Assignment 2

```
### 1) Randomization
>
> ### a) t.test for significance in difference in outcome between
treated and control pupils
>
> t.test(shr_Xfood ~ eligilitystatus, data=EL)

      Welch Two Sample t-test

data:  shr_Xfood by eligilitystatus
t = -2.2908, df = 3822.5, p-value = 0.02203
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.0778957 -0.1613853
sample estimates:
mean in group 0 mean in group 1
       58.47350        59.59314

> t.test(cpexp30_pae ~ eligilitystatus, data=EL)

      Welch Two Sample t-test

data:  cpexp30_pae by eligilitystatus
t = -6.7876, df = 3696.5, p-value = 1.323e-11
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -21733.84 -11992.05
sample estimates:
mean in group 0 mean in group 1
       81827.77        98690.72


> t.test(Xeduc_pch ~ eligilitystatus, data=EL)

      Welch Two Sample t-test

data:  Xeduc_pch by eligilitystatus
t = 2.3096, df = 2956, p-value = 0.02098
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  572.8886 7012.7013
sample estimates:
mean in group 0 mean in group 1
       15192.75        11399.96

> RegRandom <- lm(cpexp30_pae ~aeligibility,data=test_data)
> summary(RegRandom)

Call:
lm(formula = cpexp30_pae ~ aeligibility, data = test_data)

Residuals:
    Min      1Q  Median      3Q     Max
 -64721  -35440  -15527   15742 1085062
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)     56367       1382  40.797  < 2e-16 ***
aeligibility    12549       1947   6.444 1.31e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 60350 on 3840 degrees of freedom
  (70 observations deleted due to missingness)
Multiple R-squared:  0.0107,     Adjusted R-squared:  0.01044
F-statistic: 41.53 on 1 and 3840 DF,  p-value: 1.307e-10
```

b) Regression estimation of variable influence

```
> t.test(Xeduc_pch ~ eligilitystatus, data=EL)

      Welch Two Sample t-test

data:  Xeduc_pch by eligilitystatus
t = 2.3096, df = 2956, p-value = 0.02098
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  572.8886 7012.7013
sample estimates:
mean in group 0 mean in group 1
       15192.75        11399.96
```

```
summary(RegRandom)

Call:
lm(formula = Xeduc_pch ~ aeligibility, data = test_data)

Residuals:
    Min      1Q  Median      3Q     Max
 -12181  -11112   -9578   -3999 1056696

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   12181.0     1071.8  11.365   <2e-16 ***
aeligibility   -383.7     1560.8  -0.246    0.806
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 42320 on 2948 degrees of freedom
  (962 observations deleted due to missingness)
Multiple R-squared:  2.05e-05,   Adjusted R-squared:  -0.0003187
F-statistic: 0.06044 on 1 and 2948 DF,  p-value: 0.8058
```

From the regression, we have found that if a person is being treated, the impact is negative (-383.7) on the monthly education expenditure per child aged 6 to 17 (Xeduc_pch).It is not significant as the p value is very high(0.806).

c) t-test for difference between treatment and control group

```
t.test(Xeduc_pch ~ eligilitystatus, data=EL)

     Welch Two Sample t-test

data:  Xeduc_pch by eligilitystatus
t = 2.3096, df = 2956, p-value = 0.02098
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  572.8886 7012.7013
sample estimates:
mean in group 0 mean in group 1
       15192.75        11399.96

Welch Two Sample t-test

data:  rXheal by eligilitystatus
t = -1.9067, df = 2471, p-value = 0.05667
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -12566.0116    176.0879
sample estimates:
mean in group 0 mean in group 1
       19155.45        25350.42
> t.test(Xeduc ~ eligilitystatus, data=EL)

     Welch Two Sample t-test

data:  Xeduc by eligilitystatus
t = 4.2896, df = 3609, p-value = 1.837e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  6067.667 16283.503
sample estimates:
mean in group 0 mean in group 1
       35048.26        23872.67

> t.test(Xfood ~ eligilitystatus, data=EL)

     Welch Two Sample t-test

data:  Xfood by eligilitystatus
t = 4.0031, df = 3728, p-value = 6.373e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 11326.40 33070.63
sample estimates:
mean in group 0 mean in group 1
       212519.3        190320.8

t.test(shr_Xfood ~ eligilitystatus, data=EL)

     Welch Two Sample t-test

data:  shr_Xfood by eligilitystatus
t = -2.2908, df = 3822.5, p-value = 0.02203
```

```
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.0778957 -0.1613853
sample estimates:
mean in group 0 mean in group 1
       58.47350        59.59314


> RegRandom1 <-
lm(cpexp30_pae~aeligibility+rXheal+Xeduc+Xfood+shr_Xfood,data=test_d
ata)
> summary(RegRandom1)

Call:
lm(formula = cpexp30_pae ~ aeligibility + rXheal + Xeduc + Xfood +
    shr_Xfood, data = test_data)

Residuals:
    Min      1Q  Median      3Q     Max
-151093  -28206  -12179   12523 1009690

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.504e+04  4.132e+03  15.741   <2e-16 ***
aeligibility  1.480e+04  1.730e+03   8.552   <2e-16 ***
rXheal        2.221e-01  1.369e-02  16.220   <2e-16 ***
Xeduc         2.707e-02  1.129e-02   2.397   0.0166 *
Xfood         1.216e-01  5.526e-03  22.005   <2e-16 ***
shr_Xfood    -5.686e+02  5.990e+01  -9.492   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 53460 on 3834 degrees of freedom
  (72 observations deleted due to missingness)
Multiple R-squared:  0.2249,     Adjusted R-squared:  0.2239
F-statistic: 222.5 on 5 and 3834 DF,  p-value: < 2.2e-16


### 2.) Instrumental Variable Regression

### a) Use place of residence (variable: urban) as an
instrument> summary(ivreg)

Call:
ivreg(formula = cpexp30_pae ~ aeligibility + shr_Xfood +
rXdrinks +
    rXheal + Xeduc | urban, data = test_data)

Residuals:
      Min        1Q    Median        3Q       Max
-12471089 -12418365  12074445  12137051  12887693

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    12482776  751400955   0.017    0.987
```

```
aeligibility  -24535423 1487935883  -0.016    0.987

Residual standard error: 12290000 on 1099 degrees of freedom
Multiple R-Squared: -2.28e+04,     Adjusted R-squared: -
2.282e+04
Wald test: 0.0002719 on 1 and 1099 DF,  p-value: 0.9868

#Adding Additional variables
> ivreg4<-ivreg(cpexp30_pae ~
aeligibility+shr_Xfood+rXdrinks+rXheal+Xeduc|urban+rXheal+Xedu
c, data = test_data)
> Mahmud<-na.omit(test_data)
> summary(ivreg4)

Call:
ivreg(formula = cpexp30_pae ~ aeligibility + shr_Xfood +
rXdrinks +
    rXheal + Xeduc | urban + rXheal + Xeduc, data = test_data)

Residuals:
    Min      1Q  Median      3Q     Max
-548105 -375642   24664  355649 1266895

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  -9.940e+04  3.227e+05  -0.308    0.758
aeligibility  7.525e+05  5.183e+05   1.452    0.147
shr_Xfood    -3.142e+03  3.120e+03  -1.007    0.314
rXdrinks     -4.312e-01  2.771e+01  -0.016    0.988

Residual standard error: 371100 on 1097 degrees of freedom
Multiple R-Squared: -19.75,   Adjusted R-squared: -19.81
Wald test: 2.991 on 3 and 1097 DF,  p-value: 0.0301

#b) Run the same analysis with Xeduc_pch

> ivreg5<- ivreg(Xeduc_pch ~
aeligibility+shr_Xfood+rXdrinks+rXheal+Xeduc|urban, data =
test_data)
> summary(ivreg5)

Call:
ivreg(formula = Xeduc_pch ~ aeligibility + shr_Xfood +
rXdrinks +
    rXheal + Xeduc | urban, data = test_data)

Residuals:
     Min       1Q   Median       3Q      Max
-2549129 -2546844 -2513682  3059509  4141429

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)    2549129   76581073   0.033    0.973
aeligibility  -5605380  169358011  -0.033    0.974

Residual standard error: 2795000 on 845 degrees of freedom
Multiple R-Squared: -3298,    Adjusted R-squared: -3302
Wald test: 0.001095 on 1 and 845 DF,  p-value: 0.9736
```

C) Diagnostics Test

```
> summary(ivreg5, vcov = NULL, df = Inf, diagnostics = TRUE)

Call:
ivreg(formula = Xeduc_pch ~ aeligibility + shr_Xfood +
rXdrinks +
    rXheal + Xeduc | urban, data = test_data)

Residuals:
     Min       1Q   Median       3Q      Max
-2549129 -2546844 -2513682  3059509  4141429

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    2549129   76581073   0.033    0.973
aeligibility  -5605380  169358011  -0.033    0.974

Diagnostic tests:
                                df1 df2 statistic p-value
Weak instruments (aeligibility)   1 845     0.001 0.97358
Weak instruments (shr_Xfood)      1 845    10.843 0.00103 **
Weak instruments (rXdrinks)       1 845     0.003 0.95951
Weak instruments (rXheal)         1 845     0.016 0.89836
Weak instruments (Xeduc)          1 845     7.041 0.00812 **
Wu-Hausman                        1 840     0.418 0.51797
Sargan                           -4  NA        NA      NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2795000 on Inf degrees of freedom
Multiple R-Squared: -3298,    Adjusted R-squared: -3302
Wald test: 0.001095 on 1 DF,  p-value: 0.9736
```

3.) Difference-in-Differences

```
> ScieboPath <- function(NameofDataset)
+   paste0("https://hochschule-rhein-
waal.sciebo.de/s/ICXJvMpW3QIM2hu/download?path=%2F&files=",
NameofDataset)
>
> library(haven)
```

a) Create an interaction variable by interacting treatment variable

```
Reg1<-
lm(cpexp30_pae~aeligibility+Year+I(aeligibility*Year),data=test_data
)
> summary(Reg1)

Call:
lm(formula = cpexp30_pae ~ aeligibility + Year + I(aeligibility *
    Year), data = test_data)

Residuals:
    Min      1Q  Median      3Q     Max
 -87182  -25396  -11576   11612 1063555

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)               36838       1731  21.286  < 2e-16 ***
aeligibility               2884       2461   1.172    0.241
Year                      41035       2509  16.357  < 2e-16 ***
I(aeligibility * Year)    18086       3534   5.118 3.25e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 54730 on 3838 degrees of freedom
  (70 observations deleted due to missingness)
Multiple R-squared:  0.1869,     Adjusted R-squared:  0.1863
F-statistic: 294.1 on 3 and 3838 DF,  p-value: < 2.2e-16
```

b) Run the same analysis with Xeduc_pch

```
> Reg2<-Reg1<-
lm(Xeduc_pch~aeligibility+Year+I(aeligibility*Year),data=test_data)
> summary(Reg2)

Call:
lm(formula = Xeduc_pch ~ aeligibility + Year + I(aeligibility *
    Year), data = test_data)

Residuals:
    Min      1Q  Median      3Q     Max
 -13975  -10805   -9262   -3949 1056225

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)               10290       1536   6.700 2.48e-11 ***
aeligibility               1091       2187   0.499   0.6180
Year                       3686       2144   1.719   0.0857 .
I(aeligibility * Year)    -2798       3125  -0.896   0.3706
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 42310 on 2946 degrees of freedom
```

```
  (962 observations deleted due to missingness)
Multiple R-squared:  0.001074,   Adjusted R-squared:  5.715e-05
F-statistic: 1.056 on 3 and 2946 DF,  p-value: 0.3665
```

c) Now, redoing part(a), this time by controlling

```
Reg3<-Reg1<-
lm(cpexp30_pae~aeligibility+Year+shr_Xfood+rXdrinks+rXheal+Xeduc+I(a
eligibility*Year),data=test_data)
> summary(Reg3)

Call:
lm(formula = cpexp30_pae ~ aeligibility + Year + shr_Xfood +
    rXdrinks + rXheal + Xeduc + I(aeligibility * Year), data =
test_data)

Residuals:
    Min      1Q  Median      3Q     Max
-137732  -24604  -10232   11859 1061726

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)            3.144e+04  4.544e+03   6.920 5.28e-12 ***
aeligibility           2.985e+03  2.337e+03   1.277    0.202
Year                   3.650e+04  2.463e+03  14.816  < 2e-16 ***
shr_Xfood              2.301e+01  6.034e+01   0.381    0.703
rXdrinks               2.676e-01  5.426e-02   4.931 8.52e-07 ***
rXheal                 2.231e-01  1.330e-02  16.779  < 2e-16 ***
Xeduc                  9.186e-02  1.090e-02   8.427  < 2e-16 ***
I(aeligibility * Year) 1.713e+04  3.361e+03   5.096 3.63e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 51950 on 3832 degrees of freedom
  (72 observations deleted due to missingness)
Multiple R-squared:  0.2686,     Adjusted R-squared:  0.2672
F-statistic:   201 on 7 and 3832 DF,  p-value: < 2.2e-16
```

4.) Propensity score matching

```
library(lmtest)
>               library(AER)
>               library(car)
```

a) Estimating Propensity Scores

```
#   Probit regression with treatment as dependent variable

 Mahmud<-glm(eligilitystatus~regurb+urban+hsize+Xeduc,
family=binomial(link="probit"), test_data)
>               summary(Mahmud)

Call:
```

```
glm(formula = eligilitystatus ~ regurb + urban + hsize +
Xeduc,
    family = binomial(link = "probit"), data = test_data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4822  -1.1119  -0.7717   1.1199   2.1007

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  4.723e-01  1.748e-01   2.702  0.00689 **
regurb      -5.460e-04  5.101e-03  -0.107  0.91477
urban        5.683e-02  1.240e-01   0.458  0.64666
hsize       -9.256e-02  1.313e-02  -7.048 1.81e-12 ***
Xeduc        2.774e-07  4.283e-07   0.648  0.51722
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1588.9  on 1146  degrees of freedom
Residual deviance: 1533.2  on 1142  degrees of freedom
AIC: 1543.2

Number of Fisher Scoring iterations: 4

>              Scores<-
data.frame(pr_score=predict(Mahmud,type="response"),treatment=
Mahmud$model$eligilitystatus)
>              head(Scores)
   pr_score treatment
1 0.3194261         0
2 0.6439115         1
3 0.3911501         0
4 0.3608978         0
5 0.3890053         0
6 0.6068191         0


### b) Define the area of common support and interpret

>              comsup<-
c(min(Scores$pr_score[Scores$treatment==1]),max(Scores$pr_scor
e[Scores$treatment==0]))
>              (comsup)
[1] 0.1100849   0.6666073
>              summary(comsup)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.1101  0.2492  0.3883  0.3883  0.5275  0.6666
```