Stats 504

Educational Performance and School Absences

1. Introduction

Students can be absent for multiple reasons, with may lead to different potential impacts. Health issues, lack of interest in classes, unsatisfactory friendships, etc. may be suggested as reasons for the absences, and such absences may affect their academic performance. To guide the students on the right path, it is indispensable to analyze these absences. As a preliminary analysis, this research would like to examine the Student Performance data set from the UCI Machine Learning Repository on student secondary educational achievement. Specifically, two goals have been achieved via this research. First, the research would analyze the impact of absences on math final grades. Additionally, the research would like to identify student attributes that possibly contribute to absences. As a result, the research showed that there exists a causal relationship between student absences on math final grades. Further, the research stated that the math first-period grade, math final grade, intention for higher education, and absences on Portuguese language course can contribute to absences on mathematics course.

2. Methods

Since the research focused on the potential impact of absences on math grades and the impact of attributes to absences, we used the causal inference model to analyze the relationships. The first method the research used was a Pearson chi-square test. Pearson chi-square test is used to determine how likely it is that any observed difference between the sets arose by chance. The research divided each student into the data based on the 1st, 2nd, and 3rd quarters of student absences. Then, the researchers checked whether there was a significant difference in math final grade in each group of students who were absent more or less than the 1st, 2nd, and 3rd quarters using the Pearson's chi-square test. The second method the researchers used was a generalized boosting for absences. The generalized boosting method estimates the weights for each observation. Finally, the research evaluates each parameter using Survey-weighted generalized linear models with the observation weights from the generalized boosting method.

3. Result

3.1 Data Summary

The research collected 382 data with 63 variables. Among the variables, the research selected two main variable, *Absences.x* (# of school absences in math courses) and *G3.x*(Math final grade), for our first question and five more predictors, *G1.x*(Math first period grade), *Studytime.x*(weekly study time), *Absences.y*(# of school absences in Portuguese language course), *Health.x*(current health status) and *Higher_bin*(# of students who Intend to take higher education) that estimated to have the greatest impact on student absence in math course. The brief summary of the data are represented on the Table 1. Further, the research divided the data into 6 groups based on three criteria; # of student who was absent in math course more/less than 0/3/8 times.

Table 1. Brief Statistic of the data

	description	***Min***	***Q1***	***median***	***Q3***	***Max***	***s.d.***
Absences.x	*# of school absences in math course*	0	0	3	8	75	7.625
G1.x	*Math first period grade (from 0 to 20)*	3	8	10.5	13	19	3.349
G3.x	*Math final grade (from 0 to 20)*	0	8	11	14	20	4.687
Studytime.x	*weekly study time (1: <2 hours, 2: 2 to 5 hours, 3: 5 to 10 hours, or 4: >10 hours)*	1	1	2	2	4	0.846
Absences.y	*# of school absences in Portuguese language course*	0	0	2	6	32	4.906
Health.x	*current health status (from 1 - very bad to 5 - very good)*	1	3	4	5	5	1.400
	description	***True (1)***			***False (0)***		
Higher_bin	*# of students who Intend to take higher education*	364			18		
Absences.0	*# of students who was absent in math course more 3than 0 times*	266			116		
Absences.3	*# of students who was absent in math course more than 3 times*	190			192		
Absences.8	*# of students who was absent in math course more than 8 times*	76			306		

3.2 Mean and 95% CI of the data by groups

We tested whether there were significant differences in the variables in each group. In table 2, the research examined the mean and 95% confidence intervals based on the normal assumptions for six groups divided by the student absences status. Overall, the research could find the difference of G3.x and Absences.y between the group with less absences (Absences.0, Absences.3, Absences.8 = 0) and more absences (Absences.0, Absences.3, Absences.8 = 1) such that the group with less absences tend to have lower G3.x and Absences.y.

Table 2. Mean and 95% CI of the predictors for each student absences group

	Absences.0=1	***Absences.0=0***	***Absences.3=1***	***Absences.3=0***	***Absences.8=1***	***Absences.8=0***
G1.x	11.019 (4.679, 17.359)	10.5 (3.471, 17.529)	10.742 (4.233, 17.251)	10.979 (4.352,17.607)	10.5 (4.175, 16.825)	10.951 (4.330, 17.572)
G3.x	11.271 (4.840, 17.701)	8.362 (-4.346, 21.071)	10.932 (4.417, 17.446)	9.849 (-1.292, 20.989)	10.316 (3.716, 16.916)	10.405 (0.673, 20.137)
Studytime.x	2.038 (0.397, 3.678)	2.026 (0.322, 3.730)	2.026 (0.310, 3.742)	2.042 (0.439, 3.644)	1.895 (0.515, 3.274)	2.069 (0.353, 3.784)
Abesences.y	4.15 (-5.938, 14.238)	2.578 (-5.494, 10.649)	4.984 (-5.727, 15.696)	2.375 (-5.229, 9.979)	7.645 (-5.593, 20.883)	2.686 (-4.632, 10.004)
Health.x	3.53 (0.805, 6.255)	3.69 (0.901, 6.479)	3.474 (0.726, 6.221)	3.682 (0.948, 6.416)	3.526 (0.783, 6.270)	3.592 (0.843, 6.340)
Higher_bin	0.966 (0.611, 1.321)	0.922 (0.396, 1.449)	0.974 (0.659, 1.288)	0.932 (0.439, 1.426)	0.974 (0.658, 1.290)	0.948 (0.511, 1.385)

3.3 Chi-square test

The research used Pearson's chi-square test to check the causality between the absence of math courses and the math final grade. As a result, the research concluded that the p-value for the null hypothesis that the distribution between students who had never been absent (absences.0=0) and students who had been absent at least once (absences.0=1) was less than $2.2 * 10^{-16}$. For the other groups, the research resulted that the p-values for the null hypothesis with absences.3 and absences.8 were $1.2 * 10^{-9}$ and $4.859 * 10^{-5}$. Therefore, we concluded that there exists an impact of absences on math final grades.

3.4 Generalized boosting and Survey-weighted generalised linear models

The research used generalized boosting to calculate the wights of each predictors, then used the survey-weighted generalised linear models to estimate the causality between # of absences and predictors. To test more diverse associations, the research created three models for each of the three responses while maintaining the three dividing criterion of absence; 0/3/8 times of absences. Figure 1 represents the estimated weights of observations.

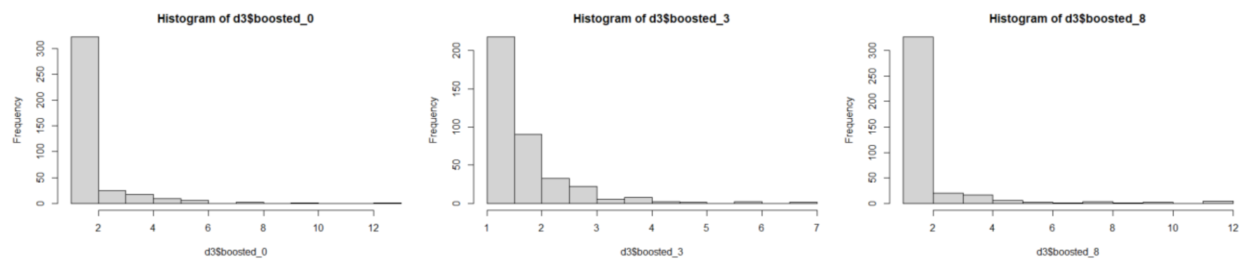


Figure 1. weights of the observations for absences.0 (left), absences.3 (middle), absences.8 (right)

Using those weights, the research designed the Survey-weighted generalised linear models to evaluate the Average treatment effect and calculate the p-value of the effect on absences for each predictor.

Table 3. Estimates and p-value of the predictors for each response

	Absence.0		*Absence.3*		*Absence.8*	
	Estimate	*p-value*	*Estimate*	*p-value*	*Estimate*	*p-value*
G1.x	-0.035	0.02460 *	-0.037	0.0071*	-0.011	0.4260
G3.x	0.043	8.35e-06 *	0.036	3.99e-05 *	0.012	0.1869
Studytime.x	0.014	0.71587	0.014	0.6815	-0.003	0.9385
Higher_bin	0.112	0.46922	0.115	0.5229	0.298	0.0007 *
Absences.y	0.021	0.00126 *	0.016	0.0104 *	0.022	0.0003 *
Health.x	0.002	0.92783	-0.013	0.5473	0.003	0.8789

(*: p-value less than 0.05)

The result states that a student with lower G1.x tends to have a higher absence, and a student with higher G3.x, Studytime.x, higher_bin, absences.y, and health.x tends to have a higher absence, while the estimate for the Studytime.x is negative on the response Absence.8 and the estimate for the health.x is negative on the response Absence.3. Focusing on the p-value, for the criteria with few absences, G1.x, G3.x, and Absences.y were the variables that were statistically significant, but for the criteria with many absences, higher_bin and absences.y were statistically significant. Contrary to our intuition, the predictors Studytime.x and health.x were not the significant predictors.

4. Conclusion

The research analyzed the causal relationship between attributes and the absences on math courses. To analyze the causal relationship between the absences and the math final grade. As a result, the research concluded that there exists an casual relationship between absences and the math final grade with the p-value $2.2 * 10^{-16}$, $1.2 * 10^{-9}$ and $4.859 * 10^{-5}$ for absences.0, absences.3 and absences.8. To analyze the causal relationship between the absences and the student attributes, the research used generalized boosting and the Survey-weighted generalized linear models. As a result, the researchers concluded that G1.x, G3.x, and Absences.y were the statistically significant predictors for the group divided by criteria with few absences, and higher_bin and absences.y were the statistically significant predictors for the group divided by criteria with many absences.

Two things remarkable is that from the second result, the higher the number of absences, the higher the final grade, and the importance of variable changes when the absence criteria changes. These might be because the attributed to various reasons for absences. Students may absence from the class because they did not want to take the class, or did not need to take the class, have some events accidentally, etc. We may think that the main reason for the absences were changed as the criteria for the absences increased. However, although we have data on the number of absences, there is no data on the reasons for the absences. Therefore, we suggest to collect the data about the reasons for the absences to analyze the attributes that can contribute to the absences precisely.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(twang)
```

```
## Warning: ÆÐÅ°Áö 'twang'´Â R ¹öÀü 4.1.2¿¡¼ ÀÛ¼ºµÇ¾ú½À´Ï´Ù
```

```
## To reproduce results from prior versions of the twang package, please see the version="leg
acy" option described in the documentation.
```

```
library(survey)
```

```
## Warning: ÆÐÅ°Áö 'survey'´Â R ¹öÀü 4.1.2¿¡¼ ÀÛ¼ºµÇ¾ú½À´Ï´Ù
```

```
## 필요한 패키지를 로딩중입니다: grid
```

```
## 필요한 패키지를 로딩중입니다: Matrix
```

```
##
## 다음의 패키지를 부착합니다: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```
## 필요한 패키지를 로딩중입니다: survival
```

```
##
## 다음의 패키지를 부착합니다: 'survey'
```

```
## The following object is masked from 'package:graphics':
##
##     dotchart
```

1. impact of absences on math final grades

```
d1=read.table("student-mat.csv",sep=";",header=TRUE)
d2=read.table("student-por.csv",sep=";",header=TRUE)

d3=merge(d1,d2,by=c("school","sex","age","address","famsize","Pstatus","Medu","Fedu","Mjob",
"Fjob","reason","nursery","internet"))
print(nrow(d3))
```

```
## [1] 382
```

```
#G1.x + G2.x +  G3.x + studytime.x + higher_bin + absences.y + health.x
summary(d3$G3.x)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    8.00   11.00   10.39   14.00   20.00
```

```
sd(d3$G3.x)
```

```
## [1] 4.687242
```

```
summary(d3$absences.x)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   0.000   3.000   5.319   8.000  75.000
```

```
sd(d3$absences.x)
```

```
## [1] 7.625251
```

```
summary(d3$G1.x)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    3.00    8.00   10.50   10.86   13.00   19.00
```

```
sd(d3$G1.x)
```

```
## [1] 3.349167
```

```
summary(d3$studytime.x)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   2.000   2.034   2.000   4.000
```

```
sd(d3$studytime.x)
```

```
## [1] 0.845798
```

```
summary(d3$higher_bin)
```

```
## Length  Class   Mode 
##      0   NULL   NULL
```

```
sd(d3$higher_bin)
```

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

```
summary(d3$absences.y)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   0.000   0.000   2.000   3.673   6.000  32.000
```

```
sd(d3$absences.y)
```

```
## [1] 4.905965
```

```
summary(d3$health.x)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   1.000   3.000   4.000   3.579   5.000   5.000
```

```
sd(d3$health.x)
```

```
## [1] 1.40036
```

```
head(d3['G3.x'])
```

```
##   G3.x
## 1   10
## 2    5
## 3   13
## 4    8
## 5   10
## 6   11
```

```
d3 = d3 %>% mutate(absences.0=ifelse(absences.x>0, 1 ,0), absences.3=ifelse(absences.x>3, 1 ,
0), absences.8=ifelse(absences.x>8, 1 ,0))
head(d3[c('absences.0','absences.3','absences.8')])
```

```
##   absences.0 absences.3 absences.8
## 1          1          0          0
## 2          1          0          0
## 3          1          1          0
## 4          1          0          0
## 5          1          1          0
## 6          1          0          0
```

```
mean = d3 %>% filter(d3$absences.0 == 0) %>% select(G3.x) %>% colMeans()
sd = d3 %>% filter(d3$absences.0 == 0) %>% select(G3.x) %>% var() %>% sqrt()
round(mean,3)
```

```
##  G3.x
## 8.362
```

```
c(round(mean - 1.96*sd,3), round(mean + 1.96*sd,3))
```

```
## [1] -4.346 21.071
```

```
chisq.test(d3$G3.x, d3$absences.0)
```

```
## Warning in chisq.test(d3$G3.x, d3$absences.0): Chi-squared approximation may be
## incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  d3$G3.x and d3$absences.0
## X-squared = 122.19, df = 17, p-value < 2.2e-16
```

```
chisq.test(d3$G3.x, d3$absences.3)
```

```
## Warning in chisq.test(d3$G3.x, d3$absences.3): Chi-squared approximation may be
## incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  d3$G3.x and d3$absences.3
## X-squared = 77.107, df = 17, p-value = 1.246e-09
```

```
chisq.test(d3$G3.x, d3$absences.8)
```

```
## Warning in chisq.test(d3$G3.x, d3$absences.8): Chi-squared approximation may be
## incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  d3$G3.x and d3$absences.8
## X-squared = 49.607, df = 17, p-value = 4.859e-05
```

2. attributes that possibly contribute to absences

```
colnames(d3)
```

```
##  [1] "school"       "sex"          "age"          "address"      "famsize"
##  [6] "Pstatus"      "Medu"         "Fedu"         "Mjob"         "Fjob"
## [11] "reason"       "nursery"      "internet"     "guardian.x"   "traveltime.x"
## [16] "studytime.x"  "failures.x"   "schoolsup.x"  "famsup.x"     "paid.x"
## [21] "activities.x" "higher.x"     "romantic.x"   "famrel.x"     "freetime.x"
## [26] "goout.x"      "Dalc.x"       "Walc.x"       "health.x"     "absences.x"
## [31] "G1.x"         "G2.x"         "G3.x"         "guardian.y"   "traveltime.y"
## [36] "studytime.y"  "failures.y"   "schoolsup.y"  "famsup.y"     "paid.y"
## [41] "activities.y" "higher.y"     "romantic.y"   "famrel.y"     "freetime.y"
## [46] "goout.y"      "Dalc.y"       "Walc.y"       "health.y"     "absences.y"
## [51] "G1.y"         "G2.y"         "G3.y"         "absences.0"   "absences.3"
## [56] "absences.8"
```

We will test G1.x, G3.x, studytime.x, higher_bin, absences.y, health.x as a potential attributes that possibly contribute to absences.

```
count(d3[d3$higher.x == 'no',])
```

```
##    n
## 1 18
```

```
d3 = d3 %>% mutate(higher_bin = ifelse(higher.x == 'no', 0, 1))
boosted.mod_0 = ps(absences.0 ~ G1.x +  G3.x + studytime.x + higher_bin + absences.y + healt
h.x,
data=d3,
estimand = "ATE",
n.trees = 10000,
interaction.depth=2,
perm.test.iters=0,
verbose=FALSE,
stop.method = c("es.mean"))
```

```
boosted.mod_3 = ps(absences.3 ~ G1.x + G3.x + studytime.x + higher_bin + absences.y + health.
x,
data=d3,
estimand = "ATE",
n.trees = 10000,
interaction.depth=2,
perm.test.iters=0,
verbose=FALSE,
stop.method = c("es.mean"))
```

```
boosted.mod_8 = ps(absences.8 ~ G1.x +  G3.x + studytime.x + higher_bin + absences.y + healt
h.x,
data=d3,
estimand = "ATE",
n.trees = 10000,
interaction.depth=2,
perm.test.iters=0,
verbose=FALSE,
stop.method = c("es.mean"))
```

```
## Warning in ps.fast(formula = formula, data = data, n.trees = nrounds,
## interaction.depth = max_depth, : Optimal number of iterations is close to the
## specified n.trees. n.trees is likely set too small and better balance might be
## obtainable by setting n.trees to be larger.
```

```
summary(boosted.mod_0)
```

```
##             n.treat n.ctrl ess.treat  ess.ctrl    max.es   mean.es    max.ks
## unw             266    116  266.0000 116.00000 0.6205372 0.2383435 0.3362069
## es.mean.ATE     266    116  253.6021  75.59422 0.2845126 0.1233851 0.1307066
##             max.ks.p   mean.ks iter
## unw               NA 0.1448773   NA
## es.mean.ATE       NA 0.0625678  859
```

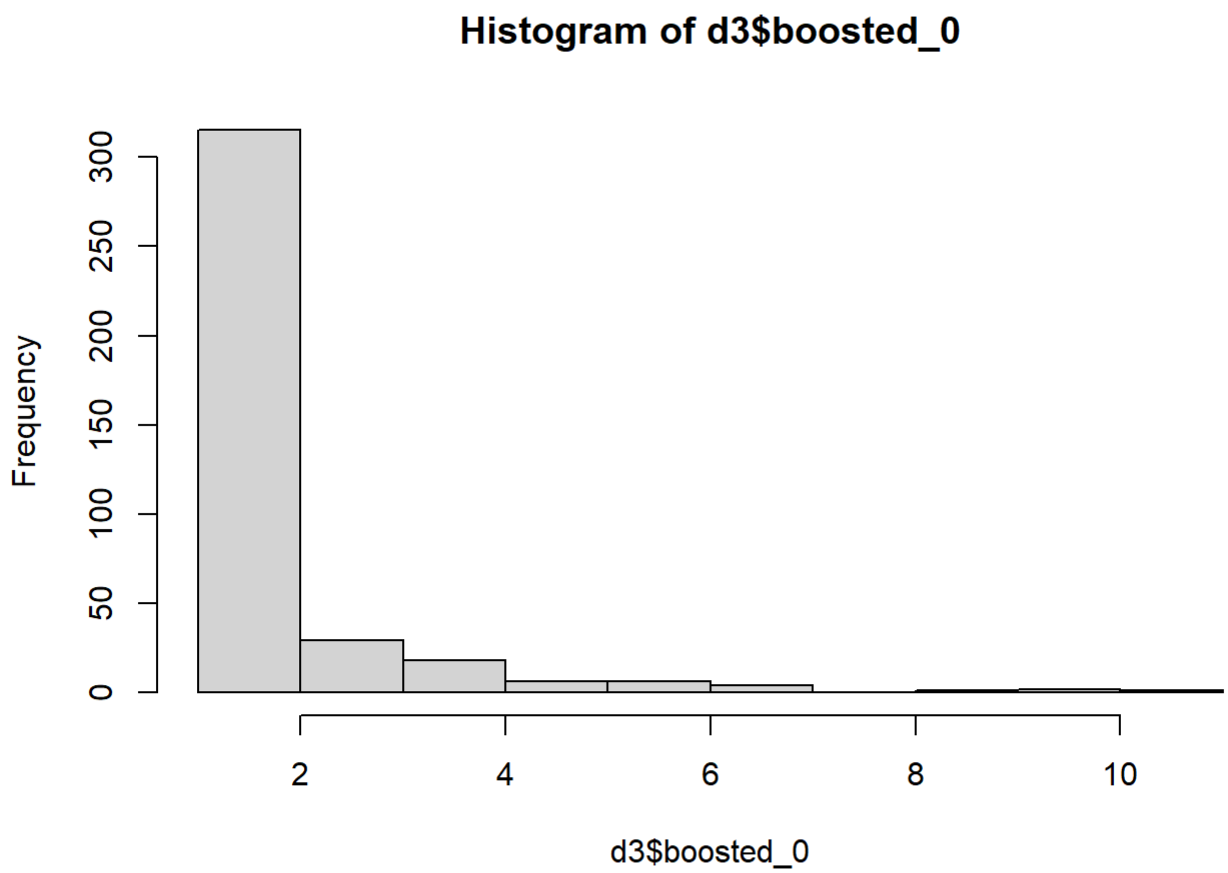
```
summary(boosted.mod_0$gbm.obj, n.trees=boosted.mod_0$desc$es.mean.ATE$n.trees, plot=F)
```

```
##                     var     rel.inf
## G3.x               G3.x 70.11423414
## absences.y   absences.y 18.43222035
## G1.x               G1.x  8.18497474
## health.x       health.x  2.38021054
## studytime.x studytime.x  0.87262597
## higher_bin   higher_bin  0.01573427
```

```
d3$boosted_0 <- get.weights(boosted.mod_0)
```

```
## Warning in get.weights(boosted.mod_0): No stop.method specified.  Using es.mean.ATE
```

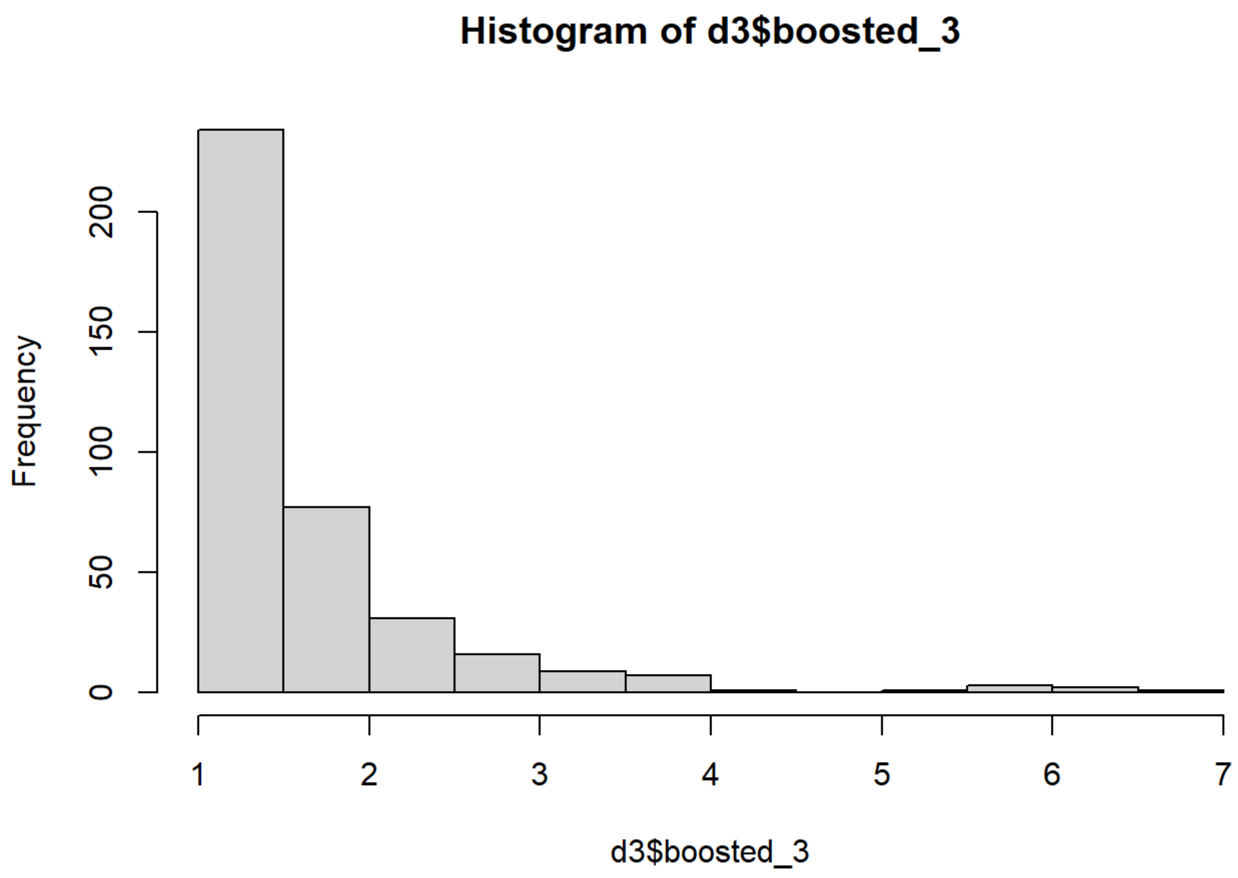
```
hist(d3$boosted_0)
```




```
d3$boosted_3 <- get.weights(boosted.mod_3)
```

```
## Warning in get.weights(boosted.mod_3): No stop.method specified.  Using es.mean.ATE
```

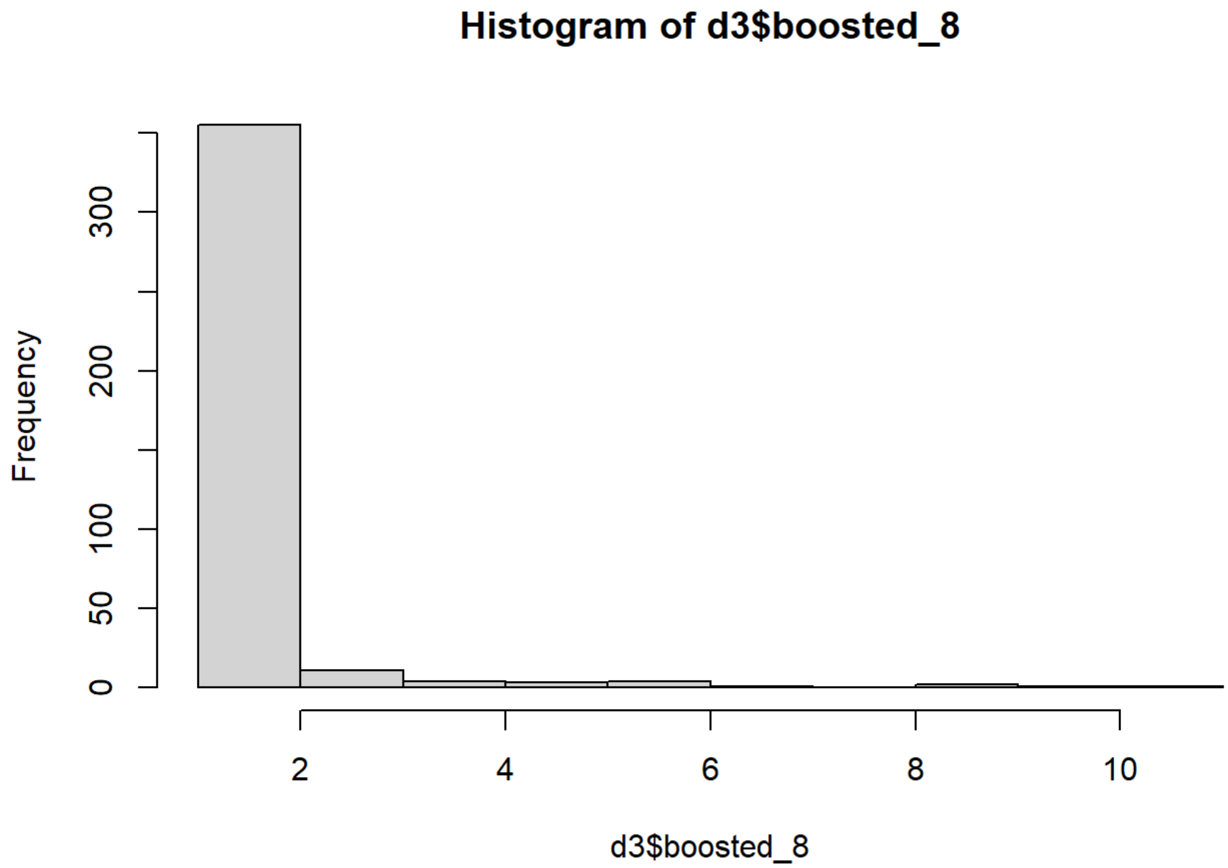
```
hist(d3$boosted_3)
```



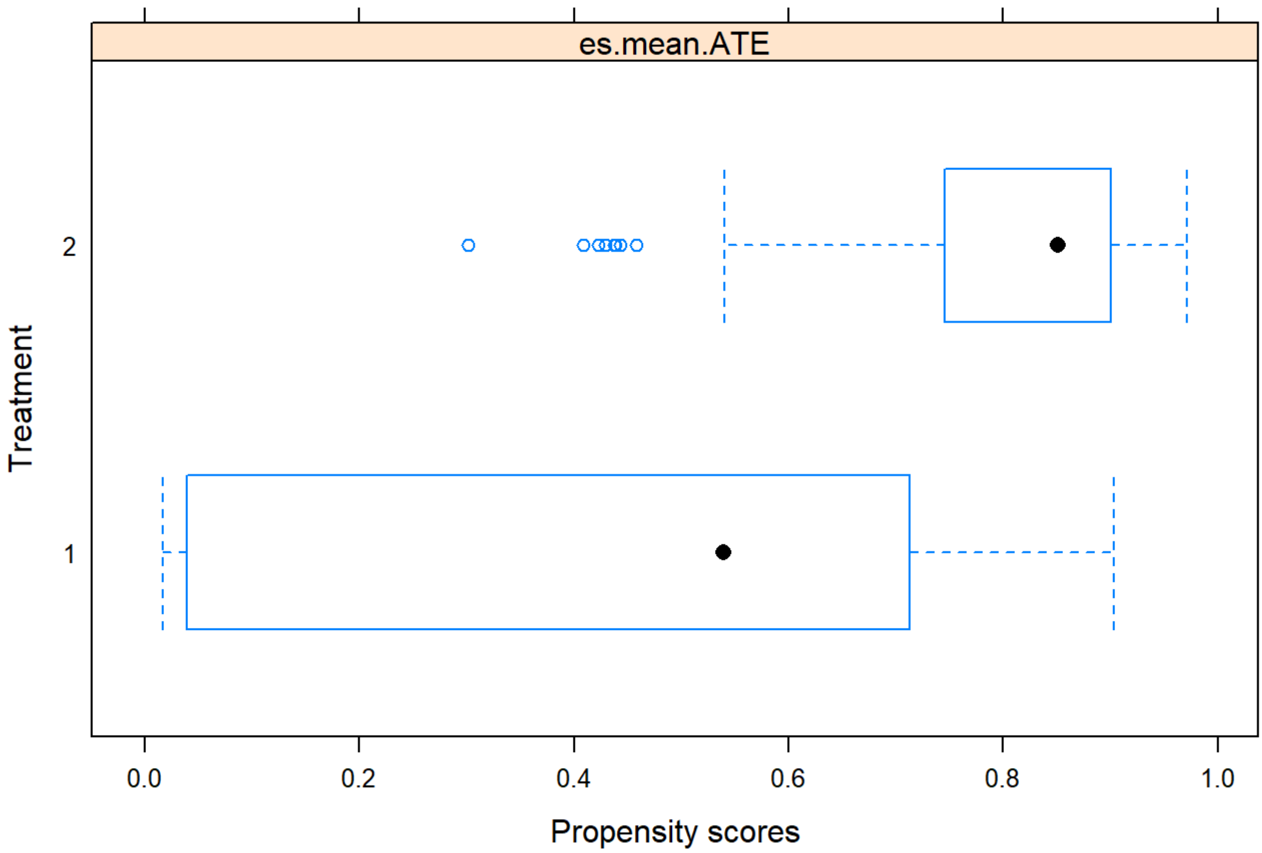

```
d3$boosted_8 <- get.weights(boosted.mod_8)
```

```
## Warning in get.weights(boosted.mod_8): No stop.method specified.  Using es.mean.ATE
```

```
hist(d3$boosted_8)
```




```
plot(boosted.mod_0, plot=2)
```



```
bal.table(boosted.mod_0)
```

```
## $unw
##              tx.mn tx.sd  ct.mn ct.sd std.eff.sz   stat     p    ks ks.pval
## G1.x        11.019 3.235 10.500 3.586      0.155  1.342 0.180 0.144   0.072
## G3.x        11.271 3.281  8.362 6.484      0.621  4.595 0.000 0.336   0.000
## studytime.x  2.038 0.837  2.026 0.869      0.014  0.123 0.902 0.034   1.000
## higher_bin   0.966 0.181  0.922 0.269      0.206  1.606 0.109 0.044   0.998
## absences.y   4.150 5.147  2.578 4.118      0.321  3.179 0.002 0.217   0.001
## health.x     3.530 1.390  3.690 1.423     -0.114 -1.017 0.310 0.095   0.461
##
## $es.mean.ATE
##              tx.mn tx.sd  ct.mn ct.sd std.eff.sz   stat     p    ks ks.pval
## G1.x        11.209 3.257 10.823 3.286      0.115  0.983 0.326 0.085   0.797
## G3.x        11.498 3.307 10.164 4.874      0.285  2.981 0.003 0.131   0.273
## studytime.x  2.045 0.837  2.031 0.838      0.016  0.135 0.893 0.015   1.000
## higher_bin   0.966 0.180  0.948 0.224      0.088  0.757 0.450 0.019   1.000
## absences.y   3.597 4.885  2.470 3.325      0.230  2.476 0.014 0.087   0.774
## health.x     3.576 1.394  3.584 1.445     -0.006 -0.045 0.964 0.039   1.000
```

```
design_0 <- svydesign(ids=~1, weights=~boosted_0, data=d3)
glm0 <- svyglm(absences.0 ~ G1.x + G3.x + studytime.x + higher_bin + absences.y + health.x, d
esign=design_0)
summary(glm0)
```

```
##
## Call:
## svyglm(formula = absences.0 ~ G1.x + G3.x + studytime.x + higher_bin +
##     absences.y + health.x, design = design_0)
##
## Survey design:
## svydesign(ids = ~1, weights = ~boosted_0, data = d3)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.231554   0.203880   1.136  0.25679
## G1.x        -0.034898   0.015463  -2.257  0.02460 *
## G3.x         0.043468   0.009619   4.519 8.35e-06 ***
## studytime.x  0.013652   0.037477   0.364  0.71587
## higher_bin   0.111795   0.154308   0.724  0.46922
## absences.y   0.020522   0.006314   3.250  0.00126 **
## health.x     0.002098   0.023150   0.091  0.92783
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2325921)
##
## Number of Fisher Scoring iterations: 2
```

```
round(glm0$coefficients,3)
```

```
## (Intercept)        G1.x        G3.x studytime.x  higher_bin  absences.y
##       0.232      -0.035       0.043       0.014       0.112       0.021
##    health.x
##       0.002
```

```
design_3 <- svydesign(ids=~1, weights=~boosted_3, data=d3)
glm3 <- svyglm(absences.3 ~ G1.x + G3.x + studytime.x + higher_bin + absences.y + health.x, d
esign=design_3)
summary(glm3)
```

```
##
## Call:
## svyglm(formula = absences.3 ~ G1.x + G3.x + studytime.x + higher_bin +
##     absences.y + health.x, design = design_3)
##
## Survey design:
## svydesign(ids = ~1, weights = ~boosted_3, data = d3)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.358130   0.223681   1.601   0.1102
## G1.x        -0.037000   0.013668  -2.707   0.0071 **
## G3.x         0.036079   0.008678   4.158 3.99e-05 ***
## studytime.x  0.014172   0.034499   0.411   0.6815
## higher_bin   0.115045   0.179905   0.639   0.5229
## absences.y   0.016093   0.006253   2.574   0.0104 *
## health.x    -0.012669   0.021033  -0.602   0.5473
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.238003)
##
## Number of Fisher Scoring iterations: 2
```

```
round(glm3$coefficients,3)
```

```
## (Intercept)        G1.x        G3.x studytime.x  higher_bin  absences.y
##       0.358      -0.037       0.036       0.014       0.115       0.016
##    health.x
##      -0.013
```

```
design_8 <- svydesign(ids=~1, weights=~boosted_8, data=d3)
glm8 <- svyglm(absences.8 ~ G1.x + G3.x + studytime.x + higher_bin + absences.y + health.x, d
esign=design_8)
summary(glm8)
```

```
## 
## Call:
## svyglm(formula = absences.8 ~ G1.x + G3.x + studytime.x + higher_bin + 
##     absences.y + health.x, design = design_8)
## 
## Survey design:
## svydesign(ids = ~1, weights = ~boosted_8, data = d3)
## 
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -0.032036   0.143407  -0.223 0.823354    
## G1.x        -0.011477   0.014400  -0.797 0.425964    
## G3.x         0.011719   0.008863   1.322 0.186862    
## studytime.x -0.003238   0.041951  -0.077 0.938509    
## higher_bin   0.297651   0.087153   3.415 0.000707 ***
## absences.y   0.022392   0.006181   3.623 0.000332 ***
## health.x     0.003495   0.022927   0.152 0.878934    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for gaussian family taken to be 0.2117372)
## 
## Number of Fisher Scoring iterations: 2
```

```
round(glm8$coefficients,3)
```

```
## (Intercept)        G1.x        G3.x studytime.x  higher_bin  absences.y 
##      -0.032      -0.011       0.012      -0.003       0.298       0.022 
##    health.x 
##       0.003
```

```
cor(d3$G3.x, d3$absences.x)
```

```
## [1] 0.02898725
```