# Quantile Regression on SAT score

Young Seok Seo

## Introduction

Education has taken paramount role in various revolutions in previous era, and it is still one of the keys to develop our world. Renaissance, Reformation, Industrial Revolution, and other more advanced revolutions were led by educated people. Education would not disappear till end of the world. Nowadays, some people argue that machines will do everything, and they will substitute human in the future. However, Machines still need to learn, and human should educate them. Even Covid-19 could not obstruct education in the world. In United States, students should take qualification exam, ACT or SAT, to enter the college. Most of high school students study hard in ACT or SAT to get into well-known university or college. Eminent economists and other professionals researched that education is affected by many aspects in our life. Actually, Quantile Regression method is used to find how quality of school or size of class may impacts on scholastic achievement of student. (Levin 2001)

In this paper, I use ap_2010, class_size, demographics, hs_directory, sat_results datasets from Kaggle to find how New York City High school' Average SAT results has relationship with independent variables at each quantile. Using Pandas package in Python, I manipulate the data and get a data frame of 363 high schools in NYC and 165 independent variables. Further, I selected 14 variables, SAT Score, Number of Exams with scores 3 4 or 5, Average Class Size, Total Enrollment, English learner Percentage, Asian, Black, White, and Hispanic Percentage, Male Percentage, Female Percentage, School District, and Safety Level. Then extract the data frame to .csv file to apply Quantile Regression in R using 'quantreg' package. Then using 'PSG' library, have confirmed that results from 'quantreg' library and that from 'PSG' is same.

## Executive Summary

Number of scores 3.4 or 5 on AP Test, Average Class Size of school, Total Enrollment, Asian and white percentage, and safety level have positive influence on SAT score. Meanwhile, English Leaner, Black and Hispanic percentage indicated negative influence on SAT score. Also, percentage of gender and location of school does not impact on Average SAT score of High School in NYC. With those conclusion, high schools those are in lower quantile on SAT score can have strategy to increase average SAT score.

## Methodology – Quantile Regression

Ordinary Least Squares estimates the conditional mean of the response variable across values of the predictor variables. Also, it minimizes the distance between the values predicted by regression line. However, assumption of normally distributed and linearity should be required to use OLS method. Meanwhile, Quantile Regression is useful when the data does not follow normal distribution and non-linearity. Moreover, Quantile Regression differentially weights the distance between the values by the regression line, then tries to minimize the weighted distance. Therefore, it is more robust against outliers in the response measurements.

In OLS, we all know that this equation is used. Here, $\epsilon_i$ is the error term which normally distributed to 0. And we could get $\beta$ by second equation.

$$y_i \ = \ x_i'\beta \ + \ \epsilon_i$$

$$\hat{\beta} = min \sum_{i=1}^{n} (y_i - x_i'\beta)^2$$

Similar concept may apply to use Quantile Regression. When independent variable $x$ is given, we can define $\tau^{\text{th}}$ value for response variable $y$.

$$Q_\tau(y_i|x_i) = x_i'\beta_\tau$$

Here, we need to get value for $\beta_\tau$.

$$\hat{\beta}_\tau \ = min \sum_{i=1}^{n} \rho_\tau(y_i - x_i'\beta_\tau)$$

$\rho_\tau$ is the check function that

$$\rho_\tau = \tau x \ I(x > 0) + (\tau - 1) \ x \ I(xn < 0)$$

Therefore, value for $y_t$ at each $\tau$ is able to get with this equation.

$$\hat{Q}_\tau(y_t|x_t) = x_t'\beta_\tau$$

## Analysis

Before moving on to modeling of quantile regression, we need to check our response variable has outliers so that the method is whether appropriate or not. According to *Figure 1*, data is normally distributed in $[-2\sigma, 2\sigma]$. However, there is extreme values in the right tail. *Figure 2* also support the exist of outliers.

Now, I would like to dive into the analysis. Goal of this quantile regression is to see how independent variables influence on each quantile of response variable. I visualize how each variable influence on SAT score. Red line indicates OLS mean value and red dotted lines shows 95% confidence interval. Black dot represents coefficients of beta at $\tau^{th}$ quantile. If coefficient of beta in any quantile is in between 95% confidence interval, we can conclude that a variable does not influence on response variable at those quantiles.

**Figure 1. QQ plot of Avg.SAT Scores of NYC Schools**

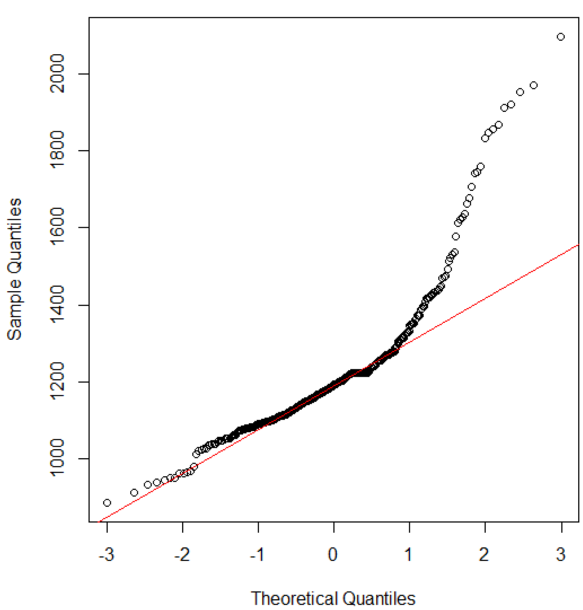
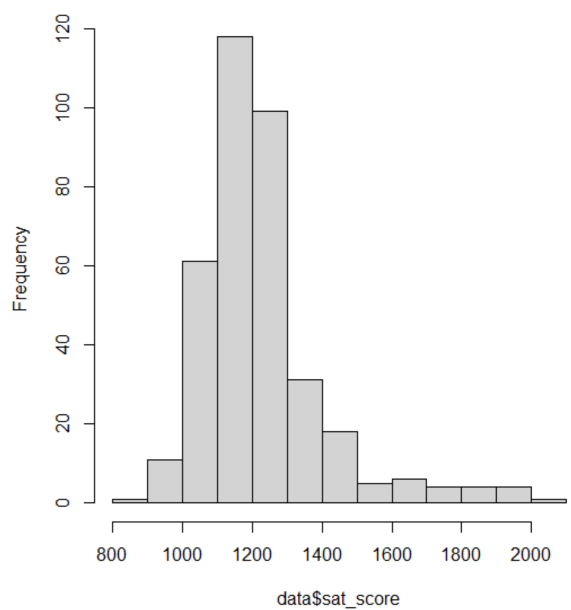


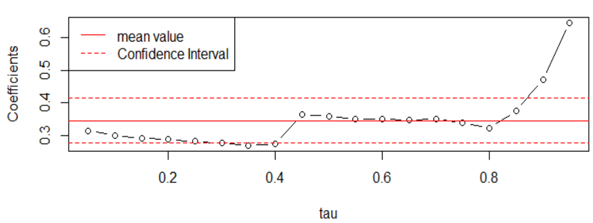

**Figure 2. Histogram of Avg.SAT scores of NYC schools**



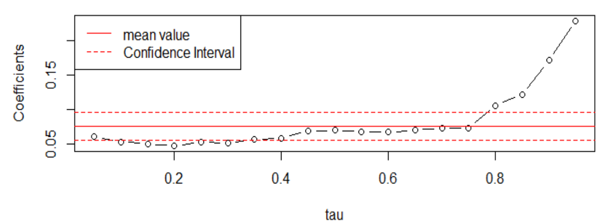

As shown in Figure *3*, influence of Number of Exams with score 3,4,or 5 on AP Exam on SAT score grows exponentially after $80^{th}$ quantile. Coefficient of $75^{th}$ quantile is 0.34 as stated in *Table 1*. However, it increases to 0.65 in $95^{th}$ quantile. With that information, we cannot say that this growth is statistically significant. ANOVA Test can provide statistical sense. $p - value$ in *ANOVA: Number of Exams with Scores 3.4 or 5* is 0.587. Although *Figure 3* shows that the degree of influence of variable in 5th quantile may be different from that of 95th quantile, statistical test does not argue that their influence has huge differences.

In *Figure 4*, influence of Average Class Size increases overall as quantile increases. Coefficient value at 95th quantile is twice of that of 5th quantile in *Table 1*. However, similar to the Number of EXAMS 3,4, or 5, $p - value$ = 0.4946. Also, we can see that most of data points in *Figure 4* are located between the confidence interval, we can assure that degree of influence of variable between 5th and 95th quantile does not have difference.

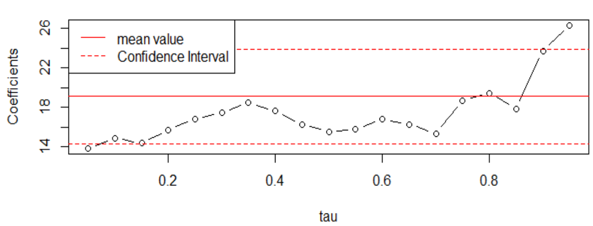
**[Figure 3] Number.of.Exams.with.scores.3.4.or.5**



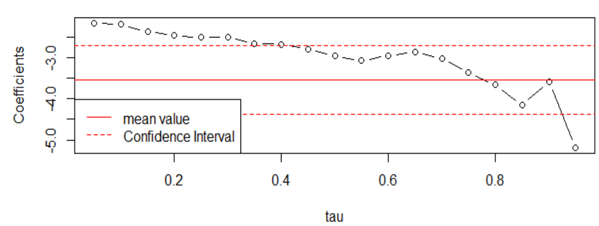

**[Figure 5] Total Enrollment**




**[Figure 4] Avg. Class Size**



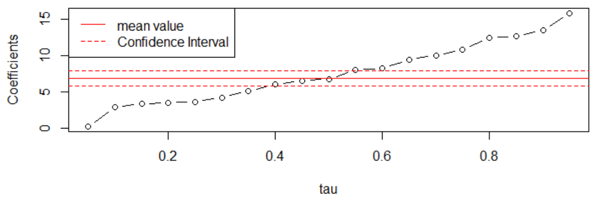

**[Figure 6] English Learner Percentage**



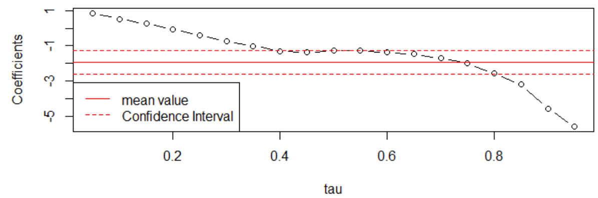

Next one is total enrollment. *Figure 5* represents how total enrollment impacts on each quantile of response variable. Comparing to the data points of coefficient, confidence interval with red-dotted is narrow. Particularly, starting $\tau = 0.8$, the coefficient value increases substantially. According to *Table 1*, coefficient value at $\tau = .75$ is only 0.073 but it increases to 0.228 at $90^{th}$ quantile. *ANOVA: Total Enrollment* also supports that there is significantly difference in degree of influence of total enrollment to SAT score.

English Learner Percentage is quite interesting. In *Figure 6*, the influence of this variable to SAT Score keeps decreasing. In lower quantile, number of English learners may impact on school's average SAT score. However, as $\tau$ increase, the influence of percentage of English learner on average SAT score diminishes. When we compare 10th and 90th quantile as shown in *ANOVA: English Learner Percentage – 10th vs 90th*, we cannot say that degree of influence has significant differences on SAT score because of large $p - value$. However, in *ANOVA: English Learner Percentage – 5th vs 95th*, $p - value$ equals to 7.99e-05. Therefore, we can argue that English Learner Percentage has significant role in SAT score in extreme quantile.

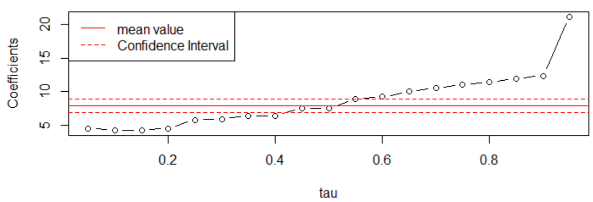
**[Figure 7] Asian Percentage**



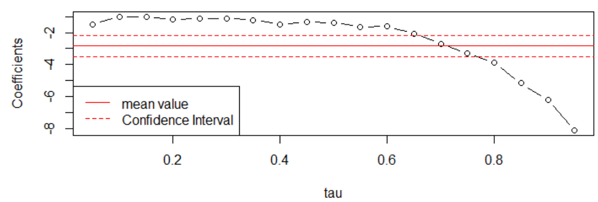

**[Figure 9] Black Percentage**




**[Figure 8] White Percentage**



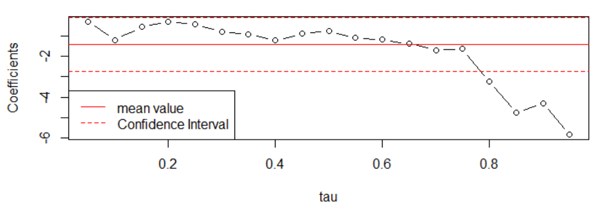

**[Figure 10] Hispanic Percentage**



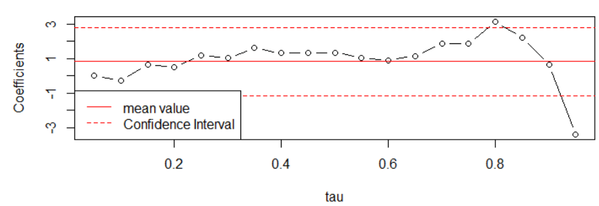

Asian and White percentage trends similarly in quantile regression. In OLS method, mean values of variable are close to each other as well. Mean value of white percentage is 7.88 and that of Asian is 6.91. as shows in *Figure 7* and *Figure 8*, those trends are moving together as τ increases. They both have positive influence on SAT score regardless of quantile. At 5th quantile, coefficient value of Asian percentage is 2.23 and increases to 15.88 at 95th quantile. Coefficient value of White percentage also grows smoothly. In *ANOVA: White Percentage*, $p - value$ is low enough to argue that degree of influence between 5th and 95th have significantly difference but not as much as other variables. Coefficients values of Hispanic and Black percentage changes homogeneously as well. Influence on both variables to school's SAT score keep decreasing. In *Figure 10*, influence does not impact noticeable till 60th quantile. However, influence of Hispanic percentage on SAT score decreases drastically after 60th quantile. In *ANOVA: Black Percentage*, and *ANOVA: Hispanic Percentage* represent that influence of Hispanic and Black percentage on SAT score at 5th quantile has statistically difference in that of at 95th quantile.

Influence on Male and Female percentage to SAT score moves the other way. In *Figure 11*, After 80th quantile, coefficient values are located under the lower bound of confidence interval and drops explicitly. Meanwhile, in *Figure 12*, coefficient values exceed the upper bound of confidence interval and grows exponentially.

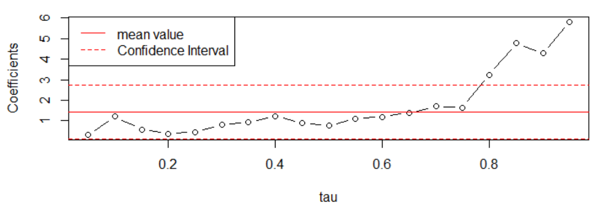
[Figure 11] Male Percentage



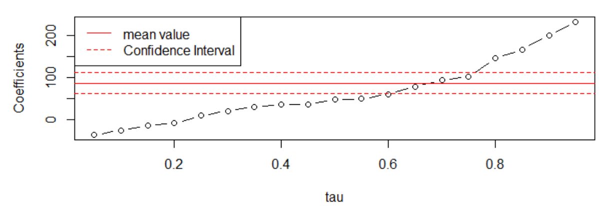

[Figure 13] School District




[Figure 12] Female Percentage




[Figure 14] Safety Level




Influence of location of school does not impacts on school's SAT score. Most of coefficients are in between confidence interval except for 95th quantile. Although *Figure 13* indicates that influence diminishes extremely after 80th quantile, *ANOVA: School District* shows that influence of location does not change statistically from 5th to 95th quantile.

Finally, Safety level has positive influence on SAT score as shown in *Figure 14.* At 5th quantile, the coefficient value for Safety level is approximately -37. However, it keeps increasing and hits 232 at 95th quantile. *ANOVA: Safety Level* also substantiates that there is statistically difference in influence to SAT score between 5th and 95th quantile.

Last but not least, using 'PSG' library in the R, confirmed that the results from both 'quantreg' library and 'PSG' library are exactly same. Beta coefficient for total enrollment for example, I could have 0.06994172 at 50th quantile and 0.170767 at 90th quantile.

```
                               PSG   quantreg
total_enrollment_at_50th 0.06994172 0.06994172

                              PSG quantreg
total_enrollment_at_90th 0.170767 0.170767
```

## Conclusion

In this report, Introduced method of Quantile Regression Estimation that Koenker and Basset(1978) already introduced and analyzed how independent variables may impacts on 363 High School's Average SAT Score in NYC at each quantile. Method of Ordinary Least Square Estimation requires various assumptions to run the regression. Quantile Regression supplement those complicated assumptions and robust against outliers. Due to COVID-19, the influence of factors on SAT score may be different in next few years. However, according to the data of 2016, some factors showed positive influence on SAT score as quantile increases. Number of scores 3.4 or 5 on AP Test, Average Class Size of school, Total Enrollment, Asian and white percentage, and safety level have positive influence on SAT score. Meanwhile, English Leaner, Black and Hispanic percentage indicated negative influence on SAT score. Also, percentage of gender and location of school does not impact on Average SAT score of High School in NYC. With those conclusion, high schools those are in lower quantile on SAT score can have strategy to increase average SAT score.

## Appendix

[Table 1. Coefficients of Quantile Regression]

| | tau = 0.05 | 0.1 | 0.25 | 0.5 | 0.75 | 0.9 | 0.95 |
|---|---|---|---|---|---|---|---|
| *Number of score 3/4/5* | 0.313788952260662 | 0.299579641214896 | 0.280633893153875 | 0.357980397266041 | 0.337962962962963 | 0.470741222366709 | 0.647281921618205 |
| *avg_class_size* | 13.8162666237198 | 14.8700173258683 | 16.8518518674554 | 15.5339805825243 | 18.6802798161648 | 23.7442922374429 | 26.340482591381 |
| *total_enrollment* | 0.0604133545310015 | 0.053495241220873 | 0.0535286284953395 | 0.0699417152373022 | 0.0730748617806195 | 0.170767004341534 | 0.228260869565217 |
| *ell_percent* | -2.14689265536723 | -2.18832891246684 | -2.5 | -2.95857988165681 | -3.35338345864659 | -3.58426966292135 | -5.1959848110832 |
| *asian_per* | 0.223285486443379 | 2.93577981651376 | 3.65517241379311 | 6.78160919540229 | 10.7835820895522 | 13.4883720930233 | 15.8862876254181 |
| *black_per* | 0.86734693877551 | 0.534918276374439 | -0.400000000000001 | -1.28 | -1.99300699300699 | -4.57013574660633 | -5.60773480662982 |
| *white_per* | 4.49516718535469 | 4.26427356962026 | 5.78595317725753 | 7.55186721991701 | 11.0288388559322 | 12.3574144486691 | 21.1811023622047 |
| *hispanic_per* | -1.50384193194292 | -0.989761092150168 | -1.10059171597633 | -1.38849929873773 | -3.31645569620253 | -6.22641509433961 | -8.14141414141414 |
| *male_per* | -0.312499999999999 | -1.2037037037037 | -0.456852791878166 | -0.77253218884121 | -1.64381218750001 | -4.27983539094651 | -5.82342954159592 |
| *female_per* | 0.312500000000003 | 1.20370370370371 | 0.456852791878172 | 0.772532188841205 | 1.64381218749999 | 4.27983539094649 | 5.82342954159593 |
| *school_dist* | -5.10702591327572e-15 | -0.272727272727272 | 1.19999999999998 | 1.34782608695652 | 1.86363636363637 | 0.631578947368433 | -3.42307692307692 |
| *saf_s_11* | -36.9230769230768 | -26.0714285714286 | 8.88888888888888 | 46.762550909091 | 101.111111111111 | 199.375 | 232 |

Quantreg vs PSG

```
                               PSG  quantreg
total_enrollment_at_95th 0.170767 0.170767
```

```
                                 PSG   quantreg
total_enrollment_at_50th 0.06994172 0.06994172
```

[ANOVA : Number of Exams with Scores 3.4.or 5]

```
Quantile Regression Analysis of Deviance Table

Model: sat_score ~ Number.of.Exams.with.scores.3.4.or.5
Joint Test of Equality of Slopes: tau in {  0.05 0.95  }

  Df Resid Df F value Pr(>F)
1  1      725  0.2948 0.5873
```

[ANOVA : Total Enrollment]

```
Quantile Regression Analysis of Deviance Table

Model: sat_score ~ total_enrollment
Joint Test of Equality of Slopes: tau in {  0.05 0.95  }

  Df Resid Df F value    Pr(>F)
1  1      725  47.953 9.631e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

[ANOVA : English Learner Percentage – 5th vs 95th]

```
Quantile Regression Analysis of Deviance Table

Model: sat_score ~ ell_percent
Joint Test of Equality of Slopes: tau in {  0.05 0.95  }

  Df Resid Df F value   Pr(>F)
1  1      725   15.74 7.99e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

[ANOVA : Asian Percentage]

```
Quantile Regression Analysis of Deviance Table

Model: sat_score ~ black_per
Joint Test of Equality of Slopes: tau in {  0.05 0.95  }

  Df Resid Df F value    Pr(>F)
1  1      725  21.442 4.323e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

[ANOVA : Black Percentage]

```
Quantile Regression Analysis of Deviance Table

Model: sat_score ~ asian_per
Joint Test of Equality of Slopes: tau in {  0.05 0.95  }

  Df Resid Df F value   Pr(>F)
1  1      725  10.114 0.001534 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

[ANOVA : White Percentage]

```
Quantile Regression Analysis of Deviance Table

Model: sat_score ~ white_per
Joint Test of Equality of Slopes: tau in {  0.05 0.95  }

  Df Resid Df F value  Pr(>F)
1  1      725  6.2467 0.01266 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

[ANOVA : Average Class Size]

```
Quantile Regression Analysis of Deviance Table

Model: sat_score ~ AVERAGE.CLASS.SIZE
Joint Test of Equality of Slopes: tau in {  0.05 0.95  }

  Df Resid Df F value Pr(>F)
1  1      725  0.4669 0.4946
```

[ANOVA : English Learner Percentage – 10th vs 90th]

```
Quantile Regression Analysis of Deviance Table

Model: sat_score ~ ell_percent
Joint Test of Equality of Slopes: tau in {  0.1 0.9  }

  Df Resid Df F value Pr(>F)
1  1      725  1.1506 0.2838
```

[ANOVA : Female Percentage]

```
Quantile Regression Analysis of Deviance Table

Model: sat_score ~ female_per
Joint Test of Equality of Slopes: tau in {  0.05 0.95  }

  Df Resid Df F value Pr(>F)
1  1      725  0.8635 0.3531
```

[ANOVA : School District]

```
Quantile Regression Analysis of Deviance Table

Model: sat_score ~ school_dist
Joint Test of Equality of Slopes: tau in {  0.05 0.95  }

  Df Resid Df F value Pr(>F)
1  1      725  0.1613 0.6881
```

[ANOVA : Safety Level]

```
Quantile Regression Analysis of Deviance Table

Model: sat_score ~ saf_s_11
Joint Test of Equality of Slopes: tau in {  0.05 0.95  }

  Df Resid Df F value   Pr(>F)
1  1      725  26.209 3.93e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## R-Code

```
library(quantreg)
library(dplyr)
library(tidyr)
library(PSG)
library(gridExtra)
library(grid)
data <- read.csv("C:/Users/ryans/Desktop/AMS/518/Project/NY
School SAT factors (Pandas)/sat.csv")
data<-data[2:14]
summary(data$sat_score)
par(mfrow=c(1,2))
qqnorm(data$sat_score, main = "Figure 1. QQ plot of Avg.SAT
Scores of NYC Schools")
qqline(data$sat_score, col ='red')
hist(data$sat_score, main = "Figure 2. Histogram of Avg.SAT
scores of NYC schools")
par(mfrow=c(1,1))
Num_score_0.05<-summary(rq(sat_score ~
Number.of.Exams.with.scores.3.4.or.5,data = data, tau = 0.05))
Num_score_0.10<-summary(rq(sat_score ~
Number.of.Exams.with.scores.3.4.or.5,data = data, tau = 0.10))
Num_score_0.25<-summary(rq(sat_score ~
Number.of.Exams.with.scores.3.4.or.5,data = data, tau = 0.25))
Num_score_0.50<-summary(rq(sat_score ~
Number.of.Exams.with.scores.3.4.or.5,data = data, tau = 0.50))
Num_score_0.75<-summary(rq(sat_score ~
Number.of.Exams.with.scores.3.4.or.5,data = data, tau = 0.75))
Num_score_0.90<-summary(rq(sat_score ~
Number.of.Exams.with.scores.3.4.or.5,data = data, tau = 0.90))
Num_score_0.95<-summary(rq(sat_score ~
Number.of.Exams.with.scores.3.4.or.5,data = data, tau = 0.95))

comparison_quantile_beta<-
data.frame(cbind(Num_score_0.05$coefficients[2],
Num_score_0.10$coefficients[2],
Num_score_0.25$coefficients[2],
Num_score_0.50$coefficients[2],
Num_score_0.75$coefficients[2],
Num_score_0.90$coefficients[2],
Num_score_0.95$coefficients[2]))
avg_class_size_0.05<-summary(rq(sat_score ~
AVERAGE.CLASS.SIZE,data = data, tau = 0.05))
avg_class_size_0.10<-summary(rq(sat_score ~
AVERAGE.CLASS.SIZE,data = data, tau = 0.10))
avg_class_size_0.25<-summary(rq(sat_score ~
AVERAGE.CLASS.SIZE,data = data, tau = 0.25))
avg_class_size_0.50<-summary(rq(sat_score ~
AVERAGE.CLASS.SIZE,data = data, tau = 0.50))
avg_class_size_0.75<-summary(rq(sat_score ~
AVERAGE.CLASS.SIZE,data = data, tau = 0.75))
avg_class_size_0.90<-summary(rq(sat_score ~
AVERAGE.CLASS.SIZE,data = data, tau = 0.90))
avg_class_size_0.95<-summary(rq(sat_score ~
AVERAGE.CLASS.SIZE,data = data, tau = 0.95))
avg_class_size <-
data.frame(cbind(avg_class_size_0.05$coefficients[2],
        avg_class_size_0.10$coefficients[2],
        avg_class_size_0.25$coefficients[2],
        avg_class_size_0.50$coefficients[2],
        avg_class_size_0.75$coefficients[2],
        avg_class_size_0.90$coefficients[2],
        avg_class_size_0.95$coefficients[2]))
total_enrollment_0.05<-summary(rq(sat_score ~
total_enrollment,data = data, tau = 0.05))
total_enrollment_0.10<-summary(rq(sat_score ~
total_enrollment,data = data, tau = 0.10))
total_enrollment_0.25<-summary(rq(sat_score ~
total_enrollment,data = data, tau = 0.25))
total_enrollment_0.50<-summary(rq(sat_score ~
total_enrollment,data = data, tau = 0.50))
total_enrollment_0.75<-summary(rq(sat_score ~
total_enrollment,data = data, tau = 0.75))
total_enrollment_0.90<-summary(rq(sat_score ~
total_enrollment,data = data, tau = 0.90))
total_enrollment_0.95<-summary(rq(sat_score ~
total_enrollment,data = data, tau = 0.95))
total_enrollment <-
data.frame(cbind(total_enrollment_0.05$coefficients[2],
            total_enrollment_0.10$coefficients[2],
            total_enrollment_0.25$coefficients[2],
            total_enrollment_0.50$coefficients[2],
            total_enrollment_0.75$coefficients[2],
            total_enrollment_0.90$coefficients[2],
            total_enrollment_0.95$coefficients[2]))
ell_percent_0.05<-summary(rq(sat_score ~ ell_percent,data = data,
tau = 0.05))
ell_percent_0.10<-summary(rq(sat_score ~ ell_percent,data = data,
tau = 0.10))
```

```
ell_percent_0.25<-summary(rq(sat_score ~ ell_percent,data = data,
tau = 0.25))
ell_percent_0.50<-summary(rq(sat_score ~ ell_percent,data = data,
tau = 0.50))
ell_percent_0.75<-summary(rq(sat_score ~ ell_percent,data = data,
tau = 0.75))
ell_percent_0.90<-summary(rq(sat_score ~ ell_percent,data = data,
tau = 0.90))
ell_percent_0.95<-summary(rq(sat_score ~ ell_percent,data = data,
tau = 0.95))
ell_percent <- data.frame(cbind(ell_percent_0.05$coefficients[2],
                                ell_percent_0.10$coefficients[2],
                                ell_percent_0.25$coefficients[2],
                                ell_percent_0.50$coefficients[2],
                                ell_percent_0.75$coefficients[2],
                                ell_percent_0.90$coefficients[2],
                                ell_percent_0.95$coefficients[2]))
a_0.05<-summary(rq(sat_score ~ asian_per,data = data, tau =
0.05))
a_0.10<-summary(rq(sat_score ~ asian_per,data = data, tau =
0.10))
a_0.25<-summary(rq(sat_score ~ asian_per,data = data, tau =
0.25))
a_0.50<-summary(rq(sat_score ~ asian_per,data = data, tau =
0.50))
a_0.75<-summary(rq(sat_score ~ asian_per,data = data, tau =
0.75))
a_0.90<-summary(rq(sat_score ~ asian_per,data = data, tau =
0.90))
a_0.95<-summary(rq(sat_score ~ asian_per,data = data, tau =
0.95))
asian_per <- data.frame(cbind(a_0.05$coefficients[2],
                                a_0.10$coefficients[2],
                                a_0.25$coefficients[2],
                                a_0.50$coefficients[2],
                                a_0.75$coefficients[2],
                                a_0.90$coefficients[2],
                                a_0.95$coefficients[2]))
a_0.05<-summary(rq(sat_score ~ black_per,data = data, tau =
0.05))
a_0.10<-summary(rq(sat_score ~ black_per,data = data, tau =
0.10))
a_0.25<-summary(rq(sat_score ~ black_per,data = data, tau =
0.25))
a_0.50<-summary(rq(sat_score ~ black_per,data = data, tau =
0.50))
a_0.75<-summary(rq(sat_score ~ black_per,data = data, tau =
0.75))
a_0.90<-summary(rq(sat_score ~ black_per,data = data, tau =
0.90))
a_0.95<-summary(rq(sat_score ~ black_per,data = data, tau =
0.95))
black_per <- data.frame(cbind(a_0.05$coefficients[2],
                              a_0.10$coefficients[2],
                              a_0.25$coefficients[2],
                              a_0.50$coefficients[2],
                              a_0.75$coefficients[2],
                              a_0.90$coefficients[2],
                              a_0.95$coefficients[2]))
a_0.05<-summary(rq(sat_score ~ white_per,data = data, tau =
0.05))
a_0.10<-summary(rq(sat_score ~ white_per,data = data, tau =
0.10))
a_0.25<-summary(rq(sat_score ~ white_per,data = data, tau =
0.25))
a_0.50<-summary(rq(sat_score ~ white_per,data = data, tau =
0.50))
a_0.75<-summary(rq(sat_score ~ white_per,data = data, tau =
0.75))
a_0.90<-summary(rq(sat_score ~ white_per,data = data, tau =
0.90))
a_0.95<-summary(rq(sat_score ~ white_per,data = data, tau =
0.95))
white_per <- data.frame(cbind(a_0.05$coefficients[2],
                              a_0.10$coefficients[2],
                              a_0.25$coefficients[2],
                              a_0.50$coefficients[2],
                              a_0.75$coefficients[2],
                              a_0.90$coefficients[2],
                              a_0.95$coefficients[2]))
a_0.05<-summary(rq(sat_score ~ hispanic_per,data = data, tau =
0.05))
a_0.10<-summary(rq(sat_score ~ hispanic_per,data = data, tau =
0.10))
a_0.25<-summary(rq(sat_score ~ hispanic_per,data = data, tau =
0.25))
a_0.50<-summary(rq(sat_score ~ hispanic_per,data = data, tau =
0.50))
a_0.75<-summary(rq(sat_score ~ hispanic_per,data = data, tau =
0.75))
a_0.90<-summary(rq(sat_score ~ hispanic_per,data = data, tau =
0.90))
```

```
a_0.95<-summary(rq(sat_score ~ hispanic_per,data = data, tau =
0.95))
hispanic_per <- data.frame(cbind(a_0.05$coefficients[2],
                          a_0.10$coefficients[2],
                          a_0.25$coefficients[2],
                          a_0.50$coefficients[2],
                          a_0.75$coefficients[2],
                          a_0.90$coefficients[2],
                          a_0.95$coefficients[2]))
a_0.05<-summary(rq(sat_score ~ male_per,data = data, tau =
0.05))
a_0.10<-summary(rq(sat_score ~ male_per,data = data, tau =
0.10))
a_0.25<-summary(rq(sat_score ~ male_per,data = data, tau =
0.25))
a_0.50<-summary(rq(sat_score ~ male_per,data = data, tau =
0.50))
a_0.75<-summary(rq(sat_score ~ male_per,data = data, tau =
0.75))
a_0.90<-summary(rq(sat_score ~ male_per,data = data, tau =
0.90))
a_0.95<-summary(rq(sat_score ~ male_per,data = data, tau =
0.95))
male_per <- data.frame(cbind(a_0.05$coefficients[2],
                          a_0.10$coefficients[2],
                          a_0.25$coefficients[2],
                          a_0.50$coefficients[2],
                          a_0.75$coefficients[2],
                          a_0.90$coefficients[2],
                          a_0.95$coefficients[2]))
a_0.05<-summary(rq(sat_score ~ female_per,data = data, tau =
0.05))
a_0.10<-summary(rq(sat_score ~ female_per,data = data, tau =
0.10))
a_0.25<-summary(rq(sat_score ~ female_per,data = data, tau =
0.25))
a_0.50<-summary(rq(sat_score ~ female_per,data = data, tau =
0.50))
a_0.75<-summary(rq(sat_score ~ female_per,data = data, tau =
0.75))
a_0.90<-summary(rq(sat_score ~ female_per,data = data, tau =
0.90))
a_0.95<-summary(rq(sat_score ~ female_per,data = data, tau =
0.95))
female_per <- data.frame(cbind(a_0.05$coefficients[2],
                          a_0.10$coefficients[2],
                          a_0.25$coefficients[2],
                          a_0.50$coefficients[2],
                          a_0.75$coefficients[2],
                          a_0.90$coefficients[2],
                          a_0.95$coefficients[2]))
a_0.05<-summary(rq(sat_score ~ school_dist,data = data, tau =
0.05))
a_0.10<-summary(rq(sat_score ~ school_dist,data = data, tau =
0.10))
a_0.25<-summary(rq(sat_score ~ school_dist,data = data, tau =
0.25))
a_0.50<-summary(rq(sat_score ~ school_dist,data = data, tau =
0.50))
a_0.75<-summary(rq(sat_score ~ school_dist,data = data, tau =
0.75))
a_0.90<-summary(rq(sat_score ~ school_dist,data = data, tau =
0.90))
a_0.95<-summary(rq(sat_score ~ school_dist,data = data, tau =
0.95))
school_dist <- data.frame(cbind(a_0.05$coefficients[2],
                          a_0.10$coefficients[2],
                          a_0.25$coefficients[2],
                          a_0.50$coefficients[2],
                          a_0.75$coefficients[2],
                          a_0.90$coefficients[2],
                          a_0.95$coefficients[2]))
a_0.05<-summary(rq(sat_score ~ saf_s_11,data = data, tau = 0.05))
a_0.10<-summary(rq(sat_score ~ saf_s_11,data = data, tau = 0.10))
a_0.25<-summary(rq(sat_score ~ saf_s_11,data = data, tau = 0.25))
a_0.50<-summary(rq(sat_score ~ saf_s_11,data = data, tau = 0.50))
a_0.75<-summary(rq(sat_score ~ saf_s_11,data = data, tau = 0.75))
a_0.90<-summary(rq(sat_score ~ saf_s_11,data = data, tau = 0.90))
a_0.95<-summary(rq(sat_score ~ saf_s_11,data = data, tau = 0.95))
saf_s_11 <- data.frame(cbind(a_0.05$coefficients[2],
                          a_0.10$coefficients[2],
                          a_0.25$coefficients[2],
                          a_0.50$coefficients[2],
                          a_0.75$coefficients[2],
                          a_0.90$coefficients[2],
                          a_0.95$coefficients[2]))
comparison_quantile_beta <-
rbind(comparison_quantile_beta,avg_class_size,total_enrollment,e
ll_percent,asian_per,
                          black_per,white_per,hispanic_per, male_per,
female_per, school_dist,saf_s_11)
rownames(comparison_quantile_beta) <- c("Number of score
3/4/5","avg_class_size","total_enrollment","ell_percent","asian_pe
r",
```

```
                                "black_per","white_per","hispanic_per",
"male_per", "female_per", "school_dist","saf_s_11")
colnames(comparison_quantile_beta) <- c("tau =
0.05",0.10,0.25,0.50,0.75,0.90,0.95)

ols_num_ex<-lm(sat_score ~
Number.of.Exams.with.scores.3.4.or.5,data = data)
ols_avg_class_size<-lm(sat_score ~
AVERAGE.CLASS.SIZE,data = data)
ols_total_enrollment<-lm(sat_score ~ total_enrollment,data = data)
ols_ell_percent<-lm(sat_score ~ ell_percent,data = data)
ols_asian_per<-lm(sat_score ~ asian_per,data = data)
ols_black_per<-lm(sat_score ~ black_per,data = data)
ols_white_per<-lm(sat_score ~ white_per,data = data)
ols_hispanic_per<-lm(sat_score ~ hispanic_per,data = data)
ols_male_per<-lm(sat_score ~ male_per,data = data)
ols_female_per<-lm(sat_score ~ female_per,data = data)
ols_school_dist<-lm(sat_score ~ school_dist,data = data)
ols_saf_s_11<-lm(sat_score ~ saf_s_11,data = data)
par(mfrow=c(3,4))
plot(seq(0.05,0.95,by = 0.05),rq(sat_score ~
Number.of.Exams.with.scores.3.4.or.5,data = data, tau =
seq(0.05,0.95, by =0.05))$coefficients[2,]
    ,main="[Figure 3] Number.of.Exams.with.scores.3.4.or.5", type
= 'br', xlab = "tau", ylab = "Coefficients")
abline(h=ols_num_ex$coefficient[2],col = 'red')
abline(h = confint(ols_num_ex, level = 0.95)[2], col = 'red', lty =
2)
abline(h = confint(ols_num_ex, level = 0.95)[4], col = 'red', lty =
2)
legend("topleft", legend =  c("mean value", "Confidence Interval"),
lty = 1:2, col = "red")
plot(seq(0.05,0.95,by = 0.05),rq(sat_score ~
AVERAGE.CLASS.SIZE,data = data, tau = seq(0.05,0.95, by
=0.05))$coefficients[2,]
    ,main="[Figure 4] Avg. Class Size", type = 'br', xlab = "tau",
ylab = "Coefficients")
abline(h=ols_avg_class_size$coefficient[2],col = 'red')
abline(h = confint(ols_avg_class_size, level = 0.95)[2], col = 'red',
lty = 2)
abline(h = confint(ols_avg_class_size, level = 0.95)[4], col = 'red',
lty = 2)
legend("topleft", legend =  c("mean value", "Confidence Interval"),
lty = 1:2, col = "red")
plot(seq(0.05,0.95,by = 0.05),rq(sat_score ~ total_enrollment,data
= data, tau = seq(0.05,0.95, by =0.05))$coefficients[2,]
    ,main="[Figure 5] Total Enrollment", type = 'br', xlab = "tau",
ylab = "Coefficients")
abline(h=ols_total_enrollment$coefficient[2],col = 'red')
abline(h = confint(ols_total_enrollment, level = 0.95)[2], col =
'red', lty = 2)
abline(h = confint(ols_total_enrollment, level = 0.95)[4], col =
'red', lty = 2)
legend("topleft", legend =  c("mean value", "Confidence Interval"),
lty = 1:2, col = "red")
plot(seq(0.05,0.95,by = 0.05),rq(sat_score ~ ell_percent,data =
data, tau = seq(0.05,0.95, by =0.05))$coefficients[2,]
    ,main="[Figure 6] English Learner Percentage", type = 'br', xlab
= "tau", ylab = "Coefficients")
abline(h=ols_ell_percent$coefficient[2],col = 'red')
abline(h = confint(ols_ell_percent, level = 0.95)[2], col = 'red', lty
= 2)
abline(h = confint(ols_ell_percent, level = 0.95)[4], col = 'red', lty
= 2)
legend("bottomleft", legend =  c("mean value", "Confidence
Interval"), lty = 1:2, col = "red")
plot(seq(0.05,0.95,by = 0.05),rq(sat_score ~ asian_per,data = data,
tau = seq(0.05,0.95, by =0.05))$coefficients[2,]
    ,main="[Figure 7] Asian Percentage", type = 'br', xlab = "tau",
ylab = "Coefficients")
abline(h=ols_asian_per$coefficient[2],col = 'red')
abline(h = confint(ols_asian_per, level = 0.95)[2], col = 'red', lty =
2)
abline(h = confint(ols_asian_per, level = 0.95)[4], col = 'red', lty =
2)
legend("topleft", legend =  c("mean value", "Confidence Interval"),
lty = 1:2, col = "red")
plot(seq(0.05,0.95,by = 0.05),rq(sat_score ~ white_per,data = data,
tau = seq(0.05,0.95, by =0.05))$coefficients[2,]
    ,main="[Figure 8] White Percentage", type = 'br', xlab = "tau",
ylab = "Coefficients")
abline(h=ols_white_per$coefficient[2],col = 'red')
abline(h = confint(ols_white_per, level = 0.95)[2], col = 'red', lty =
2)
abline(h = confint(ols_white_per, level = 0.95)[4], col = 'red', lty =
2)
legend("topleft", legend =  c("mean value", "Confidence Interval"),
lty = 1:2, col = "red")
plot(seq(0.05,0.95,by = 0.05),rq(sat_score ~ black_per,data = data,
tau = seq(0.05,0.95, by =0.05))$coefficients[2,]
    ,main="[Figure 9] Black Percentage", type = 'br', xlab = "tau",
ylab = "Coefficients")
abline(h=ols_black_per$coefficient[2],col = 'red')
```

```
abline(h = confint(ols_black_per, level = 0.95)[2], col = 'red', lty = 2)
abline(h = confint(ols_black_per, level = 0.95)[4], col = 'red', lty = 2)
legend("bottomleft", legend =  c("mean value", "Confidence Interval"), lty = 1:2, col = "red")
plot(seq(0.05,0.95,by = 0.05),rq(sat_score ~ hispanic_per,data = data, tau = seq(0.05,0.95, by =0.05))$coefficients[2,]
     ,main="[Figure 10] Hispanic Percentage", type = 'br', xlab = "tau", ylab = "Coefficients")
abline(h=ols_hispanic_per$coefficient[2],col = 'red')
abline(h = confint(ols_hispanic_per, level = 0.95)[2], col = 'red', lty = 2)
abline(h = confint(ols_hispanic_per, level = 0.95)[4], col = 'red', lty = 2)
legend("bottomleft", legend =  c("mean value", "Confidence Interval"), lty = 1:2, col = "red")
plot(seq(0.05,0.95,by = 0.05),rq(sat_score ~ male_per,data = data, tau = seq(0.05,0.95, by =0.05))$coefficients[2,]
     ,main="[Figure 11] Male Percentage", type = 'br', xlab = "tau", ylab = "Coefficients")
abline(h=ols_male_per$coefficient[2],col = 'red')
abline(h = confint(ols_male_per, level = 0.95)[2], col = 'red', lty = 2)
abline(h = confint(ols_male_per, level = 0.95)[4], col = 'red', lty = 2)
legend("bottomleft", legend =  c("mean value", "Confidence Interval"), lty = 1:2, col = "red")
plot(seq(0.05,0.95,by = 0.05),rq(sat_score ~ female_per,data = data, tau = seq(0.05,0.95, by =0.05))$coefficients[2,]
     ,main="[Figure 12] Female Percentage", type = 'br', xlab = "tau", ylab = "Coefficients")
abline(h=ols_female_per$coefficient[2],col = 'red')
abline(h = confint(ols_female_per, level = 0.95)[2], col = 'red', lty = 2)
abline(h = confint(ols_female_per, level = 0.95)[4], col = 'red', lty = 2)
legend("topleft", legend =  c("mean value", "Confidence Interval"), lty = 1:2, col = "red")
plot(seq(0.05,0.95,by = 0.05),rq(sat_score ~ school_dist,data = data, tau = seq(0.05,0.95, by =0.05))$coefficients[2,]
     ,main="[Figure 13] School District", type = 'br', xlab = "tau", ylab = "Coefficients")
abline(h=ols_school_dist$coefficient[2],col = 'red')
abline(h = confint(ols_school_dist, level = 0.95)[2], col = 'red', lty = 2)
abline(h = confint(ols_school_dist, level = 0.95)[4], col = 'red', lty = 2)
legend("bottomleft", legend =  c("mean value", "Confidence Interval"), lty = 1:2, col = "red")
plot(seq(0.05,0.95,by = 0.05),rq(sat_score ~ saf_s_11,data = data, tau = seq(0.05,0.95, by =0.05))$coefficients[2,]
     ,main="[Figure 14] Safety Level", type = 'br', xlab = "tau", ylab = "Coefficients")
abline(h=ols_saf_s_11$coefficient[2],col = 'red')
abline(h = confint(ols_saf_s_11, level = 0.95)[2], col = 'red', lty = 2)
abline(h = confint(ols_saf_s_11, level = 0.95)[4], col = 'red', lty = 2)
legend("topleft", legend =  c("mean value", "Confidence Interval"), lty = 1:2, col = "red")
par(mfrow=c(1,1))

q_5_Num_ex<-rq(sat_score ~ Number.of.Exams.with.scores.3.4.or.5, tau = 0.05, data = data)
q_95_Num_ex<-rq(sat_score ~ Number.of.Exams.with.scores.3.4.or.5, tau = 0.95, data = data)
anova(q_5_Num_ex,q_95_Num_ex)

q_5_avg_class_size<-rq(sat_score ~ AVERAGE.CLASS.SIZE, tau = 0.05, data = data)
q_95_avg_class_size<-rq(sat_score ~ AVERAGE.CLASS.SIZE, tau = 0.95, data = data)
anova(q_5_avg_class_size,q_95_avg_class_size)

q_5_total_enrollment<-rq(sat_score ~ total_enrollment, tau = 0.05, data = data)
q_95_total_enrollment<-rq(sat_score ~ total_enrollment, tau = 0.95, data = data)
anova(q_5_total_enrollment,q_95_total_enrollment)
q_5_ell_percent<-rq(sat_score ~ ell_percent, tau = 0.05, data = data)
q_95_ell_percent<-rq(sat_score ~ ell_percent, tau = 0.95, data = data)
anova(q_5_ell_percent,q_95_ell_percent)

q_10_ell_percent<-rq(sat_score ~ ell_percent, tau = 0.10, data = data)
q_90_ell_percent<-rq(sat_score ~ ell_percent, tau = 0.90, data = data)
anova(q_10_ell_percent,q_90_ell_percent)

q_5_asian_per<-rq(sat_score ~ asian_per, tau = 0.05, data = data)
```

```
q_95_asian_per<-rq(sat_score ~ asian_per, tau = 0.95, data = data)
anova(q_5_asian_per,q_95_asian_per)

q_5_black_per<-rq(sat_score ~ black_per, tau = 0.05, data = data)
q_95_black_per<-rq(sat_score ~ black_per, tau = 0.95, data = data)
anova(q_5_black_per,q_95_black_per)

q_5_white_per<-rq(sat_score ~ white_per, tau = 0.05, data = data)
q_95_white_per<-rq(sat_score ~ white_per, tau = 0.95, data =
data)
anova(q_5_white_per,q_95_white_per)

q_5_hispanic_per<-rq(sat_score ~ hispanic_per, tau = 0.05, data =
data)
q_95_hispanic_per<-rq(sat_score ~ hispanic_per, tau = 0.95, data
= data)
anova(q_5_hispanic_per,q_95_hispanic_per)

q_5_male_per<-rq(sat_score ~ male_per, tau = 0.05, data = data)
q_95_male_per<-rq(sat_score ~ male_per, tau = 0.95, data = data)
anova(q_5_male_per,q_95_male_per)

q_5_female_per<-rq(sat_score ~ female_per, tau = 0.05, data =
data)
q_95_female_per<-rq(sat_score ~ female_per, tau = 0.95, data =
data)
anova(q_5_female_per,q_95_female_per)

q_5_school_dist<-rq(sat_score ~ school_dist, tau = 0.05, data =
data)
q_95_school_dist<-rq(sat_score ~ school_dist, tau = 0.95, data =
data)
anova(q_5_school_dist,q_95_school_dist)

q_5_saf_s_11<-rq(sat_score ~ saf_s_11, tau = 0.05, data = data)
q_95_saf_s_11<-rq(sat_score ~ saf_s_11, tau = 0.95, data = data)
anova(q_5_saf_s_11,q_95_saf_s_11)
```