3rd Qu.:0   3rd Qu.:0   3rd Qu.:183.0   3rd Qu.:0
##  Max.   :0   Max.   :0   Max.   :0   Max.   :0   Max.   :672.0   Max.   :0
##    adeqcode2          adeqcode3            novisit
##  Min.   :-1.00000   Min.   :-1.000000   Min.   :-1.0000000
##  1st Qu.: 0.00000   1st Qu.: 0.000000   1st Qu.: 0.0000000
##  Median : 0.00000   Median : 0.000000   Median : 0.0000000
##  Mean   : 0.00718   Mean   : 0.002438   Mean   : 0.0008384
##  3rd Qu.: 0.00000   3rd Qu.: 0.000000   3rd Qu.: 0.0000000
##  Max.   : 1.00000   Max.   : 1.000000   Max.   : 1.0000000
##     pretri2              pretri3
##  Min.   :-1.0000000   Min.   :-1.0000000
##  1st Qu.: 0.0000000   1st Qu.: 0.0000000
##  Median : 0.0000000   Median : 0.0000000
##  Mean   : 0.0007328   Mean   : 0.0002466
##  3rd Qu.: 0.0000000   3rd Qu.: 0.0000000
##  Max.   : 1.0000000   Max.   : 1.0000000
```

```
# regression on difference. Treatment=smoke
diff_smoke = lm(dbirwt ~ smoke + dmage + gestat + male  + adeqcode2 + adeqcod
e3 + novisit + pretri2 + pretri3, data = diff)
summary(diff_smoke)
```

```
##
## Call:
## lm(formula = dbirwt ~ smoke + dmage + gestat + male + adeqcode2 +
##     adeqcode3 + novisit + pretri2 + pretri3, data = diff)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4580.4  -321.2     0.7   321.5  5288.1
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   67.214      3.027  22.204  < 2e-16 ***
## smoke       -134.293      4.465 -30.077  < 2e-16 ***
## dmage          2.938      1.090   2.695  0.00704 **
## gestat        88.777      0.504 176.158  < 2e-16 ***
## male         142.668      1.974  72.282  < 2e-16 ***
## adeqcode2    -54.771      4.249 -12.892  < 2e-16 ***
## adeqcode3   -102.541      9.916 -10.340  < 2e-16 ***
## novisit       -8.272     16.041  -0.516  0.60608
## pretri2       31.874      5.029   6.339 2.32e-10 ***
## pretri3       66.452     12.020   5.528 3.24e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 525 on 141919 degrees of freedom
## Multiple R-squared:  0.2082, Adjusted R-squared:  0.2081
## F-statistic:  4146 on 9 and 141919 DF,  p-value: < 2.2e-16

coef_smoke <- summary(diff_smoke)$coefficients["smoke", ]
tvalue_smoke <- coef_smoke["t value"]
pvalue_smoke <- coef_smoke["Pr(>|t|)"]

cat("Coefficient for smoke:", coef_smoke[1], "\n")

## Coefficient for smoke: -134.2926

cat("t-value for smoke:", tvalue_smoke, "\n")

## t-value for smoke: -30.07747

cat("p-value for smoke:", pvalue_smoke, "\n")

## p-value for smoke: 4.026753e-198

# regression on difference. Treatment=smoke
diff_cigar = lm(dbirwt ~ cigar + dmage + gestat + male  + adeqcode2 + adeqcod
e3 + novisit + pretri2 + pretri3, data = diff)
summary(diff_cigar)

##
## Call:
## lm(formula = dbirwt ~ cigar + dmage + gestat + male + adeqcode2 +
##     adeqcode3 + novisit + pretri2 + pretri3, data = diff)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4580.9  -321.0     1.3   321.3  5342.0
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)   67.9342     3.0289  22.429  < 2e-16 ***
## cigar         -7.9843     0.2881 -27.714  < 2e-16 ***
## dmage          2.8358     1.0905   2.600  0.00931 **
## gestat        88.7508     0.5042 176.018  < 2e-16 ***
## male         142.6973     1.9747  72.262  < 2e-16 ***
## adeqcode2    -54.5046     4.2507 -12.822  < 2e-16 ***
## adeqcode3   -101.9691     9.9215 -10.278  < 2e-16 ***
## novisit       -7.8548    16.0493  -0.489  0.62455
## pretri2       31.5509     5.0309   6.271 3.59e-10 ***
## pretri3       64.7515    12.0258   5.384 7.28e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 525.3 on 141919 degrees of freedom
## Multiple R-squared:  0.2074, Adjusted R-squared:  0.2074
## F-statistic:  4127 on 9 and 141919 DF,  p-value: < 2.2e-16

coef_cigar <- summary(diff_cigar)$coefficients["cigar", ]
tvalue_cigar <- coef_smoke["t value"]
pvalue_cigar <- coef_smoke["Pr(>|t|)"]

cat("Coefficient for cigar:", coef_cigar[1], "\n")

## Coefficient for cigar: -7.984308

cat("t-value for cigar:", tvalue_cigar, "\n")

## t-value for cigar: -30.07747

cat("p-value for cigar:", pvalue_cigar, "\n")

## p-value for cigar: 4.026753e-198
```

#Method3:PSM Propensity Score Matching

```
library(dplyr)
library(MatchIt)
vars_to_match <- c("dmage", "nlbnl", "gestat", "male", "married", "hsgrad", "
somecoll", "collgrad", "black", "adeqcode2", "adeqcode3", "novisit", "pretri2
", "pretri3")
data_match <- data %>% dplyr::select(dbirwt, smoke, one_of(vars_to_match))
m.out <- matchit(smoke ~ dmage + nlbnl + gestat + male + married + hsgrad + s
omecoll + collgrad + black + adeqcode2 + adeqcode3 + novisit + pretri2 + pret
ri3, data = data_match, method = "nearest")
matched_data <- match.data(m.out)
PSM_smoke<- lm(dbirwt ~ smoke, data = matched_data)
summary(PSM_smoke)

##
## Call:
## lm(formula = dbirwt ~ smoke, data = matched_data)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3197.9  -308.9    20.1   346.1  4595.1
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3424.915      2.893 1184.05   <2e-16 ***
## smoke       -223.996      4.091  -54.76   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 555.9 on 73866 degrees of freedom
## Multiple R-squared:  0.03901,    Adjusted R-squared:  0.039
## F-statistic:  2998 on 1 and 73866 DF,  p-value: < 2.2e-16
```

#Method4:Instrument variable: dmeduc

```
# Treatment=smoke
library(AER)

## Loading required package: car

## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode

## Loading required package: lmtest

## Loading required package: zoo

##
## Attaching package: 'zoo'

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

## Loading required package: sandwich

## Loading required package: survival

iv_smoke <- ivreg(dbirwt ~ smoke + dmage + nlbnl + gestat + male + married +
black + adeqcode2 + adeqcode3 + novisit + pretri2 + pretri3 | dmeduc+ dmage +
nlbnl + gestat + male + married + black + adeqcode2 + adeqcode3 + novisit + p
retri2 + pretri3, data = data)

summary(iv_smoke, vcov = sandwich, diagnostics = TRUE)
```

```
##
## Call:
## ivreg(formula = dbirwt ~ smoke + dmage + nlbnl + gestat + male +
##     married + black + adeqcode2 + adeqcode3 + novisit + pretri2 +
##     pretri3 | dmeduc + dmage + nlbnl + gestat + male + married +
##     black + adeqcode2 + adeqcode3 + novisit + pretri2 + pretri3,
##     data = data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -3047.816  -303.398    -7.266   297.779  4743.916
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -729.4895    25.7269 -28.355  < 2e-16 ***
## smoke       -475.1948    12.5941 -37.732  < 2e-16 ***
## dmage          3.1736     0.2072  15.318  < 2e-16 ***
## nlbnl         42.6659     0.8780  48.596  < 2e-16 ***
## gestat       103.3081     0.6012 171.844  < 2e-16 ***
## male         138.6004     1.7696  78.323  < 2e-16 ***
## married        2.9649     4.4804   0.662    0.508
## black       -211.5949     4.1842 -50.570  < 2e-16 ***
## adeqcode2    -61.1848     3.7174 -16.459  < 2e-16 ***
## adeqcode3   -102.7594     9.3461 -10.995  < 2e-16 ***
## novisit        6.2013    15.1354   0.410    0.682
## pretri2       38.0846     4.5488   8.372  < 2e-16 ***
## pretri3       70.7650    11.3327   6.244 4.26e-10 ***
##
## Diagnostic tests:
##                      df1    df2 statistic p-value
## Weak instruments       1 283845   13070.4  <2e-16 ***
## Wu-Hausman             1 283844     381.6  <2e-16 ***
## Sargan                 0     NA        NA      NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 471.5 on 283845 degrees of freedom
## Multiple R-Squared: 0.2354,  Adjusted R-squared: 0.2353
## Wald test:  4292 on 12 and 283845 DF,  p-value: < 2.2e-16
```

```
# Treatment=cigar
iv_cigar <- ivreg(dbirwt ~ cigar + dmage + nlbnl + gestat + male + married +
black + adeqcode2 + adeqcode3 + novisit + pretri2 + pretri3 | dmeduc+ dmage +
nlbnl + gestat + male + married + black + adeqcode2 + adeqcode3 + novisit + p
retri2 + pretri3, data = data)

summary(iv_cigar, vcov = sandwich, diagnostics = TRUE)
```

```
##
## Call:
```

```
## ivreg(formula = dbirwt ~ cigar + dmage + nlbnl + gestat + male +
##     married + black + adeqcode2 + adeqcode3 + novisit + pretri2 +
##     pretri3 | dmeduc + dmage + nlbnl + gestat + male + married +
##     black + adeqcode2 + adeqcode3 + novisit + pretri2 + pretri3,
##     data = data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -3056.432  -305.725    -9.338   297.523  4762.710
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -758.1379    25.6859 -29.516  < 2e-16 ***
## cigar        -33.9131     0.9116 -37.201  < 2e-16 ***
## dmage          3.6717     0.2043  17.969  < 2e-16 ***
## nlbnl         44.6336     0.9070  49.210  < 2e-16 ***
## gestat       103.2548     0.6060 170.401  < 2e-16 ***
## male         138.8358     1.7843  77.811  < 2e-16 ***
## married       12.4487     4.3827   2.840  0.00451 **
## black       -222.5816     4.3290 -51.417  < 2e-16 ***
## adeqcode2    -61.6802     3.7558 -16.423  < 2e-16 ***
## adeqcode3    -97.3247     9.6613 -10.074  < 2e-16 ***
## novisit       12.4154    15.7103   0.790  0.42937
## pretri2       39.0313     4.6161   8.456  < 2e-16 ***
## pretri3       65.8020    11.6765   5.635 1.75e-08 ***
##
## Diagnostic tests:
##                     df1    df2 statistic p-value
## Weak instruments      1 283845    9775.6  <2e-16 ***
## Wu-Hausman            1 283844     487.1  <2e-16 ***
## Sargan                0     NA        NA      NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 475.4 on 283845 degrees of freedom
## Multiple R-Squared: 0.2227,  Adjusted R-squared: 0.2226
## Wald test:  4220 on 12 and 283845 DF,  p-value: < 2.2e-16
```