Barrera-Osorio et al 2011

Dimitrios & Eva

2023-03-08

Contents

1 Motivation

1.1 Why is this research question relevant?

Education in Colombia and other middle-income countries face challenges such as high dropout rates among low-income students, and the reasons behind them, such as the high cost of education. Conditional cash transfers (CCTs) are an evidence-based intervention to increase participation in education. However, the authors highlight that there is little variability in the structure of programmes. The paper investigates if changes in the timing of payments affect the outcomes of interest: Attendance and re-enrollment. Optimising the structure of CCTs may contribute to improved education outcomes and reduce disparities in access to education.

1.2 What are the main hypotheses?

- The savings model will improve outcomes compared to the basic programme by relaxing possible savings constraints.
- The tertiary model will improve rates of graduation and tertiary enrollment compared to the basic programme by providing direct incentives for continuation of education.

2 Data sources

We investigated performing the replication with the data provided as part of the lecture. However, we soon discovered that the file did not contain all required variables, nor was there any meta data or other information about the variables in the dataset. A brief search revealed that the data and STATA scripts used to obtain the authors' results are freely available here.

For this project, we reference the following files: * Data file: Public_Data_AEJApp_2010-0132.dta * STATA script for Table 3: Table_03_Attendance.do * Meta data: AEJApp_2010-0132_Data_ReadMe.pdf

2.1 Where does the data come from (country, time period, source)?

Data were collected in San Christobal ("Basic" and "Savings" experiments) and in Suba ("Tertiary" experiment) and combined from six different sources: 1. SISBEN surveys 2003 and 2004: Baseline data on eligible families 2. Programme registration data: Basic information on students 3. Administrative records: Enrollment records 4. Direct observation in 68 out of 251 schools: Attendance data in last quarter of 2005 for 7,158 students. 5. Survey in 68 schools: Baseline data collection in 2005. 6. Survey in 68 schools: Follow-up in 2006.

2.2 What are the key variables and ow are these measured?

Key variables for the replication of Table 3 are the outcome variable, *at_msamean*. This measures the percentage of days absent using a verified attendance measure (see metadata doc) and takes values between 0 and 1 (scale).

Additionally,

3 Method

3.1 Research design

The research paper describes three interventions designed to improve attendance and educational outcomes for students in Colombia.

The first intervention ("basic") is similar to the PROGRESA/OPORTUNIDADES program, a conditional cash transfer program in Mexico that operated from 1997 to 2012. It pays participants 30,000 pesos per month (approximately USD 15) if the child attends at least 80% of the days in that month. Payments are made bi-monthly through a dedicated debit card, and students will be removed from the program if they fail to meet attendance targets or are expelled from school.

The second intervention, called the savings treatment, pays two-thirds of the monthly amount (20,000 pesos or USD 10) to students' families on a bi-monthly basis, while the remaining one-third is held in a bank account. The accumulated funds are then made available to students' families during the period in which students prepare to enroll for the next school year, with 100,000 pesos (US$50) available to them in December if they reach the attendance target every month.

The third intervention, called the tertiary treatment, incentivizes students to graduate and matriculate to a higher education institution. The monthly transfer for good attendance is reduced from 30,000 pesos per month to 20,000 pesos, but upon graduating, the student earns the right to receive a transfer of 600,000 pesos (USD 300) if they enroll in a tertiary institution, and after a year if they fail to enroll upon graduation.

Students were removed from the program if they fail to meet attendance targets, fail to matriculate to the next grade twice, or are expelled from school.

In our replication, we focus on the first and second intervention.

The eligibility criteria for the "basic" and "savings" experiments were as follows:

- Children had to have finished grade 5 and be enrolled in grades 6 - 10.
- The children's families had to be classified into the bottom two categories on Colombia's poverty index (SISBEN).
- Only households living in San Cristobal prior to 2004 were eligible to participate.

The paper investigates differences in enrollment and graduation / progression to tertiary education for the three treatment groups compared to untreated controls. Randomization to treatment vs control group was stratified by location, school public vs private, gender and grade.

3.2 Data preparation

We imported the data file from STATA format and prepared it for analysis by first turning categorical variables into factors. For convenience when producing graphs, we combined the three treatment indicators into a single factor variable with four expressions (0 = control group, 1 = T1, 2 = T2, 3 = T3).

We then translated the STATA commands to filter the data in line with the inclusion criteria:

- Dropping ineligible cases from Suba: Drop if suba == 1
- Keeping only those who were selected for the survey in schools: survey_selected == 1
- Drop if grade is < 6 or grade is 11

The dataset for our analysis is called *filtered_barrera*.

3.3 Analysis

3.3.1 What are the assumptions of the method?

The authors initially use simple linear regression to compare treatment groups. They model the relationship between a dependent variable (outcome; attendance) and two independent variables (whether participant is allocated to treatment "basic", and whether participant is allocated to treatment "savings".)

The assumptions about the data underlying linear regression are:

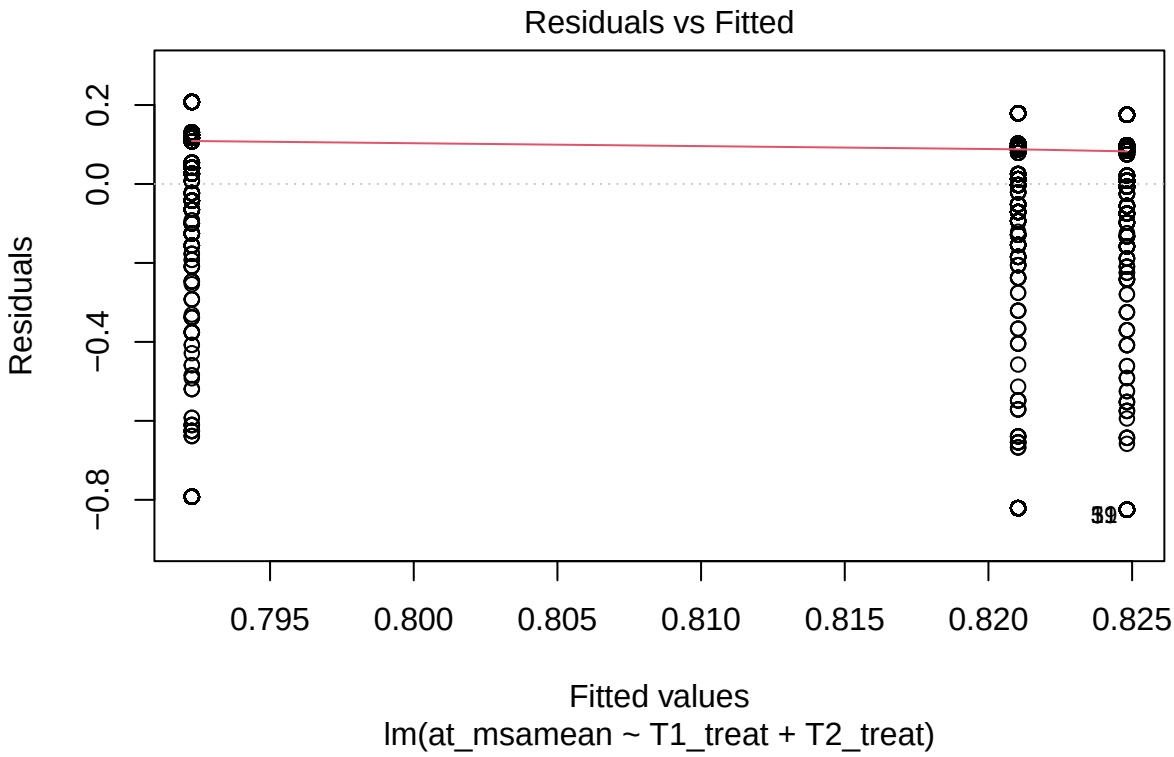
1. Linearity: There should be a linear relationship between the independent and dependent variables.
2. Independence: The observations used in the regression analysis should be independent of each other. In other words, the value of one observation should not be influenced by the value of another observation.
3. Homoscedasticity: The variance of the dependent variable should be constant across all values of the independent variable(s).
4. Normality: The dependent variable should be normally distributed at each level of the independent variable(s).
5. No multicollinearity: If there are multiple independent variables in the regression model, there should be no high correlation between these independent variables.

If these assumptions are not met, this can lead to unreliable estimators (regression coefficients) and / or biased standard errors, i.e. standard errors that are systematically smaller or larger than the "true" standard error. This means that the relationship between dependent and independent variables is not estimated correctly by the model.

3.3.2 Are these assumptions plausible in this example?

We test the assumptions of the simplest regression model using the procedure detailed here.

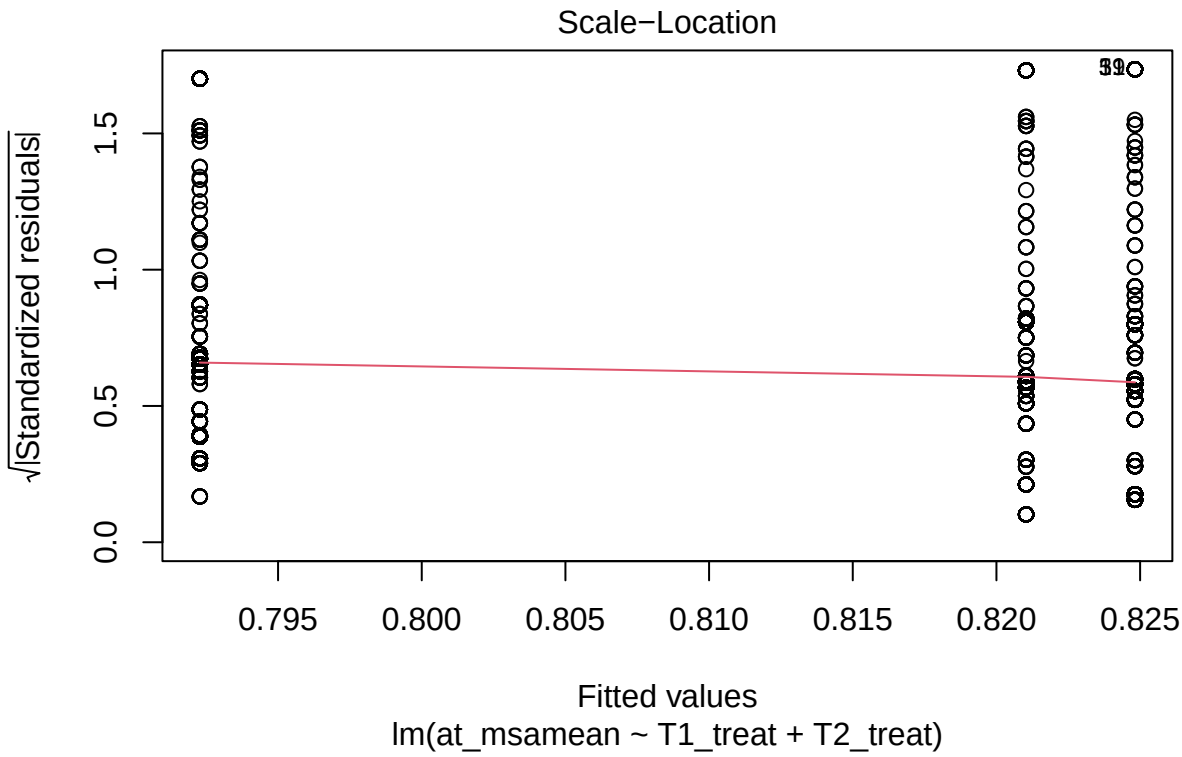
```
# Setting up model
mod0 <- lm(data = filtered_barrera, at_msamean ~ T1_treat + T2_treat)

# 1. Linearity and 3. heteroskedasticity
plot(mod0, 1)
```




```
# The plot is not what we would typically expect if these assumptions were fulfilled.

# 2. Independence
durbinWatsonTest(mod0)
```

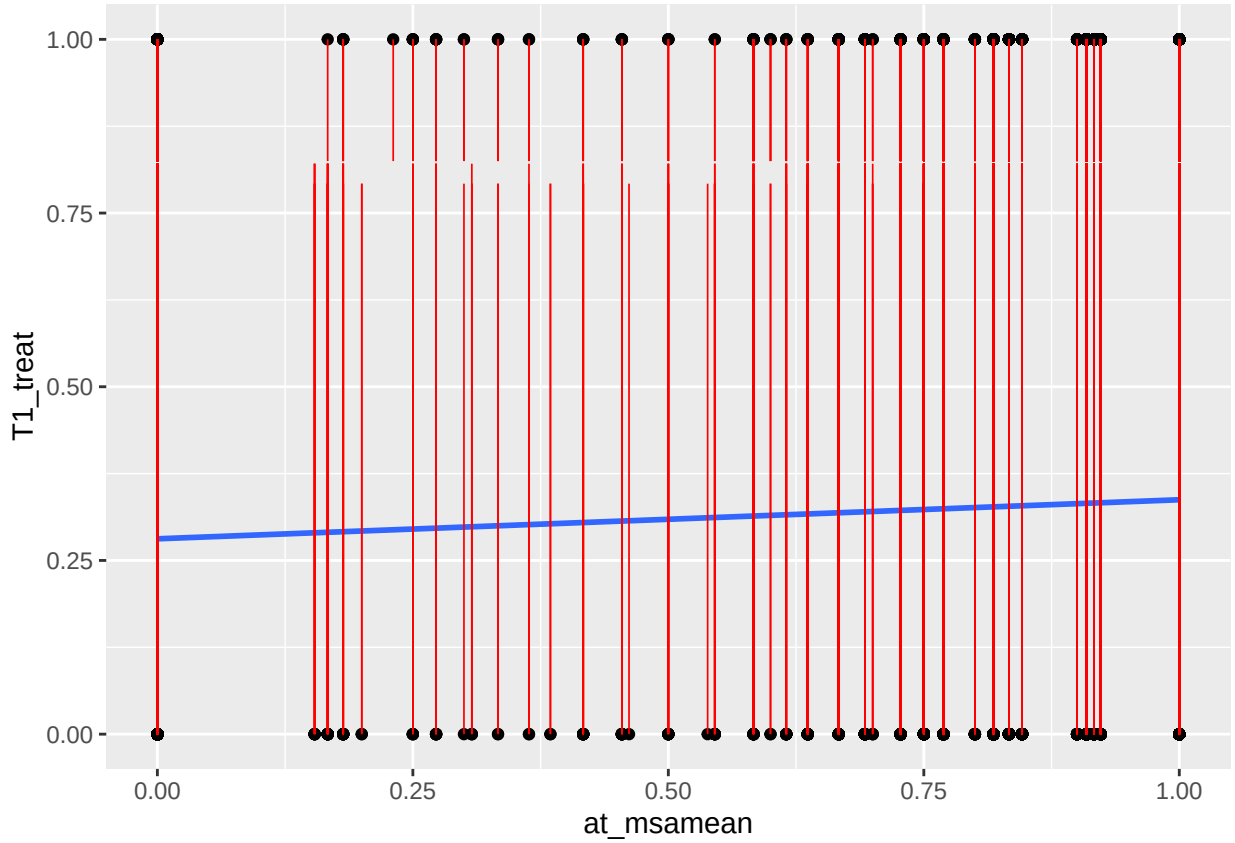
```
##  lag Autocorrelation D-W Statistic p-value
##    1     0.004010807      1.990513   0.748
##  Alternative hypothesis: rho != 0
```

```
# A result for the p-value > 0.05 would suggest we can reject the Null hypothesis and the assumption is

# 4. Normality
plot(mod0, 3)
```




```
# This is again not a typical plot.


# Plotting fitted vs actual values for T1

# Souce: http://www.sthda.com/english/articles/39-regression-model-diagnostics/161-linear-regression-as

model.diag.mod0 <- augment(mod0)

ggplot(model.diag.mod0, aes(at_msamean, T1_treat)) +
  geom_point() +
  stat_smooth(method = lm, se = FALSE) +
  geom_segment(aes(xend = at_msamean, yend = .fitted), color = "red", linewidth = 0.3)
```

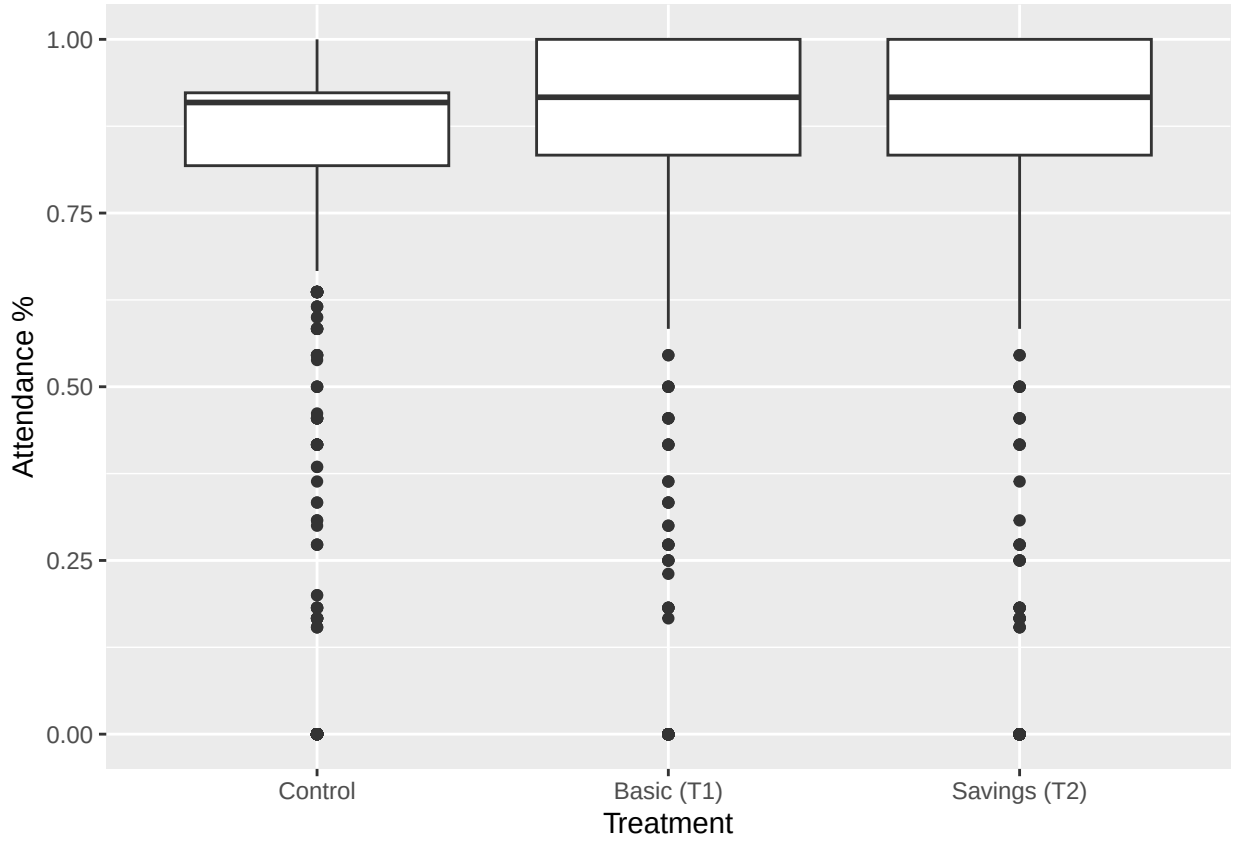
```
## `geom_smooth()` using formula = 'y ~ x'
```



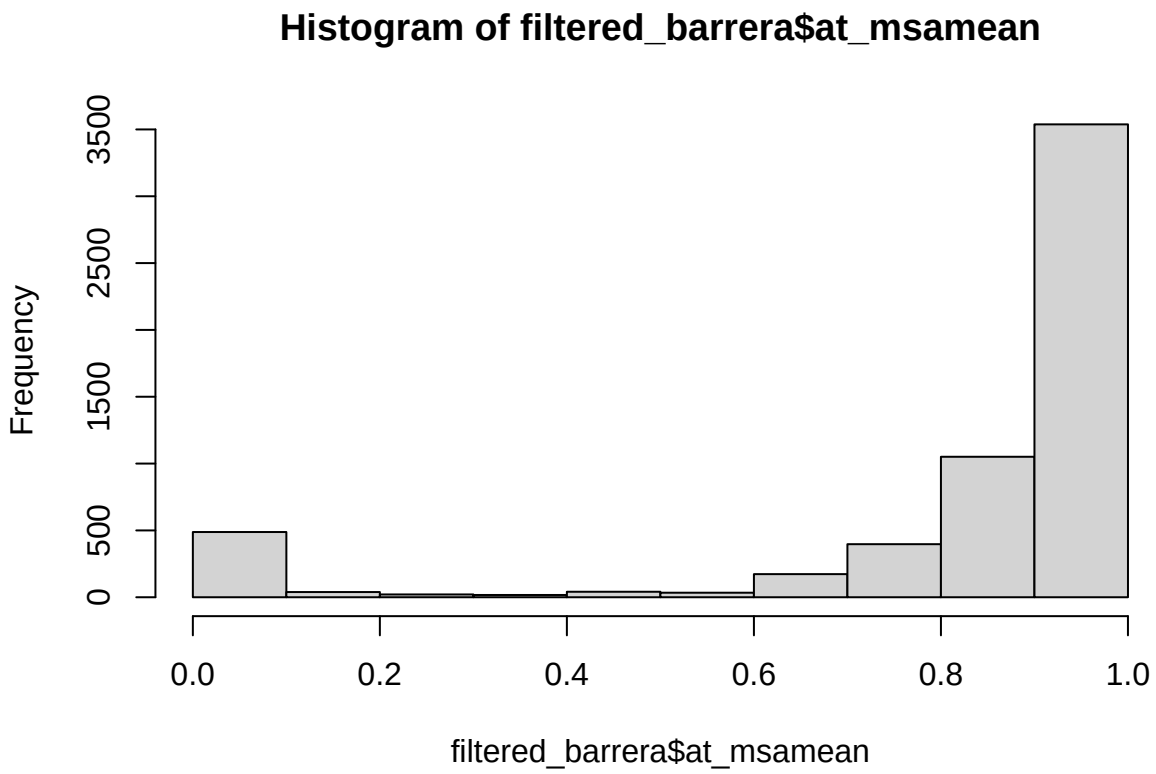

```
# And another atypical plot.
```

The plots reflect the model set-up, where averages are estimated by treatment group. There appears to be little variation within groups. This is further explored graphically below.

```
# Plotting the outcome variable

# A boxplot for each group
ggplot(filtered_barrera, aes(x = T1T2T3, y = at_msamean)) +
  geom_boxplot() +    # Box plot for visualization
  labs(x = "Treatment", y = "Attendance %")  # Label the axes
```

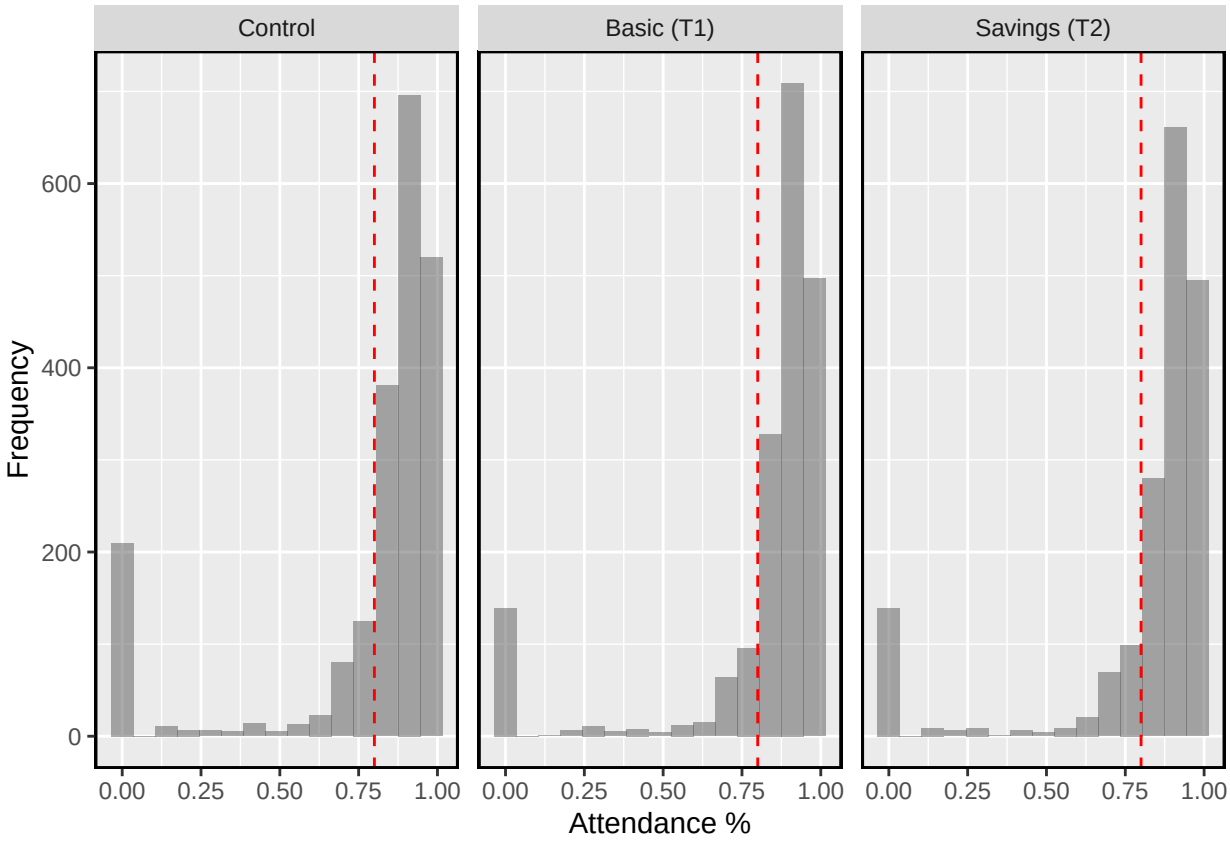



```
# Histogram of the outcome variable

hist(filtered_barrera$at_msamean)
```




```
# Create separate histograms of at_msamean for each level of T1T2T3

ggplot(filtered_barrera, aes(x = at_msamean)) +
  geom_histogram(binwidth = 0.07, alpha = 0.5, position = "identity") +
  labs(x = "Attendance %", y = "Frequency") +
  facet_wrap(~T1T2T3, ncol = 3) +
  geom_vline(xintercept = 0.8, color = "red", linetype = "dashed")+
  theme(panel.border = element_rect(colour = "black", fill = NA, size = 1),
        panel.spacing = unit(0.5, "lines"))
```

```
## Warning: The ‘size‘ argument of ‘element_rect()‘ is deprecated as of ggplot2 3.4.0.
## i Please use the ‘linewidth‘ argument instead.
```



We can see that a large proportion of each sample has an attendance record above the 80% requirement, explaining the skew of the distribution observed in the box plots. It is therefore unlikely that

3.4 Model specifications

In addition to the above, one violation that should be expected based on the data is that of independence. It is likely that there are unobserved characteristics at the school level (e.g. school culture and rules) that affect the outcome. The following equations show formally how the analyses were conceptualised(where i is the individual and j is the school). While the authors state that standard errors are clustered also clustered within the individual, the STATA code suggests that the only clustering variable was the school (*school code*), and this is what we have replicated below.

Model 1:

$$y_{ij} = \beta_0 + \beta_B Basic_i + \beta_S Savings_i + \epsilon_{ij}$$

Model 1 is a simple linear model with only treatment allocation as the dependent variable, while model 2 also includes a collection of student and household characteristics.

Model 2:

$$y_{ij} = \beta_0 + \beta_B Basic_i + \beta_S Savings_i + \delta X_{ijk} + \theta_j + \epsilon_{ij}$$

Model 3 builds on model two, but includes a fixed effect for the school level. Fixed effect models are used to control for unobserved factors that affect the outcome variable. In this case, the school was chosen as the fixed effect.

4 Descriptive statistics

- Describe sample

	Control	Basic (T1)	Savings (T2)	Overall
	(N=2096)	(N=1895)	(N=1808)	(N=5799)
factor(f_teneviv)				
1	1160 (55.3%)	1013 (53.5%)	992 (54.9%)	3165 (54.6%)
2	150 (7.2%)	133 (7.0%)	141 (7.8%)	424 (7.3%)
3	570 (27.2%)	524 (27.7%)	503 (27.8%)	1597 (27.5%)
4	216 (10.3%)	225 (11.9%)	172 (9.5%)	613 (10.6%)
s_utilities				
Mean (SD)	4.64 (1.42)	4.62 (1.40)	4.69 (1.39)	4.65 (1.40)
Median [Min, Max]	5.00 [1.00, 6.00]	5.00 [1.00, 6.00]	5.00 [1.00, 6.00]	5.00 [1.00, 6.00]
s_durables				
Mean (SD)	1.35 (0.883)	1.32 (0.881)	1.39 (0.871)	1.35 (0.879)
Median [Min, Max]	1.00 [0, 4.00]	1.00 [0, 4.00]	1.00 [0, 4.00]	1.00 [0, 4.00]
s_infraest_hh				
Mean (SD)	11.6 (1.75)	11.5 (1.82)	11.7 (1.63)	11.6 (1.74)
Median [Min, Max]	12.0 [3.00, 19.0]	12.0 [3.00, 18.0]	12.0 [3.00, 17.0]	12.0 [3.00, 19.0]
s_age_sorteo				
Mean (SD)	14.1 (5.42)	14.2 (5.56)	13.9 (5.04)	14.1 (5.35)
Median [Min, Max]	13.0 [4.00, 72.0]	13.0 [1.00, 76.0]	13.0 [3.00, 78.0]	13.0 [1.00, 78.0]
factor(f_sexo)				
Female	1055 (50.3%)	931 (49.1%)	916 (50.7%)	2902 (50.0%)
Male	1041 (49.7%)	964 (50.9%)	892 (49.3%)	2897 (50.0%)
Years of Education				
Mean (SD)	5.34 (1.72)	5.27 (1.70)	5.29 (1.69)	5.30 (1.70)
Median [Min, Max]	5.00 [0, 14.0]	5.00 [0, 16.0]	5.00 [0, 12.0]	5.00 [0, 16.0]
factor(f_single)				
No	1492 (71.2%)	1334 (70.4%)	1270 (70.2%)	4096 (70.6%)
Yes	604 (28.8%)	561 (29.6%)	538 (29.8%)	1703 (29.4%)
Age of Jefe del Hogar				
Mean (SD)	45.6 (10.3)	45.5 (9.74)	45.8 (9.80)	45.6 (9.97)
Median [Min, Max]	43.0 [19.0, 91.0]	43.0 [23.0, 98.0]	44.0 [24.0, 84.0]	43.0 [19.0, 98.0]
Years of Education Jefe				
Mean (SD)	5.59 (2.92)	5.56 (2.80)	5.49 (2.89)	5.55 (2.87)
Median [Min, Max]	5.00 [0, 22.0]	5.00 [0, 15.0]	5.00 [0, 16.0]	5.00 [0, 22.0]
Number of people in the household				
Mean (SD)	5.40 (1.94)	5.44 (1.93)	5.41 (1.93)	5.42 (1.93)
Median [Min, Max]	5.00 [2.00, 19.0]	5.00 [2.00, 19.0]	5.00 [2.00, 19.0]	5.00 [2.00, 19.0]
Number of kids 18 and under				
Mean (SD)	2.63 (1.32)	2.71 (1.35)	2.66 (1.33)	2.67 (1.33)
Median [Min, Max]	2.00 [0, 11.0]	3.00 [0, 12.0]	2.00 [0, 12.0]	2.00 [0, 12.0]
factor(f_estrato)				
0	440 (21.0%)	408 (21.5%)	379 (21.0%)	1227 (21.2%)
1	292 (13.9%)	256 (13.5%)	267 (14.8%)	815 (14.1%)
2	1364 (65.1%)	1231 (65.0%)	1162 (64.3%)	3757 (64.8%)
SISBEN score				
Mean (SD)	11.7 (4.64)	11.5 (4.51)	11.5 (4.52)	11.6 (4.56)
Median [Min, Max]	12.4 [1.92, 21.9]	12.4 [2.28, 21.8]	12.3 [1.82, 22.0]	12.3 [1.82, 22.0]
Household Income				
Mean (SD)	367 (239)	358 (240)	368 (226)	364 (235)
Median [Min, Max]	332 [0, 3320]	330 [0, 4000]	332 [0, 1730]	332 [0, 4000]

5 Results

- Are these results plausible?
- How robust are the results to changing the sample?

```
## The variables 'f_estrato3', 'f_grade10' and two others have been removed because of collinearity (se
## The variables 'f_estrato3', 'f_grade10' and two others have been removed because of collinearity (se
```

```
## Warning in kable_styling(kable_input, "none", htmltable_class = light_class, :
## Please specify format in kable. kableExtra can customize either HTML or LaTeX
## outputs. See https://haozhu233.github.io/kableExtra/ for details.
```

```
## Warning in pack_rows(kable_classic(modelsummary(list(feols_m1, feols_m2, :
## Please specify format in kable. kableExtra can customize either HTML or LaTeX
## outputs. See https://haozhu233.github.io/kableExtra/ for details.
```

```
## Warning in add_header_above(pack_rows(kable_classic(modelsummary(list(feols_m1,
## : Please specify format in kable. kableExtra can customize either HTML or LaTeX
## outputs. See https://haozhu233.github.io/kableExtra/ for details.
```

```
## Warning in
## footnote(add_header_above(pack_rows(kable_classic(modelsummary(list(feols_m1, :
## Please specify format in kable. kableExtra can customize either HTML or LaTeX
## outputs. See https://haozhu233.github.io/kableExtra/ for details.
```

Table 1: Table 3 - Effects on Monitored School Attendance Rates

	(1)	(2)	(3)
Basic treatment	0.033***	0.032***	0.032***
	(0.007)	(0.008)	(0.007)
Savings treatment	0.029**	0.027**	0.027***
	(0.008)	(0.008)	(0.007)
Chi-squared	0.31	0.40	0.48
p-value	0.58	0.52	0.49
Num.Obs.	5799	5799	5799
R2	0.003	0.037	0.089
Std.Errors	by: school_code	by: school_code	by: school_code
FE: school_code			X

Note: ^^ + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

6 Graph

ggplot(data=filtered_barrera, aes(x=at_baseline, y=at_msamean, color=factor(T1T2T3))) + geom_point()
+ geom_smooth(method="lm", se=FALSE)

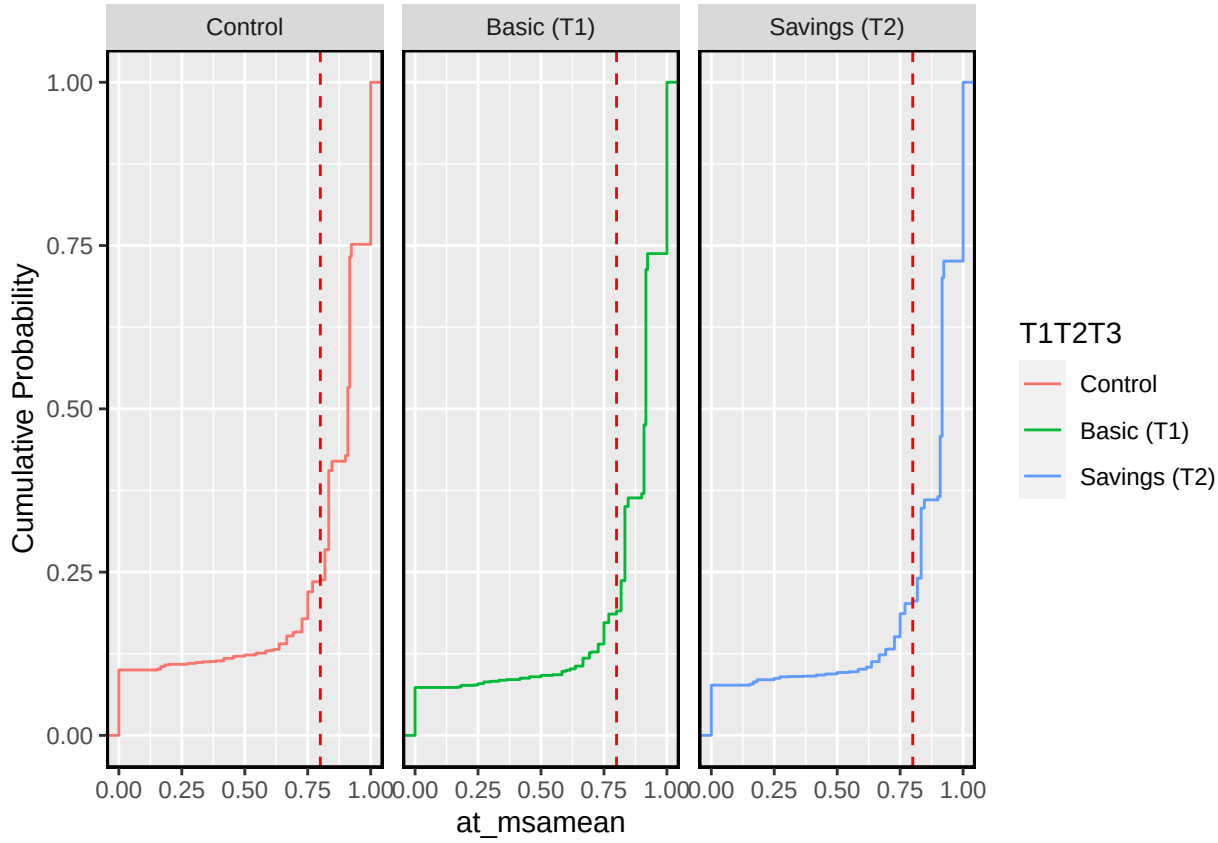
ggplot(data=filtered_barrera, aes(x=at_baseline, y=at_msamean, color=factor(T1T2T3))) + geom_smooth(method="lm",
se=FALSE)+ xlim(0.65, NA) + ylim(0.5, NA)

7 Some exploration

Part of the problem with this model is the outcome variable, which has a ceiling effect (can't go above 100%, and many people have high attendance, with target attendance also being high at 80%.)

It may be worthwhile investigating whether there is a significant difference in the proportion above the cut-off between samples.

```
# Create separate cumulative distribution plots of at_msamean for each level of T1T2T3

ggplot(filtered_barrera, aes(x = at_msamean)) +
  stat_ecdf(aes(color = T1T2T3)) +
  labs(x = "at_msamean", y = "Cumulative Probability") +
  facet_wrap(~T1T2T3, ncol = 3) +
  geom_vline(xintercept = 0.8, color = "red", linetype = "dashed") +
  theme(panel.border = element_rect(colour = "black", fill = NA, size = 1),
        panel.spacing = unit(0.5, "lines"))
```




```
# What proportion in each group falls at or above the target of 80% attendance?

cutoff <- 0.8  # set the cutoff value

filtered_barrera %>%
  group_by(T1T2T3) %>%
  summarize(prop_cutoff = sum(at_msamean >= cutoff) / n())
```

```
## # A tibble: 3 x 2
##   T1T2T3       prop_cutoff
##   <fct>              <dbl>
## 1 Control            0.765
## 2 Basic (T1)         0.814
## 3 Savings (T2)       0.798
```

Variability in the outcome variable is limited because most students attend at least 80% of the time. An alternative model specification may be to analyse differences in proportion of attendance (above / below cut-off).

Alternative way of approaching this: GLM for skewed data (e.g. log link and gamma function –> would need to fit this more carefully,) Binary variable: Whether or not student achieved 80% attendance

```
## Loading required package: AER
```

```
## 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
```

```
## Loading required package: Formula
```

```
## Loading required package: plm
```

```
##
## Attaching package: 'plm'
```

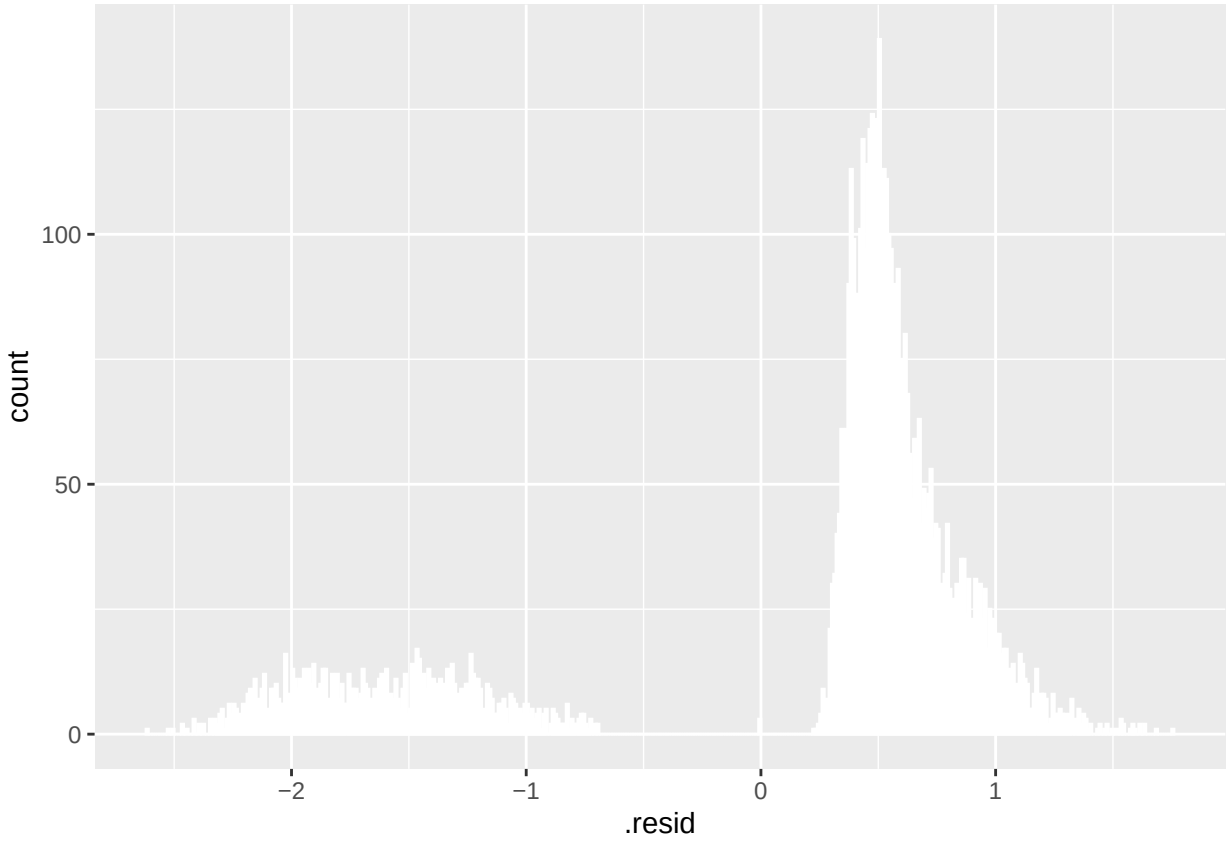
```
## The following objects are masked from 'package:dplyr':
##
##     between, lag, lead
```

```
##
##  When using this package, cite:
##
##  Justin Esarey and Andrew Menger (2019).
##  "Practical and Effective Approaches to Dealing with Clustered Data."
##  Political Science Research and Methods 7(3): 541-549.
##  URL: https://doi.org/10.1017/psrm.2017.42.
```

```
##
## Call:
## glm(formula = above_cutoff ~ T1_treat + T2_treat + f_teneviv +
##     s_utilities + s_durables + s_infraest_hh + s_age_sorteo +
##     s_age_sorteo2 + s_years_back + s_sexo + f_estcivil + s_single +
##     s_edadhead + s_yrshead + s_tpersona + s_num18 + f_estrato +
##     s_puntaje + s_ingtotal + f_grade + suba + s_over_age + factor(school_code),
##     family = binomial(), data = filtered_barrera)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6115   0.3473   0.4945   0.6623   1.7580
##
## Coefficients: (2 not defined because of singularities)
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)             1.805e+00  7.585e-01   2.379 0.017337 *
## T1_treat                3.364e-01  8.434e-02   3.988 6.65e-05 ***
## T2_treat                2.192e-01  8.424e-02   2.602 0.009264 **
## f_teneviv2              3.867e-01  1.503e-01   2.572 0.010104 *
## f_teneviv3             -4.368e-02  9.304e-02  -0.469 0.638749
## f_teneviv4              3.989e-01  1.255e-01   3.178 0.001484 **
## s_utilities             1.744e-02  4.227e-02   0.413 0.679931
## s_durables              8.622e-02  5.139e-02   1.678 0.093422 .
## s_infraest_hh           8.976e-03  2.311e-02   0.388 0.697661
## s_age_sorteo           -4.467e-02  3.740e-02  -1.194 0.232345
## s_age_sorteo2           1.296e-03  6.548e-04   1.980 0.047719 *
## s_years_back           -3.825e-02  4.121e-02  -0.928 0.353362
## s_sexo                 -7.022e-02  7.020e-02  -1.000 0.317216
## f_estcivilMarried       1.731e-01  5.545e-01   0.312 0.754886
## f_estcivilWidow(er)    -9.852e-01  1.088e+00  -0.905 0.365203
## f_estcivilDivorced      2.064e-01  6.826e-01   0.302 0.762365
## f_estcivilSingle       -1.058e+00  5.086e-01  -2.079 0.037572 *
## s_single               -7.570e-02  8.064e-02  -0.939 0.347903
## s_edadhead             -1.842e-04  4.201e-03  -0.044 0.965035
## s_yrshead              -2.774e-02  1.397e-02  -1.986 0.046988 *
## s_tpersona             -3.457e-02  3.503e-02  -0.987 0.323799
## s_num18                 3.593e-02  4.531e-02   0.793 0.427776
## f_estrato1             -6.434e-02  1.552e-01  -0.415 0.678382
## f_estrato2              8.660e-02  1.832e-01   0.473 0.636402
## s_puntaje              -2.136e-02  1.872e-02  -1.141 0.253982
## s_ingtotal              1.299e-05  1.793e-04   0.072 0.942246
## f_grade7                1.244e-01  9.512e-02   1.307 0.191071
## f_grade8                9.769e-02  1.008e-01   0.970 0.332260
## f_grade9               -1.588e-02  1.124e-01  -0.141 0.887622
## f_grade10                      NA         NA      NA       NA
## suba                           NA         NA      NA       NA
## s_over_age             -7.678e-01  1.115e-01  -6.888 5.64e-12 ***
## factor(school_code)56   1.913e+00  2.981e-01   6.417 1.39e-10 ***
## factor(school_code)57   1.983e+00  2.518e-01   7.876 3.39e-15 ***
## factor(school_code)61   1.258e+00  3.262e-01   3.858 0.000114 ***
## factor(school_code)78   6.700e-01  2.309e-01   2.902 0.003705 **
## factor(school_code)79   1.604e+00  1.993e-01   8.048 8.44e-16 ***
## factor(school_code)80   1.435e+00  2.089e-01   6.869 6.47e-12 ***
## factor(school_code)86   2.348e+00  3.382e-01   6.942 3.88e-12 ***
```

```
## factor(school_code)87   1.650e+00  2.934e-01   5.624 1.87e-08 ***
## factor(school_code)88   7.537e-01  2.673e-01   2.820 0.004800 **
## factor(school_code)89   1.322e+00  2.955e-01   4.474 7.66e-06 ***
## factor(school_code)90   1.374e+00  2.859e-01   4.805 1.55e-06 ***
## factor(school_code)97   1.273e+00  2.200e-01   5.786 7.19e-09 ***
## factor(school_code)100 -1.262e+01  3.247e+02  -0.039 0.969001
## factor(school_code)105  1.340e+00  2.390e-01   5.608 2.04e-08 ***
## factor(school_code)113  1.031e+00  3.360e-01   3.070 0.002144 **
## factor(school_code)114  8.393e-03  1.790e-01   0.047 0.962598
## factor(school_code)117  1.947e+00  2.207e-01   8.819  < 2e-16 ***
## factor(school_code)122 -4.901e-01  2.022e-01  -2.424 0.015350 *
## factor(school_code)125  1.297e+00  2.330e-01   5.567 2.60e-08 ***
## factor(school_code)126  1.154e+00  2.298e-01   5.019 5.19e-07 ***
## factor(school_code)135  2.010e+00  3.708e-01   5.419 5.99e-08 ***
## factor(school_code)149  1.733e+00  2.756e-01   6.288 3.21e-10 ***
## factor(school_code)153  2.295e+00  4.496e-01   5.104 3.33e-07 ***
## factor(school_code)166 -1.322e+01  3.247e+02  -0.041 0.967534
## factor(school_code)172 -1.309e+01  3.247e+02  -0.040 0.967847
## factor(school_code)261  3.475e-01  3.068e-01   1.133 0.257274
## factor(school_code)262  7.817e-01  1.970e-01   3.968 7.25e-05 ***
## factor(school_code)276  1.479e+00  2.707e-01   5.463 4.69e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5940.2  on 5798  degrees of freedom
## Residual deviance: 5235.7  on 5741  degrees of freedom
## AIC: 5351.7
##
## Number of Fisher Scoring iterations: 11

##
## z test of coefficients:
##
##                           Estimate  Std. Error  z value  Pr(>|z|)
## (Intercept)             1.8049e+00  1.1089e+00   1.6277 0.1035987
## T1_treat                3.3638e-01  7.1491e-02   4.7053 2.535e-06 ***
## T2_treat                2.1920e-01  9.2287e-02   2.3752 0.0175408 *
## f_teneviv2              3.8672e-01  2.5275e-01   1.5301 0.1260029
## f_teneviv3             -4.3675e-02  1.0227e-01  -0.4271 0.6693408
## f_teneviv4              3.9890e-01  1.1584e-01   3.4435 0.0005741 ***
## s_utilities             1.7441e-02  4.8517e-02   0.3595 0.7192412
## s_durables              8.6218e-02  3.9954e-02   2.1579 0.0309327 *
## s_infraest_hh           8.9763e-03  2.1486e-02   0.4178 0.6761054
## s_age_sorteo           -4.4668e-02  5.7011e-02  -0.7835 0.4333344
## s_age_sorteo2           1.2964e-03  9.7280e-04   1.3327 0.1826337
## s_years_back           -3.8246e-02  4.5090e-02  -0.8482 0.3963154
## s_sexo                 -7.0218e-02  6.4640e-02  -1.0863 0.2773506
## f_estcivilMarried       1.7312e-01  5.1762e-01   0.3345 0.7380311
## f_estcivilWidow(er)    -9.8517e-01  1.0079e+00  -0.9774 0.3283469
## f_estcivilDivorced      2.0641e-01  5.6248e-01   0.3670 0.7136454
## f_estcivilSingle       -1.0576e+00  6.4264e-01  -1.6458 0.0998103 .
## s_single               -7.5698e-02  6.4143e-02  -1.1802 0.2379375
```

```
## s_edadhead             -1.8415e-04  4.8891e-03  -0.0377 0.9699539
## s_yrshead              -2.7744e-02  1.5790e-02  -1.7571 0.0789034 .
## s_tpersona             -3.4565e-02  2.9569e-02  -1.1690 0.2424179
## s_num18                 3.5933e-02  4.7117e-02   0.7626 0.4456831
## f_estrato1             -6.4340e-02  1.3934e-01  -0.4617 0.6442621
## f_estrato2              8.6601e-02  1.5579e-01   0.5559 0.5783004
## s_puntaje              -2.1357e-02  1.9242e-02  -1.1099 0.2670243
## s_ingtotal              1.2991e-05  1.5546e-04   0.0836 0.9334034
## f_grade7                1.2437e-01  9.0883e-02   1.3684 0.1711767
## f_grade8                9.7689e-02  1.6256e-01   0.6009 0.5478774
## f_grade9               -1.5879e-02  1.6674e-01  -0.0952 0.9241272
## s_over_age             -7.6778e-01  1.1712e-01  -6.5554 5.549e-11 ***
## factor(school_code)56   1.9130e+00  2.8814e-02  66.3912 < 2.2e-16 ***
## factor(school_code)57   1.9833e+00  3.5846e-02  55.3298 < 2.2e-16 ***
## factor(school_code)61   1.2584e+00  5.7256e-02  21.9793 < 2.2e-16 ***
## factor(school_code)78   6.7004e-01  3.9355e-02  17.0256 < 2.2e-16 ***
## factor(school_code)79   1.6043e+00  3.2763e-02  48.9665 < 2.2e-16 ***
## factor(school_code)80   1.4349e+00  2.9572e-02  48.5201 < 2.2e-16 ***
## factor(school_code)86   2.3478e+00  3.3772e-02  69.5192 < 2.2e-16 ***
## factor(school_code)87   1.6501e+00  4.6719e-02  35.3186 < 2.2e-16 ***
## factor(school_code)88   7.5370e-01  3.9832e-02  18.9219 < 2.2e-16 ***
## factor(school_code)89   1.3223e+00  7.0130e-02  18.8551 < 2.2e-16 ***
## factor(school_code)90   1.3737e+00  7.9874e-02  17.1991 < 2.2e-16 ***
## factor(school_code)97   1.2732e+00  4.0527e-02  31.4165 < 2.2e-16 ***
## factor(school_code)100 -1.2620e+01  1.0369e+00 -12.1709 < 2.2e-16 ***
## factor(school_code)105  1.3402e+00  4.3415e-02  30.8702 < 2.2e-16 ***
## factor(school_code)113  1.0312e+00  5.3005e-02  19.4548 < 2.2e-16 ***
## factor(school_code)114  8.3933e-03  3.2363e-02   0.2593 0.7953694
## factor(school_code)117  1.9467e+00  3.0113e-02  64.6449 < 2.2e-16 ***
## factor(school_code)122 -4.9008e-01  7.6791e-02  -6.3821 1.747e-10 ***
## factor(school_code)125  1.2968e+00  3.3757e-02  38.4170 < 2.2e-16 ***
## factor(school_code)126  1.1535e+00  2.6981e-02  42.7532 < 2.2e-16 ***
## factor(school_code)135  2.0096e+00  2.7194e-02  73.8985 < 2.2e-16 ***
## factor(school_code)149  1.7327e+00  4.9423e-02  35.0589 < 2.2e-16 ***
## factor(school_code)153  2.2946e+00  3.8730e-02  59.2477 < 2.2e-16 ***
## factor(school_code)166 -1.3217e+01  1.0401e+00 -12.7083 < 2.2e-16 ***
## factor(school_code)172 -1.3090e+01  1.0393e+00 -12.5955 < 2.2e-16 ***
## factor(school_code)261  3.4755e-01  9.4199e-02   3.6895 0.0002247 ***
## factor(school_code)262  7.8173e-01  3.6269e-02  21.5538 < 2.2e-16 ***
## factor(school_code)276  1.4789e+00  1.0492e-01  14.0950 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



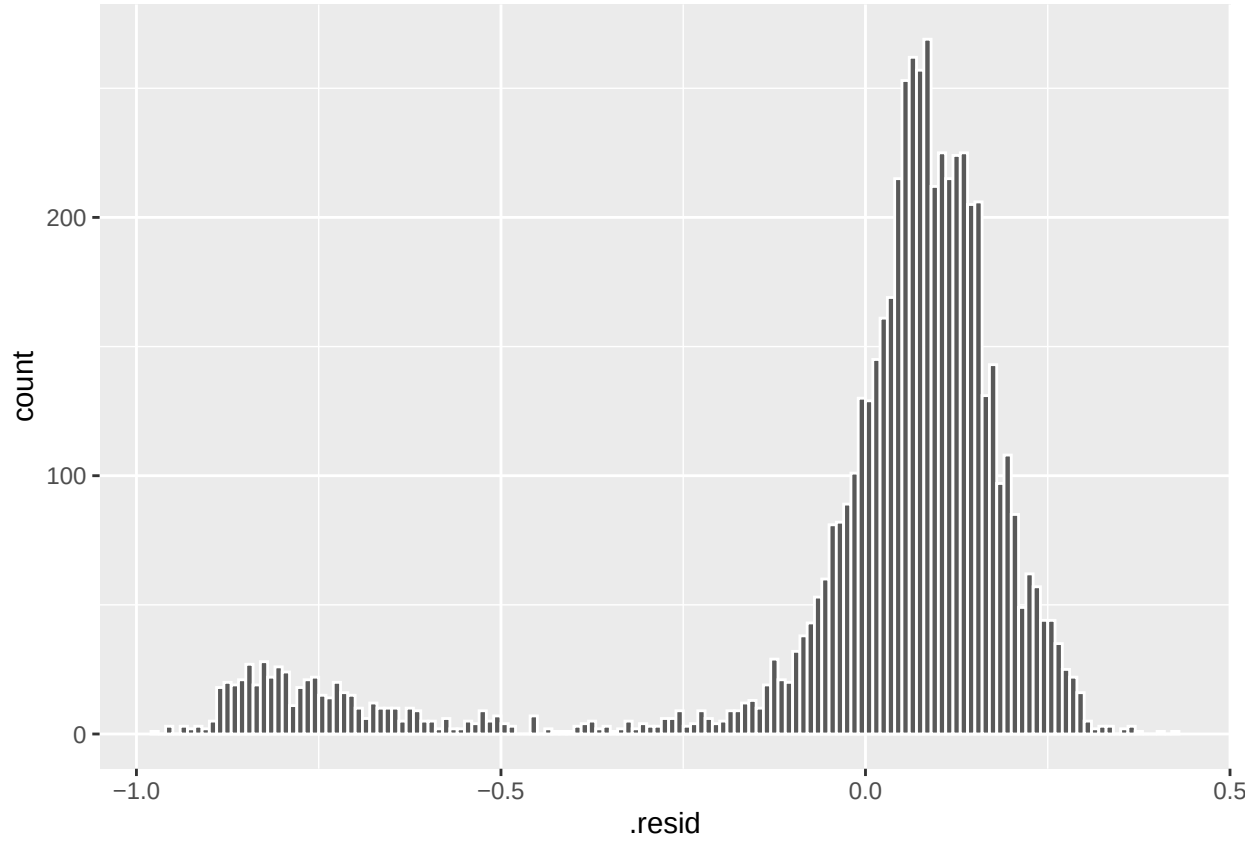

```
##
## Call:
## glm(formula = at_msamean ~ T1_treat + T2_treat + f_teneviv +
##     s_utilities + s_durables + s_infraest_hh + s_age_sorteo +
##     s_age_sorteo2 + s_years_back + s_sexo + f_estcivil + s_single +
##     s_edadhead + s_yrshead + s_tpersona + s_num18 + f_estrato +
##     s_puntaje + s_ingtotal + f_grade + suba + s_over_age + factor(school_code),
##     family = gaussian(), data = filtered_barrera)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.97310  -0.00327   0.07251   0.13531   0.42150
##
## Coefficients: (2 not defined because of singularities)
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             8.795e-01  7.512e-02  11.708  < 2e-16 ***
## T1_treat                3.157e-02  8.368e-03   3.773 0.000163 ***
## T2_treat                2.704e-02  8.497e-03   3.183 0.001466 **
## f_teneviv2              4.199e-02  1.427e-02   2.943 0.003259 **
## f_teneviv3             -1.028e-03  9.456e-03  -0.109 0.913456
## f_teneviv4              3.391e-02  1.180e-02   2.875 0.004056 **
## s_utilities            -1.519e-03  4.286e-03  -0.354 0.723009
## s_durables              6.446e-03  5.135e-03   1.255 0.209415
## s_infraest_hh          -1.342e-03  2.363e-03  -0.568 0.570224
## s_age_sorteo           -6.359e-03  3.545e-03  -1.794 0.072884 .
## s_age_sorteo2           1.816e-04  5.656e-05   3.212 0.001327 **
```

```
## s_years_back           -7.134e-03  4.023e-03  -1.773 0.076227 .
## s_sexo                 -7.171e-03  6.988e-03  -1.026 0.304861
## f_estcivilMarried       2.107e-03  5.026e-02   0.042 0.966572
## f_estcivilWidow(er)    -2.556e-01  1.096e-01  -2.332 0.019719 *
## f_estcivilDivorced      2.996e-02  6.289e-02   0.476 0.633803
## f_estcivilSingle       -1.715e-01  4.841e-02  -3.542 0.000401 ***
## s_single               -5.252e-03  8.127e-03  -0.646 0.518181
## s_edadhead              1.494e-04  4.214e-04   0.355 0.722895
## s_yrshead              -1.890e-03  1.410e-03  -1.340 0.180170
## s_tpersona             -1.625e-03  3.567e-03  -0.455 0.648790
## s_num18                 6.351e-03  4.599e-03   1.381 0.167310
## f_estrato1              5.453e-03  1.590e-02   0.343 0.731680
## f_estrato2              3.836e-02  1.844e-02   2.080 0.037573 *
## s_puntaje              -3.901e-03  1.881e-03  -2.074 0.038109 *
## s_ingtotal             -1.031e-05  1.807e-05  -0.571 0.568154
## f_grade7                1.919e-02  9.452e-03   2.030 0.042358 *
## f_grade8                1.985e-02  1.002e-02   1.981 0.047621 *
## f_grade9               -1.820e-05  1.110e-02  -0.002 0.998692
## f_grade10                      NA         NA      NA       NA
## suba                           NA         NA      NA       NA
## s_over_age             -8.525e-02  1.169e-02  -7.293 3.44e-13 ***
## factor(school_code)56   1.910e-01  2.745e-02   6.961 3.76e-12 ***
## factor(school_code)57   2.621e-01  2.442e-02  10.729  < 2e-16 ***
## factor(school_code)61   1.720e-01  3.479e-02   4.945 7.82e-07 ***
## factor(school_code)78   1.095e-01  2.782e-02   3.937 8.36e-05 ***
## factor(school_code)79   1.968e-01  2.217e-02   8.877  < 2e-16 ***
## factor(school_code)80   2.074e-01  2.328e-02   8.912  < 2e-16 ***
## factor(school_code)86   2.715e-01  2.770e-02   9.801  < 2e-16 ***
## factor(school_code)87   2.197e-01  2.914e-02   7.541 5.38e-14 ***
## factor(school_code)88   1.116e-01  3.087e-02   3.615 0.000303 ***
## factor(school_code)89   1.668e-01  3.144e-02   5.305 1.17e-07 ***
## factor(school_code)90   1.856e-01  2.978e-02   6.232 4.95e-10 ***
## factor(school_code)97   1.426e-01  2.462e-02   5.793 7.27e-09 ***
## factor(school_code)100 -5.950e-01  2.645e-01  -2.249 0.024550 *
## factor(school_code)105  2.140e-01  2.599e-02   8.235  < 2e-16 ***
## factor(school_code)113  1.490e-01  3.593e-02   4.146 3.43e-05 ***
## factor(school_code)114  1.086e-01  2.247e-02   4.832 1.39e-06 ***
## factor(school_code)117  2.306e-01  2.266e-02  10.176  < 2e-16 ***
## factor(school_code)122  8.054e-02  2.505e-02   3.215 0.001312 **
## factor(school_code)125  1.961e-01  2.578e-02   7.606 3.28e-14 ***
## factor(school_code)126  1.631e-01  2.611e-02   6.248 4.46e-10 ***
## factor(school_code)135  2.707e-01  3.262e-02   8.298  < 2e-16 ***
## factor(school_code)149  1.997e-01  2.744e-02   7.278 3.84e-13 ***
## factor(school_code)153  2.608e-01  3.266e-02   7.984 1.69e-15 ***
## factor(school_code)166 -6.670e-01  2.642e-01  -2.525 0.011613 *
## factor(school_code)172 -6.602e-01  2.641e-01  -2.499 0.012476 *
## factor(school_code)261  5.837e-02  3.692e-02   1.581 0.113945
## factor(school_code)262  1.394e-01  2.367e-02   5.888 4.12e-09 ***
## factor(school_code)276  1.936e-01  2.847e-02   6.799 1.16e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.06924162)
##
```

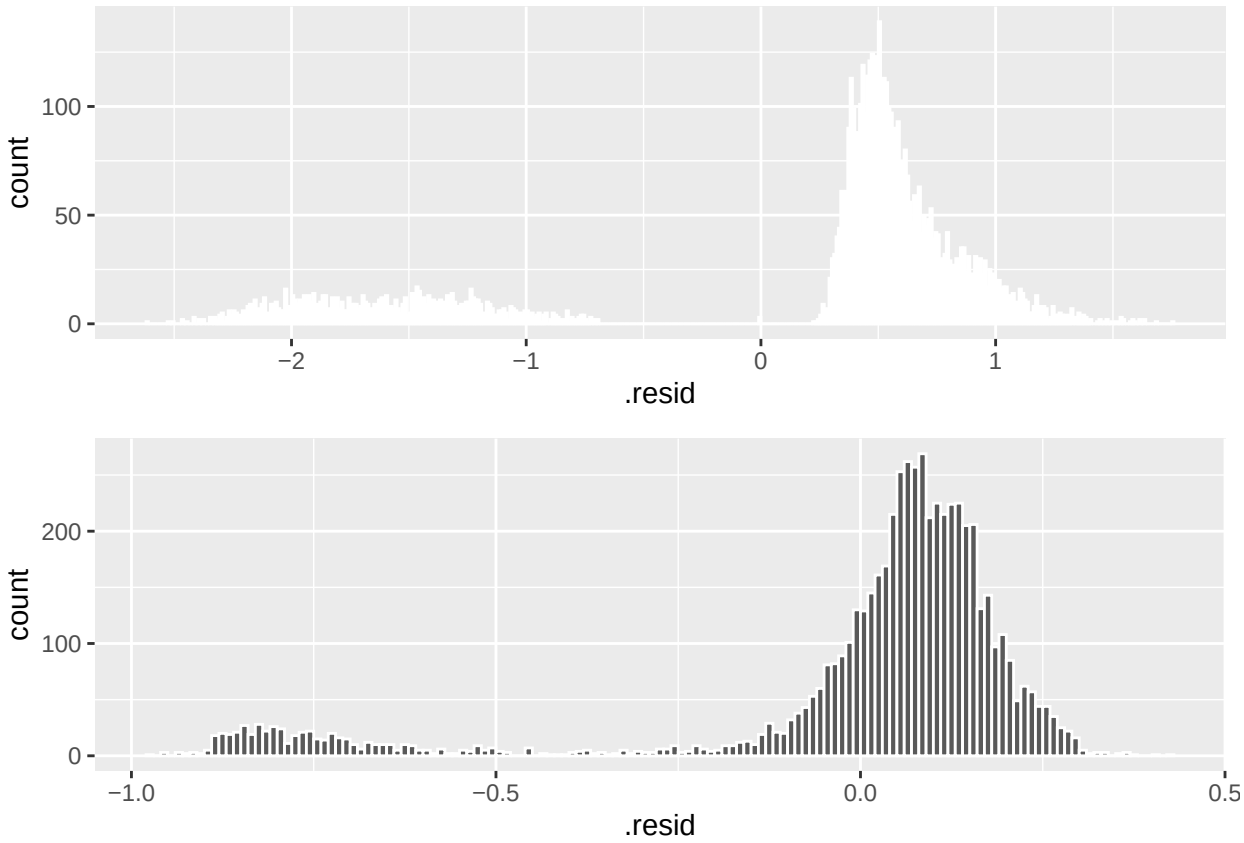
```
##     Null deviance: 436.59  on 5798  degrees of freedom
## Residual deviance: 397.52  on 5741  degrees of freedom
## AIC: 1032.3
##
## Number of Fisher Scoring iterations: 2
```

```
##
## z test of coefficients:
##
##                            Estimate  Std. Error  z value  Pr(>|z|)
## (Intercept)             8.7954e-01  1.0878e-01   8.0855 6.190e-16 ***
## T1_treat                3.1572e-02  7.3851e-03   4.2751 1.910e-05 ***
## T2_treat                2.7044e-02  7.0955e-03   3.8114 0.0001382 ***
## f_teneviv2              4.1993e-02  1.7916e-02   2.3438 0.0190877 *
## f_teneviv3             -1.0277e-03  1.3568e-02  -0.0757 0.9396194
## f_teneviv4              3.3910e-02  1.2750e-02   2.6595 0.0078248 **
## s_utilities            -1.5191e-03  5.5335e-03  -0.2745 0.7836735
## s_durables              6.4455e-03  4.6332e-03   1.3912 0.1641794
## s_infraest_hh          -1.3419e-03  2.5402e-03  -0.5283 0.5973189
## s_age_sorteo           -6.3586e-03  4.2982e-03  -1.4794 0.1390467
## s_age_sorteo2           1.8165e-04  7.6326e-05   2.3799 0.0173196 *
## s_years_back           -7.1340e-03  4.6221e-03  -1.5434 0.1227216
## s_sexo                 -7.1712e-03  6.0288e-03  -1.1895 0.2342487
## f_estcivilMarried       2.1065e-03  4.1669e-02   0.0506 0.9596811
## f_estcivilWidow(er)    -2.5560e-01  1.2362e-01  -2.0676 0.0386741 *
## f_estcivilDivorced      2.9960e-02  3.5565e-02   0.8424 0.3995757
## f_estcivilSingle       -1.7146e-01  6.8887e-02  -2.4890 0.0128112 *
## s_single               -5.2517e-03  6.8330e-03  -0.7686 0.4421474
## s_edadhead              1.4943e-04  4.2965e-04   0.3478 0.7279855
## s_yrshead              -1.8896e-03  1.4531e-03  -1.3003 0.1934866
## s_tpersona             -1.6246e-03  3.1933e-03  -0.5088 0.6109274
## s_num18                 6.3514e-03  5.1794e-03   1.2263 0.2200929
## f_estrato1              5.4526e-03  1.6976e-02   0.3212 0.7480699
## f_estrato2              3.8364e-02  1.6167e-02   2.3730 0.0176445 *
## s_puntaje              -3.9014e-03  2.1026e-03  -1.8555 0.0635233 .
## s_ingtotal             -1.0315e-05  1.5101e-05  -0.6831 0.4945689
## f_grade7                1.9192e-02  1.0313e-02   1.8610 0.0627512 .
## f_grade8                1.9848e-02  6.6118e-03   3.0019 0.0026833 **
## f_grade9               -1.8205e-05  1.1316e-02  -0.0016 0.9987164
## s_over_age             -8.5245e-02  1.5108e-02  -5.6422 1.679e-08 ***
## factor(school_code)56   1.9105e-01  3.3398e-03  57.2026 < 2.2e-16 ***
## factor(school_code)57   2.6207e-01  3.1067e-03  84.3542 < 2.2e-16 ***
## factor(school_code)61   1.7204e-01  5.6048e-03  30.6945 < 2.2e-16 ***
## factor(school_code)78   1.0953e-01  3.5188e-03  31.1263 < 2.2e-16 ***
## factor(school_code)79   1.9683e-01  2.8198e-03  69.8021 < 2.2e-16 ***
## factor(school_code)80   2.0742e-01  2.9762e-03  69.6935 < 2.2e-16 ***
## factor(school_code)86   2.7146e-01  2.8557e-03  95.0604 < 2.2e-16 ***
## factor(school_code)87   2.1973e-01  3.9279e-03  55.9405 < 2.2e-16 ***
## factor(school_code)88   1.1161e-01  3.8900e-03  28.6905 < 2.2e-16 ***
## factor(school_code)89   1.6676e-01  5.9783e-03  27.8942 < 2.2e-16 ***
## factor(school_code)90   1.8558e-01  4.3146e-03  43.0116 < 2.2e-16 ***
## factor(school_code)97   1.4265e-01  2.6114e-03  54.6254 < 2.2e-16 ***
## factor(school_code)100 -5.9495e-01  1.7061e-02 -34.8714 < 2.2e-16 ***
## factor(school_code)105  2.1398e-01  4.7869e-03  44.7025 < 2.2e-16 ***
```

```
## factor(school_code)113  1.4898e-01  5.1185e-03  29.1073 < 2.2e-16 ***
## factor(school_code)114  1.0860e-01  3.8378e-03  28.2978 < 2.2e-16 ***
## factor(school_code)117  2.3056e-01  3.7657e-03  61.2267 < 2.2e-16 ***
## factor(school_code)122  8.0537e-02  7.5196e-03  10.7103 < 2.2e-16 ***
## factor(school_code)125  1.9612e-01  3.6754e-03  53.3605 < 2.2e-16 ***
## factor(school_code)126  1.6311e-01  2.8207e-03  57.8247 < 2.2e-16 ***
## factor(school_code)135  2.7070e-01  2.8411e-03  95.2822 < 2.2e-16 ***
## factor(school_code)149  1.9972e-01  5.6547e-03  35.3202 < 2.2e-16 ***
## factor(school_code)153  2.6079e-01  3.7524e-03  69.4993 < 2.2e-16 ***
## factor(school_code)166 -6.6696e-01  1.6192e-02 -41.1914 < 2.2e-16 ***
## factor(school_code)172 -6.6015e-01  1.9061e-02 -34.6328 < 2.2e-16 ***
## factor(school_code)261  5.8366e-02  7.5393e-03   7.7415 9.822e-15 ***
## factor(school_code)262  1.3939e-01  2.5691e-03  54.2546 < 2.2e-16 ***
## factor(school_code)276  1.9356e-01  1.0075e-02  19.2112 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```




```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine
```



GLM with binary outcome variable performs much better on AIC. But I'm not sure the residual plots help us very much as it's not clustered SE!

8 Conclusion

- Replication of a research paper: How do results compare to results of research paper?

Models all bad, R-sq low.

Compare coefficients, standard errors Why might they be different? * Software: Possible to get STATA standard errors in R

Source for clustered standard errors in R: https://evalf21.classes.andrewheiss.com/example/standard-errors/