Test-taking with Maimonides

Estimating the Effect of Class Size on 4th Grade Achievement
using an Instrumental Variable

Dipika Gawande | S&DS 530 | 16 December 2021

1. Introduction

1a. Background and Motivation

A big policy question in education is the effect of class size on student outcomes. It is often argued that smaller class size is better, both for teachers and students (Mueller and Walden 1988) as well as for policy-makers (Robinson 1990) who find that class size is easier to change than other policy targets (such as lack of resources). However, empirical evidence for this belief is very weak. Angrist and Lavy, in their seminal 1999 paper, argue that research has not consistently shown that class size affects student achievement. The reason for the lack of evidence is a special case of omitted variable bias, called the "endogeneity problem."

Endogeneity of school quality variables is a long-standing issue in education research (Nascimento 2009). The problem is that school quality is determined by many factors that are unobservable. Interactions between parents, teachers, schools, and local government affect students' outcomes in ways that are difficult to measure. These "school interactions" manifest in unobserved "variables" like wealth and advantage, which are correlated **both** with class size (school quality in general) *and* student outcomes. For example, wealthy parents can pay for tutors and can also choose to send their kids to schools with smaller classes. The source of variation in class size (parental wealth, among other factors) is therefore not independent of the source of variation in the student's outcomes (also parental wealth, among other factors).

This "endogeneity" problem is resolved by a technique called Instrumental Variable (IV) estimation (Steele, Vignoles, and Jenkins 2007). The idea is to find a variable that is highly correlated with Class Size, but whose source of variation is independent of student outcomes. It is very difficult to find such IVs in practice. Angrist and Lavy (1999) find such a variable in an unusual rule that governs the division of school enrollment into class size in Israeli public schools today. The "Rule of 40," interpreted from the Talmud by 12th century Rabbinic scholar Maimonides, says the size of any class should be no more than 40 students (Hyamson 1937). Maimonides' rule has been used by the Israeli Education Ministry to determine the division of enrollment cohorts into classes in Jewish public schools since 1969. Angrist and Lavy use this rule to to estimate the effects of class size on achievement of 4th and 5th graders.

1b. Data source

The data set used in this project is the administrative class-level replication data for "Using Maimonides Rule to Estimate the Effect of Class Size on Scholastic Achievement" (Angrist and Lavy 1999), downloaded from Harvard DataVerse.[1] Class average test score data in the paper's data set comes from administrative school and class data on average math and reading scores for 4th and 5th graders in 1991. The average test score data comes from a national testing program in Israeli public elementary schools in 1991 and 1992. The class and school variables (such as school enrollment size, class size, percent disadvantaged students, gender breakdown in each grade) come from annual reports from the Israeli Ministry of Education and Central

[1] https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/XRSUJU.

Bureau of Statistics. The data coverage is the population of **all** Jewish public elementary schools in Israel covered by the Central Bureau of Statistics [1991, 1993] Censuses of Schools.

1c. Project Uniqueness

The approach in my analysis of Angrist and Lavy's data is unique in the following aspects: I focus on the data for only 4th graders throughout this paper. Second, I test hypotheses regarding the relevance of economic disadvantage to 4th grade Class Average Test Score. Third, I conduct the Instrumental Variable (IV) estimation by writing my own R code for 2-stage linear regression in separate parts. Fourth, I test for interaction effects between Class Size and Enrollment.

2. Visualizing the Data

Our main predictor/covariate of interest is `class_size`. The average class has 30 students. The `class_size` variable has a standard deviation of 6.4 students. 10% of classes have less than 22 students, and 10% of classes have more than 38 students. The response variables of interest are `class_avg_math` and `class_avg_verbal`. Before visualizing Test Scores, we drop 4 observations for which these variables have missing values. The mean of `class_avg_verbal` is 72.5%, with a standard deviation of 8.0 percentage points. The mean of `class_avg_math` is 68.9%, with a standard deviation of 8.8 percentage points. *See Appendix for figure code.*

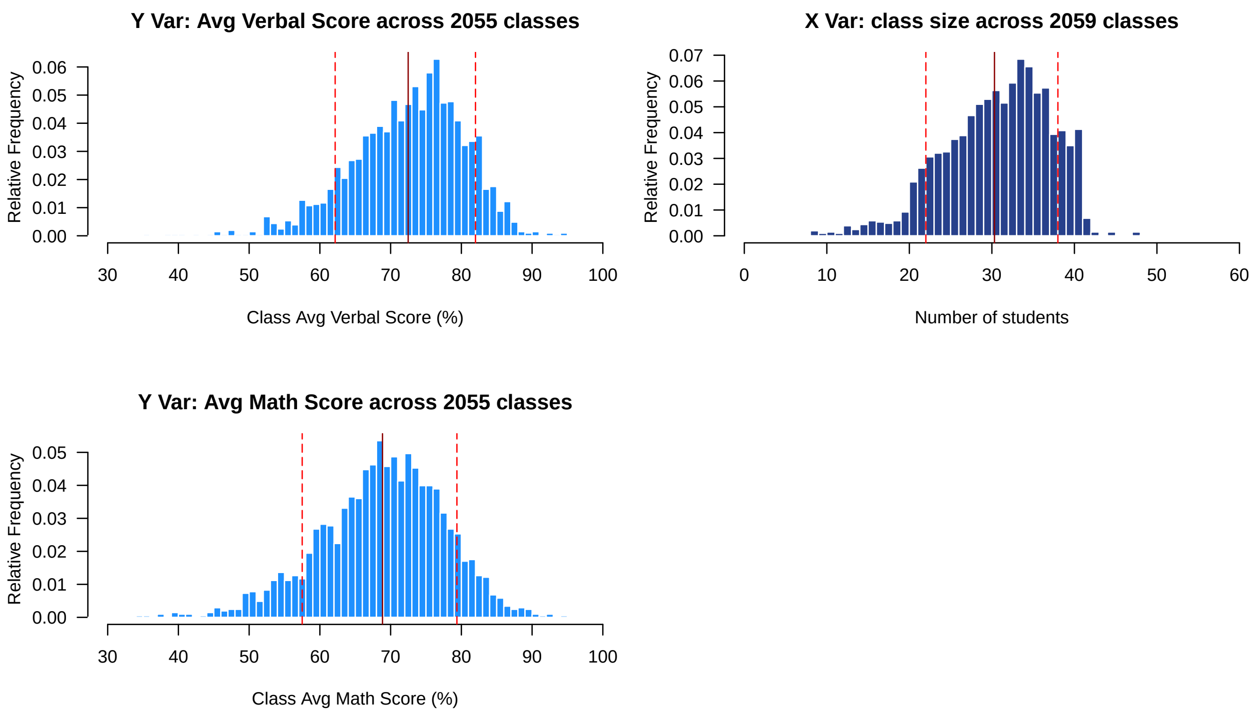


Figure 1: *Left: Distribution of Class Average Verbal Score and Class Average Math Score (response variables) from 2055 non-missing 4th grade Jewish public school classes. Right: Distribution of Class Size (predictor variable) across the population of 2059 4th grade Jewish public school classes. Lowest and highest decile denoted by red dashed lines. Mean denoted by maroon solid line.*

From the histograms and Q-Q plots, the two response variables (`class_avg_math` and `class_avg_verbal`) look to be approximately normally distributed. The distribution of `class_avg_verbal` is more left-skewed, deviating slightly from the 45 degree line of the Q-Q plot. However, we will proceed with ANOVA and linear regression to estimate the effect of Class Size on Test Scores. *See Appendix for figure code.*

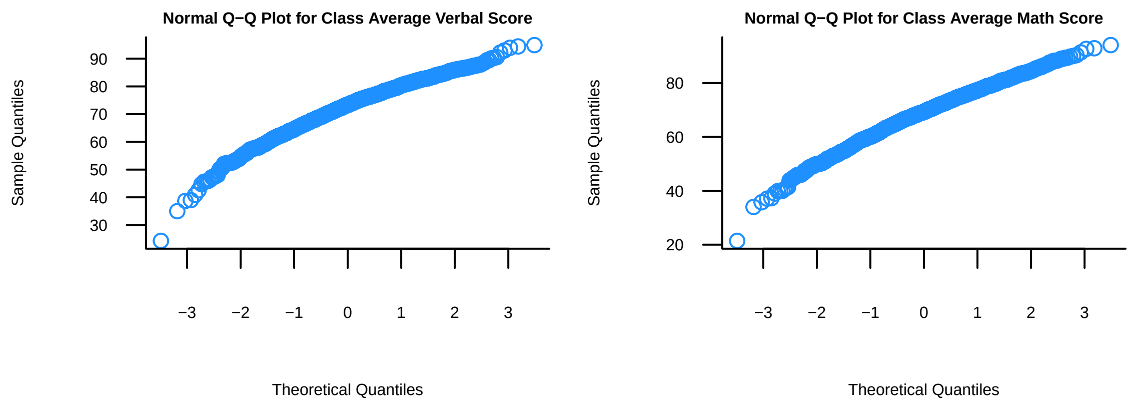


Figure 2: *Normal Q-Q Plots of the response variables Class Average Verbal Score and Class Average Math Score. Class Average Math is nearly normally distributed. Class Average Verbal is slightly left-skewed.*

3. Some pre-analysis

3a. A First Look at the Effect of Class Size on Test Scores

To see whether class size has an effect on 4th graders' test scores, we conduct ANOVA on a linear model. Model 1 includes only `class_size`, while Model 2 includes `class_size` and `pct_disadvantaged`.

```
fit1_verbal <- lm(data = g4, class_avg_verbal ~ class_size)
summary(fit1_verbal)$coefficients[ ,1:4]
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 68.3798393 0.85015110 80.432571 0.000000e+00
## class_size   0.1352977 0.02743036  4.932406 8.775126e-07
```

```
anova(fit1_verbal)
```

```
## Analysis of Variance Table
##
## Response: class_avg_verbal
##              Df Sum Sq Mean Sq F value    Pr(>F)
## class_size    1   1534 1534.20  24.329 8.775e-07 ***
## Residuals  2053 129465   63.06
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
fit2_verbal <- lm(data = g4, class_avg_verbal ~ class_size + pct_disadvantaged + enrollment)
summary(fit2_verbal)$coefficients[ ,1:4]
```

```
##                      Estimate  Std. Error     t value      Pr(>|t|)
## (Intercept)       78.78026097 0.794296053  99.1824908  0.000000e+00
## class_size        -0.04176875 0.028264256  -1.4777940  1.396165e-01
## pct_disadvantaged -0.34073046 0.011674057 -29.1869801 6.327194e-157
## enrollment        -0.00403053 0.004875752  -0.8266478  4.085329e-01
```

```r
anova(fit2_verbal)
```

```
## Analysis of Variance Table
##
## Response: class_avg_verbal
##                      Df Sum Sq Mean Sq  F value    Pr(>F)
## class_size            1   1534    1534  34.7792 4.308e-09 ***
## pct_disadvantaged     1  38960   38960 883.2040 < 2.2e-16 ***
## enrollment            1     30      30   0.6833    0.4085
## Residuals          2051  90475      44
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Two conclusions are apparent from the ANOVA and regression results:

- **Good news:** Adding more variables to the model improves the model's fit and reduces the Mean Squared Error of the residuals. Including `pct_disadvantaged` in the model along with our covariate of interest `class_size` allows us to explain more of the variation in test scores.
- **Bad news:** The coefficient on `class_size` disappears (reduces from 0.135 to -0.042) when we include `pct_disadvantaged`. The effect of what we thought was the most important explanator of 4th-grade achievement has disappeared. Are the data telling us that Class Size has no effect on test scores?

3b. The Endogeneity Problem

The `class_size` variable is a victim of omitted variable bias. This is because the variation in `class_size` is actually caused by important unobserved or unmeasurable factors (child's socioeconomic background, wealth of the school district, parental involvement and influence over school choice), which we cannot account for in our model. These omitted variables are responsible for the variation in **both** class size and test scores! Not including them leads to bias in our estimate of the effect of class size on test scores.

3c. Effect of Percent Disadvantaged Students on Test Scores

Parametric Hypothesis Test: Welch's t-Test for the Difference of Two Means

From our discussion of the endogeneity problem, a natural question arises: do test scores (`class_avg_verbal` and `class_avg_math`) differ significantly between disadvantaged vs non-disadvantaged 4th grade classes? Since our data are approximately normally distributed, we will use a t-test for the difference of two means. We'll do this test twice, once for `class_avg_verbal` and once for `class_avg_math`.

The `pct_disadvantaged` variable in the data set is a function of pupils' fathers' education and continent of birth, and family size" (Angrist and Lavy 1999). It reports the percentage of students in the class who come from what is defined to be a disadvantaged background.

```r
c(mean(g4$pct_disadvantaged), quantile(g4$pct_disadvantaged, c(.50, .75, .90, .95)))
```

```
##                50%      75%      90%      95%
## 13.84313  9.00000 19.00000 35.00000 43.00000
```

We will define a "disadvantaged class" as one in which `pct_disadvantaged` is above the 75th percentile. **These are classes in which at least 19% of the students are from disadvantaged backgrounds**. We define a "non-disadvantaged class" as having less than 19% (the 75th percentile of) disadvantaged students. Creating the two groups this way yields sample sizes of: 523 "disadvantaged" classes and 1532 "non-disadvantaged" classes.

Our null hypothesis is: *There is no difference in the mean Class Avg Verbal (Math) Score between disadvantaged and non-disadvantaged classes.* Our alternate hypothesis is: *There is a difference between the mean Class Avg Verbal (Math) Score in disadvantaged classes and that of non-disadvantaged classes.*

- $H_0 : \mu_{disadvantaged} = \mu_{non-disadvantaged}$
- $H_0 : \mu_{disadvantaged} \neq \mu_{non-disadvantaged}$

Where μ is the mean of the Class Average Verbal or Math Score. *See appendix for full output of t-Test.*

```
## Create the "disadvantaged" dummy variable
g4_disad <- g4 %>%
  filter(is.na(class_avg_math) == 0, is.na(class_avg_verbal) == 0) %>%
  mutate(disad = (pct_disadvantaged >= quantile(pct_disadvantaged, .75))) %>%
  group_by(disad)

## Make vectors of the two groups
disad_verb <- g4_disad[g4_disad$disad == TRUE, ]$class_avg_verbal
disad_math <- g4_disad[g4_disad$disad == TRUE, ]$class_avg_math
nondisad_verb <- g4_disad[g4_disad$disad == FALSE, ]$class_avg_verbal
nondisad_math <- g4_disad[g4_disad$disad == FALSE, ]$class_avg_math

## Welch's T-Test for difference in mean Class Average Verbal Score
ttest_verb <- t.test(disad_verb, nondisad_verb)
c(format(ttest_verb$statistic, scientific = FALSE), signif(ttest_verb$p.value, 4))
```

```
##            t
## "-22.20026" "1.036e-84"
```

```
## Welch's T-Test for difference in mean Class Average Math Score
ttest_math <- t.test(disad_math, nondisad_math)
c(format(ttest_math$statistic, scientific = FALSE), signif(ttest_math$p.value, 4))
```

```
##            t
## "-18.45343" "1.475e-63"
```

The p-value is near-zero for both tests so we reject H_0 in favor of H_A. Mean Class Average Verbal and Math Scores differ significantly between disadvantaged and non-disadvantaged classes. Therefore, `pct_disadvantaged` is an important control variable in the analysis.

4. Data Preparation: the Instrumental Variable

4a. Expected Class Size as per Maimonides' Rule

To solve the endogeneity problem we will use the technique of creating an **instrumental variable** – a variable highly correlated with our predictor `class_size`, but whose source of variation is independent from that of our response variable, Test Score. Maimonides' rule gives a function for creating just such a variable (Angrist and Lavy 1999). Whereas the variation in actual `class_size` is problematically endogenous, the variation in *Maimonides' Expected Class Size* comes from an arbitrary formula unrelated to Test Score.

Mamonides' rule is a function for class size that depends only on grade `enrollment` and the number 40:[2]

$$ExpectedClassSize_{sc} = enrollment_s/integer\left[\left(\frac{enrollment_s - 1}{40}\right) + 1\right]$$

For a school s with a 4th grade cohort of 90 students ($enrollment_s = 90$), the expected size of each 4th grade class sc is $90/integer\left[\left(\frac{90-1}{40}\right) + 1\right] =$ **30 students**. **We use this function to create our instrumental variable `class_size_maim`, the Maimonides' Expected Class Size.**

```
## Define the Maimonides Function
maim_func <- function(x) {
  x/(as.integer((x-1)/40) + 1)  }
## Create class_size_maim (the expected Class Size as per Maimonides' rule).
g4_maim <- g4 %>%
  dplyr::select(schlcode, enrollment, class_size, pct_disadvantaged, class_avg_math, class_avg_verbal) %
  mutate(class_size_maim = (sapply(enrollment, maim_func))) %>%
  arrange(enrollment)
```

The figure shows that `class_size` in the population of Israeli Jewish public school 4th grade classes seems to follow the Maimonides' rule of 40 (it is the law, though executed imperfectly). We note the zig-zag pattern of `class_size`, which drops every time the school's 4th grade enrollment reaches a multiple of 40. Actual `class_size` (in grey) follows the Maimonides' Expected Class Size (the instrumental variable `class_size_maim`, in midblue) as grade enrollment increases. This figure mirrors a key visualization from (Angrist and Lavy 1999), but we have created it independently. *See appendix for figure code.*

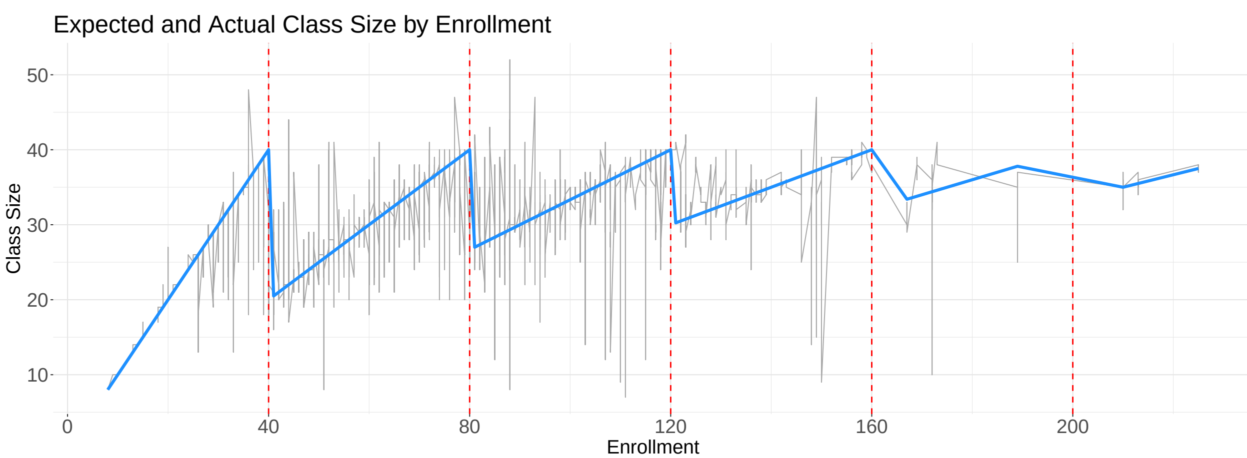


Figure 3: *The zig-zag pattern of Class Size as Grade Enrollment increases in the population of Jewish public school 4th grade cohorts in Israel. Both axes show number of students. Enrollment multiples of 40 marked by red dashed lines. Actual Class Size in light grey. The instrumental variable, Expected Class Size as per Maimonides' Rule, in midblue.*

The instrumental variable `class_size_maim` is highly correlated with actual `class_size` (r = 0.73). This is the first property of a good instrument: `class_size_maim` will be able to predict actual `class_size` with high precision. We will further elaborate on this in the next section.

[2] Interpreting the Talmud, Maimonides writes that "Twenty-five children may be put in charge of one teacher. If the number in the class exceeds twenty-five but is not more than forty, he should have an assistant to help with the instruction. If there are more than forty, two teachers must be appointed." – from Angrist and Lavy (1999)

```
cor(g4_maim$class_size, g4_maim$class_size_maim)
```

```
## [1] 0.7386586
```

4b. Constructing the Instrumental Variable

As cited by Nascimento, the instrumental variable (IV) is "a variable correlated with the intervention but otherwise unrelated to the test score (Webbink 2005). The net effect of class size can be estimated with Two Stage Least Squares (2SLS): first, regressing class size on the IV and other exogenous variables, and then regressing the test score on…" *Predicted Class Size due to the IV*, "together with other controls." Accordingly, we will carry out the Instrumental Variable technique in two stages:

1. Make a **Predicted** Class Size variable by regressing actual `class_size` on the instrumental variable `class_size_maim`, the Maimonides' Expected Class Size. Include other controls.
2. Use this `pred_class_size` as the explanatory variable for Test Scores. Unlike the original `class_size`, the variation in `pred_class_size` is independent of school interactions, parental involvement, socioeconomic leverage, or any other unobservable factors also responsible for variation in Test Scores.

Once we predict `class_size` using the Maimonides' Rule, the resulting `pred_class_size` variable has two properties: 1) it is highly correlated with the original `class_size` and 2) its **only** source of variation is the arbitrary formula of Maimonides' rule. The Instrumental Variable method gives us a covariate whose source of variation is independent of Test Scores, but whose variation can explain variation in Test Scores!

```
## STEP 1: CREATE INSTRUMENTAL VARIABLE
## Regress actual class size on class_size_maim (and other desired covariates)
iv_fit <- lm(data = g4_maim, class_size ~ class_size_maim + pct_disadvantaged + enrollment)
## Then predict class size using this model
g4_maim <- g4_maim %>%
  mutate(pred_class_size = predict.lm(iv_fit))
cor(g4_maim$class_size, g4_maim$pred_class_size)
```

```
## [1] 0.7555464
```

5. Analysis and Results

5a. Instrumental Variable Estimates of Class Size Effects

Now, we can replace the problematic `class_size` with our new variable `pred_class_size` as the predictor of Test Score. If the coefficient on `pred_class_size` is still zero, we will have strong evidence to believe this coefficient. Whatever variation in Test Score that is explained by the variation in **Predicted** Class Size will now be an unbiased estimate of the effect of Class Size on Test Score.

```
## STEP 2: FIT MODEL USING IV-PREDICTED CLASS SIZE
fit2_verbal_iv <- lm(data = g4_maim, class_avg_verbal ~ pred_class_size + pct_disadvantaged + enrollmen
summary(fit2_verbal_iv)$coefficients[,1:4]
```

```
##                        Estimate  Std. Error     t value      Pr(>|t|)
## (Intercept)        80.842079838 1.211397929  66.7345369  0.000000e+00
## pred_class_size    -0.128963177 0.047914615  -2.6915207  7.170488e-03
## pct_disadvantaged -0.344482962 0.011778219 -29.2474575 1.814333e-157
## enrollment          0.004056786 0.006050397   0.6704992  5.026151e-01
```

```r
fit2_math_iv <- lm(data = g4_maim, class_avg_math ~ pred_class_size + pct_disadvantaged + enrollment)
summary(fit2_math_iv)$coefficients[,1:4]
```

```
##                      Estimate  Std. Error     t value     Pr(>|t|)
## (Intercept)       72.64326868 1.429154256  50.8295507 0.000000e+00
## pred_class_size   -0.04587316 0.056527565  -0.8115184 4.171622e-01
## pct_disadvantaged -0.28286065 0.013895427 -20.3563839 4.737562e-84
## enrollment         0.01936223 0.007137994   2.7125590 6.732365e-03
```

The coefficient on `pred_class_size` is statistically significant for Class Average Verbal Score but not for Math Score. The instrumental variable estimate of the effect of Class Size on Class Average Verbal Score is -0.129 percentage points *per unit increase in Class Size*. We interpret this as a -0.129 percentage reduction in Class Average Verbal Score *for every additional student*. If Class Size increases by 20 students, then the Class Average Verbal Score is estimated to decrease by 2.58 percentage points. This is a sizeable decrease worthy of policy attention.

Interestingly, Class Size matters for Verbal Scores but not for Math Scores. It is possible that Verbal aptitude benefits from a discussion-based environment, in which there is more student-student or student-teacher interaction. This may not hold true for a math learning environment. Indeed, even at Yale many higher level math classes are taught as large lectures, whereas English or writing classes often utilize smaller seminar environments.

5b. Interaction Effects

Now let's see if the effect of Class Size changes when interacted with another variable, `enrollment`. *See appendix for full regression output.*

```r
regverb <- lm(data = g4_maim, class_avg_verbal ~ pred_class_size*enrollment + pct_disadvantaged)
summary(regverb)$coefficients[c(2,3,5),1:4]
```

```
##                                Estimate   Std. Error   t value     Pr(>|t|)
## pred_class_size            -0.249593236 0.0584768752 -4.268238 0.0000206004
## enrollment                 -0.102179933 0.0303010397 -3.372159 0.0007596687
## pred_class_size:enrollment  0.003054988 0.0008539035  3.577673 0.0003547094
```

```r
regmath <- lm(data = g4_maim, class_avg_math ~ pred_class_size*enrollment + pct_disadvantaged)
summary(regmath)$coefficients[c(2,3,5),1:4]
```

```
##                                Estimate  Std. Error    t value  Pr(>|t|)
## pred_class_size            -0.092484669 0.069180461 -1.3368611 0.1814163
## enrollment                 -0.021687686 0.035847331 -0.6050014 0.5452450
## pred_class_size:enrollment  0.001180449 0.001010202  1.1685276 0.2427299
```

Both the regression output and interaction plot show a statistically significant interaction effect between `enrollment` and `pred_class_size` on Class Average Verbal Score. However, the magnitude of the interaction coefficient is small. A one-student increase in `enrollment` slightly mitigates the negative effect of increasing Class Size on Class Average Verbal Score. A similar analysis for Math Score shows no evidence of an interaction. A possible explanation for this may the fact that in Israel, larger schools (with higher `enrollment`) are located in cities and urban centers, where student achievement may be higher on average. Smaller schools may be located in less developed towns, where average test scores are lower.

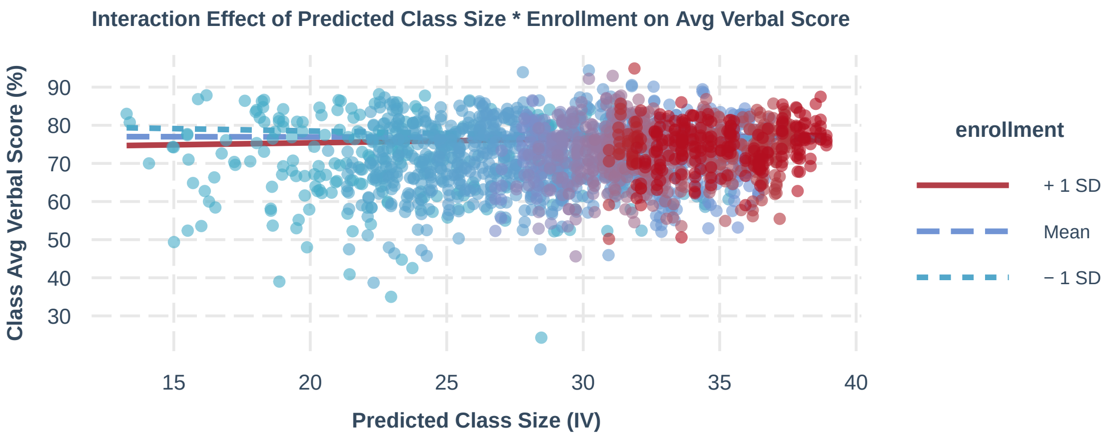


Figure 4: Interaction Effect of Predicted Class Size and Enrollment on Class Avg Verbal Score. The interaction effect exists for Verbal Score but not for Math Score.

Reflection

This project cemented my understanding of hypothesis testing and practical applications of one-way ANOVA. Implementing the dplyr and ggplot techniques we learned in class to arrange the data frame for my analysis and make the specific visualizations I wanted was an intensive and good practice with coding in R. I felt satisfied being able to implement a difficult technique, namely Instrumental Variable (IV) estimation using 2-stage linear regression in R. I spent 30-35 hours on this project.

References

Angrist, Joshua D., and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *The Quarterly Journal of Economics* 114 (2): 533–75. https://doi.org/10.1162/003355399556061.

Hyamson, Moses. 1937. "Annotated English Translation of Maimonides' Mishneh Torah." *Book I (The Book of Knowledge)(New York: Jewish Theological Seminary, 1937)*.

Mueller, C. I. Chase, D., and J. D. Walden. 1988. "Effects of Reduced Class Size in Primary Classes." *Educational Leadership* 45 (5): 48–50.

Nascimento, Paulo A Meyer M. 2009. "How Good Is Maimonides' Rule? Strengths and Pitfalls of Using Institutional Rules as Instruments to Assess Class Size Effects." *SciELO Books*.

Robinson, Glen E. 1990. "Synthesis of Research on the Effects of Class Size." *Educational Leadership* 47 (7): 80–90.

Steele, Fiona, Anna Vignoles, and Andrew Jenkins. 2007. "The Effect of School Resources on Pupil Attainment: A Multilevel Simultaneous Equation Modelling Approach." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 170 (3): 801–24.

Webbink, Dinand. 2005. "Causal Effects in Education." *Journal of Economic Surveys* 19 (4): 535–60.

Appendix

2. Loading the data

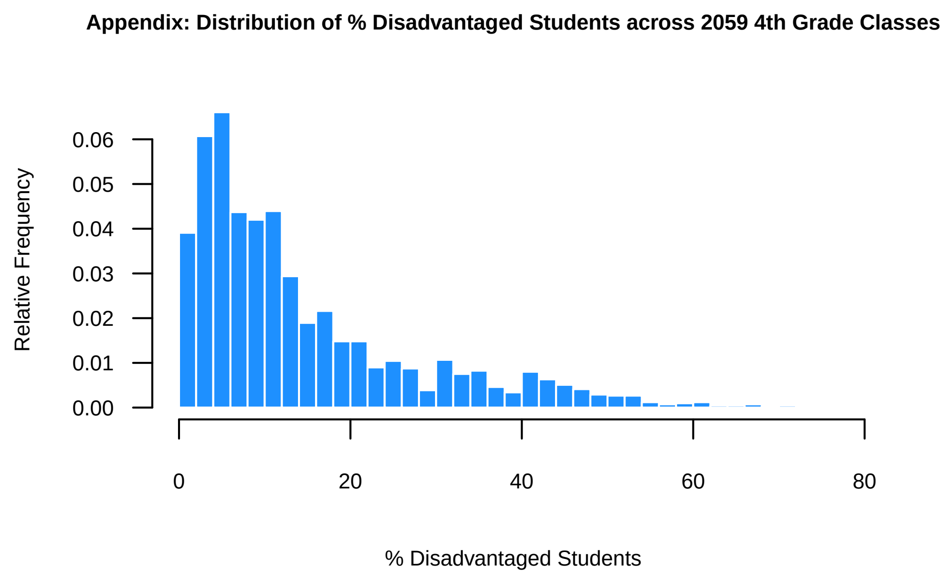
```
library(ggplot2)
library(tidyr)
```

```r
library(dplyr)
library(MASS)
library(foreign)

g4 <- read.dta("final4.dta")
g4 <- g4 %>%
  rename(class_avg_math = avgmath) %>%
  rename(class_avg_verbal = avgverb) %>%
  rename(class_size = classize) %>%
  rename(pct_disadvantaged = tipuach) %>%
  rename(enrollment = c_size)
```

3a. Pre-analysis on Percent Disadvantaged Students

- Distribution of Percent Disadvantaged Students

```r
truehist(g4_maim$pct_disadvantaged, col = "dodgerblue", border = "white", las = 1, h = 2,
         main = "Appendix: Distribution of % Disadvantaged Students across 2059 4th Grade Classes",
         xlab = "% Disadvantaged Students", ylab = "Relative Frequency",
         cex.main = 0.65, cex.axis = 0.65, cex.lab = 0.65)
```




```r
## Welch's T-Test for Class Average Verbal Score
t.test(disad_verb, nondisad_verb)
```

```
##
##  Welch Two Sample t-test
##
## data:  disad_verb and nondisad_verb
## t = -22.2, df = 766.72, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -9.532028 -7.983237
## sample estimates:
```

```
## mean of x mean of y
##  65.95434  74.71197
```

```r
## Welch's T-Test for Class Average Math Score
t.test(disad_math, nondisad_math)
```

```
##
##  Welch Two Sample t-test
##
## data:  disad_math and nondisad_math
## t = -18.453, df = 794.82, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.986386 -7.258376
## sample estimates:
## mean of x mean of y
##  62.80054  70.92292
```

4a. ANOVA output for regression of class average Math score on class size

```r
fit1_math <- lm(data = g4_maim, class_avg_math ~ class_size)
anova(fit1_math)
```

```
## Analysis of Variance Table
##
## Response: class_avg_math
##              Df Sum Sq Mean Sq F value    Pr(>F)
## class_size    1   3734  3733.8  49.669 2.475e-12 ***
## Residuals  2053 154332    75.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
fit2_math <- lm(data = g4, class_avg_math ~ class_size + pct_disadvantaged + enrollment)
anova(fit2_math)
```

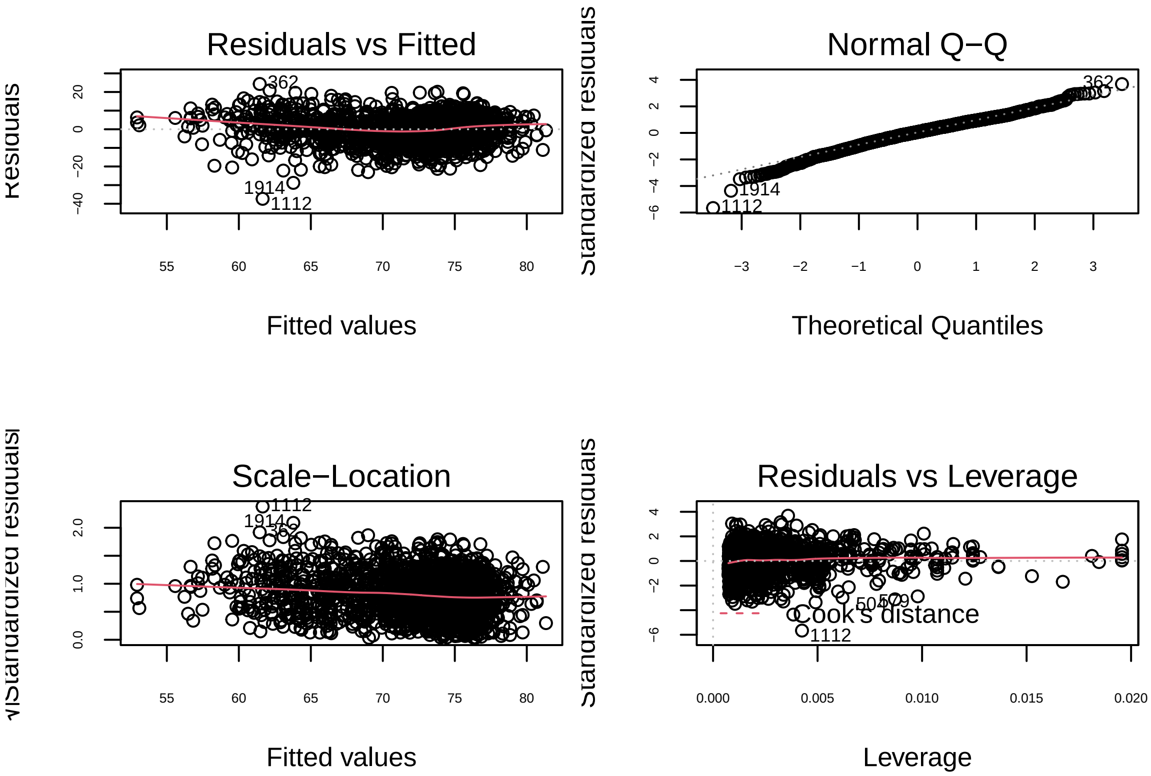
```
## Analysis of Variance Table
##
## Response: class_avg_math
##                     Df Sum Sq Mean Sq  F value    Pr(>F)
## class_size           1   3734  3733.8  60.9448 9.284e-15 ***
## pct_disadvantaged    1  28270 28270.5 461.4441 < 2.2e-16 ***
## enrollment           1    406   406.3   6.6318   0.01009 *
## Residuals         2051 125655    61.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Plotting the interaction regressions

```r
par(mfrow = c(2,2), mai = c(1, .6, 0.5, 0.1))

summary(regverb)
```

```
##
## Call:
## lm(formula = class_avg_verbal ~ pred_class_size * enrollment +
##     pct_disadvantaged, data = g4_maim)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -37.364  -3.871   0.452   4.408  24.303
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)               85.1619482  1.7079317  49.863  < 2e-16 ***
## pred_class_size           -0.2495932  0.0584769  -4.268 2.06e-05 ***
## enrollment                -0.1021799  0.0303010  -3.372 0.000760 ***
## pct_disadvantaged         -0.3460114  0.0117523 -29.442  < 2e-16 ***
## pred_class_size:enrollment  0.0030550  0.0008539   3.578 0.000355 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.615 on 2050 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.3153, Adjusted R-squared:  0.314
## F-statistic:   236 on 4 and 2050 DF,  p-value: < 2.2e-16
```

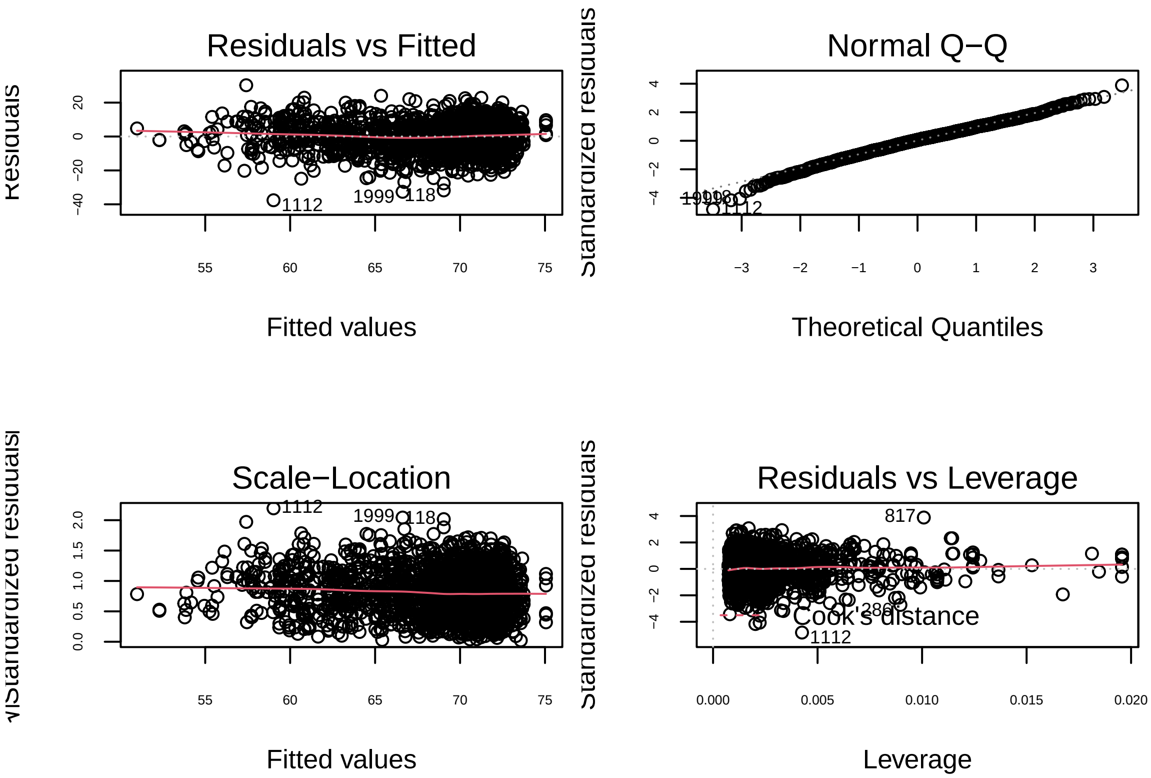
```r
plot(regverb, cex.main = 0.5, cex.axis = 0.5)
```




```r
summary(regmath)
```

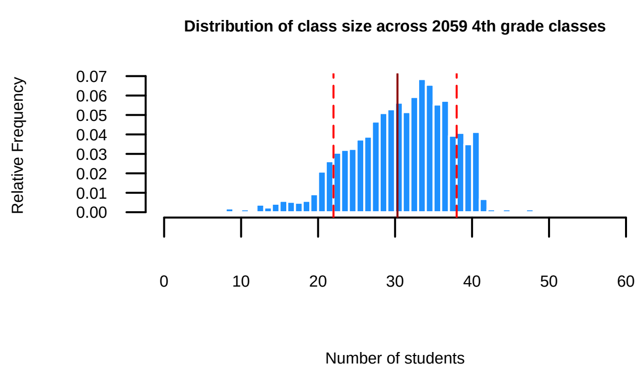
```
##
## Call:
## lm(formula = class_avg_math ~ pred_class_size * enrollment +
##     pct_disadvantaged, data = g4_maim)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -37.612  -4.926   0.464   5.236  30.256
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 74.31247    2.02055  36.778   <2e-16 ***
## pred_class_size             -0.09248    0.06918  -1.337    0.181
## enrollment                  -0.02169    0.03585  -0.605    0.545
## pct_disadvantaged           -0.28345    0.01390 -20.387   <2e-16 ***
## pred_class_size:enrollment  0.00118    0.00101   1.169    0.243
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.825 on 2050 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.2058, Adjusted R-squared:  0.2043
## F-statistic: 132.8 on 4 and 2050 DF,  p-value: < 2.2e-16
```

```r
plot(regmath, cex.main = 0.5, cex.axis = 0.5)
```



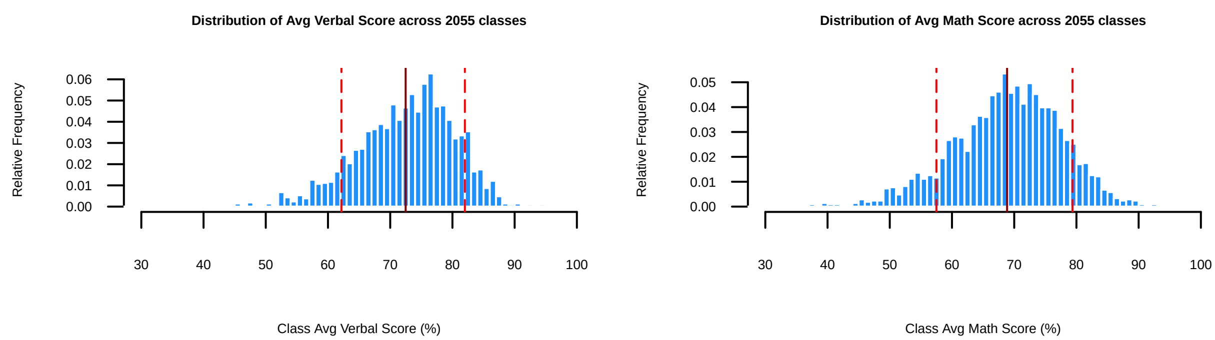
- Code for Figure 1

```r
# c(mean(g4$class_size), sd(g4$class_size))

par(mfrow = c(2, 1), mai = c(1, .8, 0.5, 0.1))
truehist(g4$class_size, h = 1, las = 1, xlim = c(0, 60),
         border = "white", col = "dodgerblue",
         main = "Distribution of class size across 2059 4th grade classes",
         ylab = "Relative Frequency", xlab = "Number of students",
         cex.lab = 0.5, cex.main = 0.5, cex.axis = 0.5)
abline(v = mean(g4$class_size), col = "darkred")
abline(v = quantile(g4$class_size, c(0.1, 0.9)), col = "red", lty = 5)
```



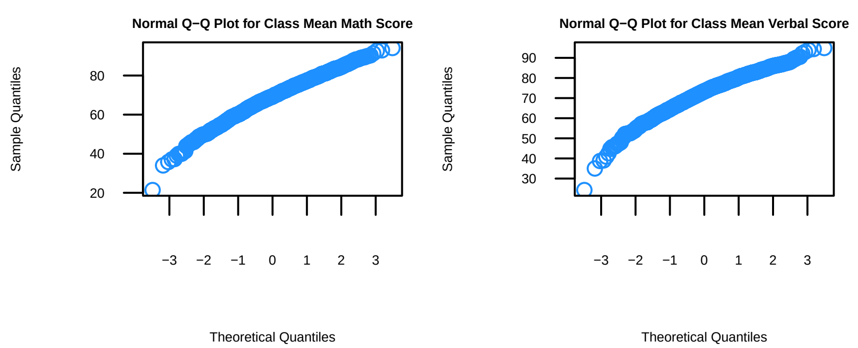
- Code for Figure 2

```r
g4_yvar <- g4 %>%
  filter(is.na(class_avg_math) == 0, is.na(class_avg_verbal) == 0)
#c(mean(g4_yvar$class_avg_verbal), sd(g4_yvar$class_avg_verbal))
#c(mean(g4_yvar$class_avg_math), sd(g4_yvar$class_avg_math))

par(mfrow = c(2, 2), mai = c(1, .7, 0.5, 0.1))
truehist(g4_yvar$class_avg_verbal, h = 1, las = 1, xlim = c(30, 100),
         border = "white", col = "dodgerblue",
         main = "Distribution of Avg Verbal Score across 2055 classes",
         ylab = "Relative Frequency", xlab = "Class Avg Verbal Score (%)",
         cex.main = 0.5, cex.axis = 0.5, cex.lab = 0.5)
abline(v = mean(g4_yvar$class_avg_verbal), col = "darkred")
abline(v = quantile(g4_yvar$class_avg_verbal, c(0.1, 0.9)), col = "red", lty = 5)

truehist(g4$class_avg_math, h = 1, las = 1, xlim = c(30, 100),
         border = "white", col = "dodgerblue",
         main = "Distribution of Avg Math Score across 2055 classes",
         ylab = "Relative Frequency", xlab = "Class Avg Math Score (%)",
         cex.main = 0.5, cex.axis = 0.5, cex.lab = 0.5)
abline(v = mean(g4_yvar$class_avg_math), col = "darkred")
abline(v = quantile(g4_yvar$class_avg_math, c(0.1, 0.9)), col = "red", lty = 5)
```



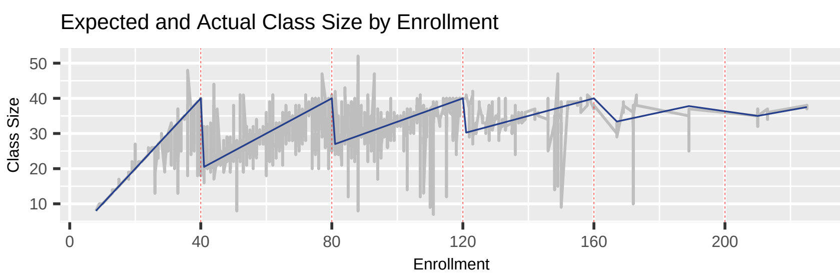
- Code for Figure 3

```r
par(mfrow = c(1:2),  mai = c(1, .8, 0.2, 0.1))
qqnorm(g4$class_avg_math, col = "dodgerblue", las = 1,
       cex.axis = 0.4, cex.main = 0.4, cex.lab = 0.4,
       main = "Normal Q-Q Plot for Class Mean Math Score")
qqnorm(g4$class_avg_verbal, col = "dodgerblue", las = 1,
       cex.axis = 0.4, cex.main = 0.4, cex.lab = 0.4,
       main = "Normal Q-Q Plot for Class Mean Verbal Score")
```



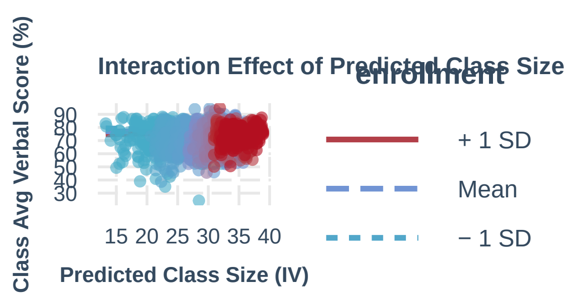
- Code for Figure 4

```
## Zig-zag pattern of Class Size as Enrollment Increases
g4 <- g4 %>%
  arrange(enrollment, class_size)

ggplot(data = g4_maim, aes(x=enrollment, y=class_size)) + geom_line(color = "grey") +
  scale_x_continuous(name = "Enrollment", breaks = seq(0, 200, by = 40)) +
  geom_vline(xintercept = c(40, 80, 120, 160, 200), color = "red", linetype="dashed", size=.1) +
  geom_line(data = g4_maim, aes(x = enrollment, y = class_size_maim),
            color = "royalblue4", linetype="solid", size=.3) +
  theme(axis.text.x = element_text(size = 6), axis.text.y = element_text(size = 6),
        axis.title = element_text(size = 6), plot.title  = element_text(size = 8)) +
  labs(title = "Expected and Actual Class Size by Enrollment",
       x = "Enrollment", y = "Class Size")
```

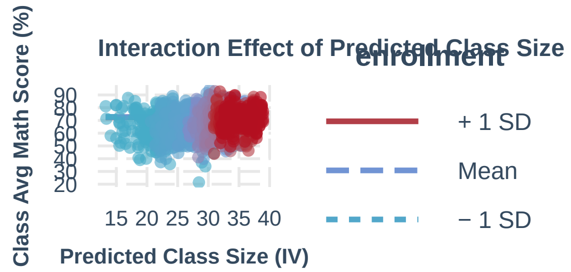


- Code for Figure 5

```
library(interactions)
library(wesanderson)

interact_plot(regverb, pred = pred_class_size, modx = enrollment, centered = "none", plot.points = TRUE
  scale_y_continuous(name = "Class Avg Verbal Score (%)", breaks = seq(10, 100, by = 10)) +
  theme(axis.text.x = element_text(size = 8), axis.text.y = element_text(size = 8),
        axis.title = element_text(size = 8), plot.title  = element_text(size = 9)) +
  labs(title = "Interaction Effect of Predicted Class Size * Enrollment on Avg Verbal Score")
```




```
interact_plot(regmath, pred = pred_class_size, modx = enrollment, centered = "none", plot.points = TRUE
  scale_y_continuous(name = "Class Avg Math Score (%)", breaks = seq(10, 100, by = 10)) +
  theme(axis.text.x = element_text(size = 8), axis.text.y = element_text(size = 8),
        axis.title = element_text(size = 8), plot.title  = element_text(size = 9)) +
  labs(title = "Interaction Effect of Predicted Class Size * Enrollment on Avg Math Score")
```


Interaction Effect of Predicted Class Size *
enrollment
Class Avg Math Score (%)
90
80
70
60
50
40
30
20
15 20 25 30 35 40
Predicted Class Size (IV)
+ 1 SD
Mean
− 1 SD