590 HW1

Austin Turvy Derek Holste Mason Carhart

4/13/2021

```
# Load packages
library(pacman)
p_load(readr, data.table, dplyr, janitor, haven, here,
       tidyverse, skimr, lfe, stargazer, quantreg)
```

```
# Load data
af = read_dta(here("afghanistan_anonymized_data.dta"))
```

1. An outlier observation is someone who had more than 50 sheep or goats owned by the household in 2007 and in the fall of 2007 or had more than 10 jeribs of land owned by the household in the fall of 2007 or had more than 20 poeple in the houdcount in fall of 2007 or a person that has number of people in a household greater than 20 and are observed in summer of 2008. Or if number of jerribs of land counted is more than 10 and are observed in summer of 2008. Lastly if the number of sheep and goats is above 50 and they are observed in summer of 2008.

```
# Create nonoutlier column
af <- af %>% mutate(nonoutlier= ifelse((f07_num_ppl_hh_cnt > 20 & f07_observed == 1) |
                                       (f07_jeribs_cnt > 10 & f07_observed == 1)|
                                       (f07_num_sheep_cnt > 50 & f07_observed == 1)|
                                         (s08_num_ppl_hh_cnt > 20 & s08_observed == 1)|
                                         (s08_jeribs_cnt > 10 & s08_observed == 1) |
                                         (s08_num_sheep_cnt > 50 & s08_observed == 1)
                                         ,0,1))


# Tabulating nonoutliers
non_outlier_tab <- af %>% group_by(nonoutlier) %>%
    summarise(count = n(), percentage = n()/nrow(.))
```

```
mytable <- table(af$nonoutlier,af$nonoutlier)
mytable
```

```
##
##        0    1
##   0   76    0
##   1    0 1728
```

There are 76 non-outliers in the data given the specifications prior.

Table 1: Regression Results

	Dependent variable:	
	Girl Enrolled in Formal School Fall 2007	
	(1)	(2)
Treatment	0.517***	0.372***
	(0.091)	(0.094)
chagcharan	0.178**	0.086
	(0.085)	(0.090)
Constant	0.087*	0.328***
	(0.051)	(0.057)
Observations	667	707
R^2	0.335	0.167
Note:		*p<0.1; **p<0.05; ***p<0.01
		S.E Clustered by village

```
# Run girls regression table 4 column 1 including outliers
reg3 = felm(f07_formal_school ~ treatment + chagcharan | 0 | 0 | clustercode,
            data = af %>% filter(f07_girl_cnt == 1 &
                                 f07_test_observed == 1))

# Run boys regression table 4 column 1including outliers
reg4 = felm(f07_formal_school ~ treatment + chagcharan | 0 | 0 | clustercode,
            data = af %>% filter(f07_girl_cnt == 0 &
                                 f07_test_observed == 1))
```

```
# Run girls regression table 4 column 1 with controls
reg5 = felm(f07_formal_school ~ treatment + chagcharan + f07_heads_child_cnt +
              f07_age_cnt + f07_duration_village_cnt +
              f07_farsi_cnt + f07_tajik_cnt + f07_farmer_cnt +
              f07_age_head_cnt + f07_yrs_ed_head_cnt + f07_num_ppl_hh_cnt +
              f07_jeribs_cnt + f07_num_sheep_cnt +
              f07_nearest_scl| 0 | 0 | clustercode,
            data = af %>% filter(f07_girl_cnt == 1 &
                                 nonoutlier == 1 &
                                 f07_test_observed == 1))

# Run girls regression table 4 column 1 with controls
reg6 = felm(f07_formal_school ~ treatment + chagcharan + f07_heads_child_cnt +
              f07_age_cnt + f07_duration_village_cnt +
              f07_farsi_cnt + f07_tajik_cnt + f07_farmer_cnt +
              f07_age_head_cnt + f07_yrs_ed_head_cnt + f07_num_ppl_hh_cnt +
              f07_jeribs_cnt + f07_num_sheep_cnt +
              f07_nearest_scl| 0 | 0 | clustercode,
            data = af %>% filter(f07_girl_cnt == 0 &
                                 nonoutlier == 1 &
                                 f07_test_observed == 1))
```

```
# Create table with regressions with controls
stargazer(reg1, reg2, reg5, keep.stat = c('n', 'rsq'),title = "Regression Results",
          covariate.labels = c('Treatment', "chagcharan"),
          dep.var.labels = c('Girls', 'Boys', 'Girls'),
          omit = c("f07_heads_child_cnt", "f07_girl_cnt", "f07_age_cnt", "f07_duration_village_cnt", "f
          type = 'latex', header = FALSE, float = TRUE,  notes = "S.E Clustered by village")
```

Table 2: Regression Results

	Dependent variable:		
	Girls		
	(1)	(2)	(3)
Treatment	0.517***	0.372***	0.511***
	(0.091)	(0.094)	(0.083)
chagcharan	0.178**	0.086	0.156*
	(0.085)	(0.090)	(0.082)
Constant	0.087*	0.328***	−0.159
	(0.051)	(0.057)	(0.256)
Observations	667	707	667
R^2	0.335	0.167	0.366
Note:	*p<0.1; **p<0.05; ***p<0.01		
	S.E Clustered by village		

Question 3:

```
# girls regression for column 3
reg7 = felm(f07_both_norma_total ~ treatment + chagcharan | 0 | 0 | clustercode,
            data = af %>% filter(f07_girl_cnt == 1 &
                                 nonoutlier == 1 &
                                 f07_test_observed == 1))

# boys regression for column 3
reg8 = felm(f07_both_norma_total ~ treatment + chagcharan | 0 | 0 | clustercode,
            data = af %>% filter(f07_girl_cnt == 0 &
                                 nonoutlier == 1 &
                                 f07_test_observed == 1))
```

Question 4:

```
# girls regression column 4
reg9 = felm(f07_both_norma_total ~ treatment + chagcharan + f07_heads_child_cnt +
```

```
              f07_age_cnt + f07_duration_village_cnt +
              f07_farsi_cnt + f07_tajik_cnt + f07_farmer_cnt +
              f07_age_head_cnt + f07_yrs_ed_head_cnt + f07_num_ppl_hh_cnt +
              f07_jeribs_cnt + f07_num_sheep_cnt +
              f07_nearest_scl| 0 | 0 | clustercode,
            data = af %>% filter(f07_girl_cnt == 1 &
                                 nonoutlier == 1 &
                                 f07_test_observed == 1))

# boys regression column 4
reg10 = felm(f07_both_norma_total ~ treatment + chagcharan + f07_heads_child_cnt +
              f07_age_cnt + f07_duration_village_cnt +
              f07_farsi_cnt + f07_tajik_cnt + f07_farmer_cnt +
              f07_age_head_cnt + f07_yrs_ed_head_cnt + f07_num_ppl_hh_cnt +
              f07_jeribs_cnt + f07_num_sheep_cnt +
              f07_nearest_scl| 0 | 0 | clustercode,
            data = af %>% filter(f07_girl_cnt == 0 &
                                 nonoutlier == 1 &
                                 f07_test_observed == 1))
```

Question 4b:

```
# girls regression for column 6
reg11 = felm(s08_both_norma_total ~ treatment + chagcharan | 0 | 0 | clustercode,
            data = af %>% filter(s08_girls_cnt == 1 &
                                 nonoutlier == 1 &
                                 s08_test_observed == 1))

# boys regression for column 6
reg12 = felm(s08_both_norma_total ~ treatment + chagcharan | 0 | 0 | clustercode,
            data = af %>% filter(s08_girls_cnt == 0 &
                                 nonoutlier == 1 &
                                 s08_test_observed == 1))
# girls regression for column 7
reg13 = felm(s08_both_norma_total ~ treatment + chagcharan +
               s08_heads_child_cnt + s08_age_cnt +
               s08_duration_village_cnt + s08_farsi_cnt + s08_tajik_cnt +
               s08_farmer_cnt + s08_age_head_cnt + s08_yrs_ed_head_cnt +
               s08_num_ppl_hh_cnt + s08_jeribs_cnt + s08_num_sheep_cnt +
               s08_nearest_scl | 0 | 0 | clustercode,
            data = af %>% filter(s08_girls_cnt == 1 &
                                 nonoutlier == 1 &
                                 s08_test_observed == 1))

# boys regression for column 7
reg14 = felm(s08_both_norma_total ~ treatment + chagcharan +
               s08_heads_child_cnt  + s08_age_cnt +
               s08_duration_village_cnt + s08_farsi_cnt + s08_tajik_cnt +
               s08_farmer_cnt + s08_age_head_cnt + s08_yrs_ed_head_cnt +
               s08_num_ppl_hh_cnt + s08_jeribs_cnt + s08_num_sheep_cnt +
               s08_nearest_scl | 0 | 0 | clustercode,
```

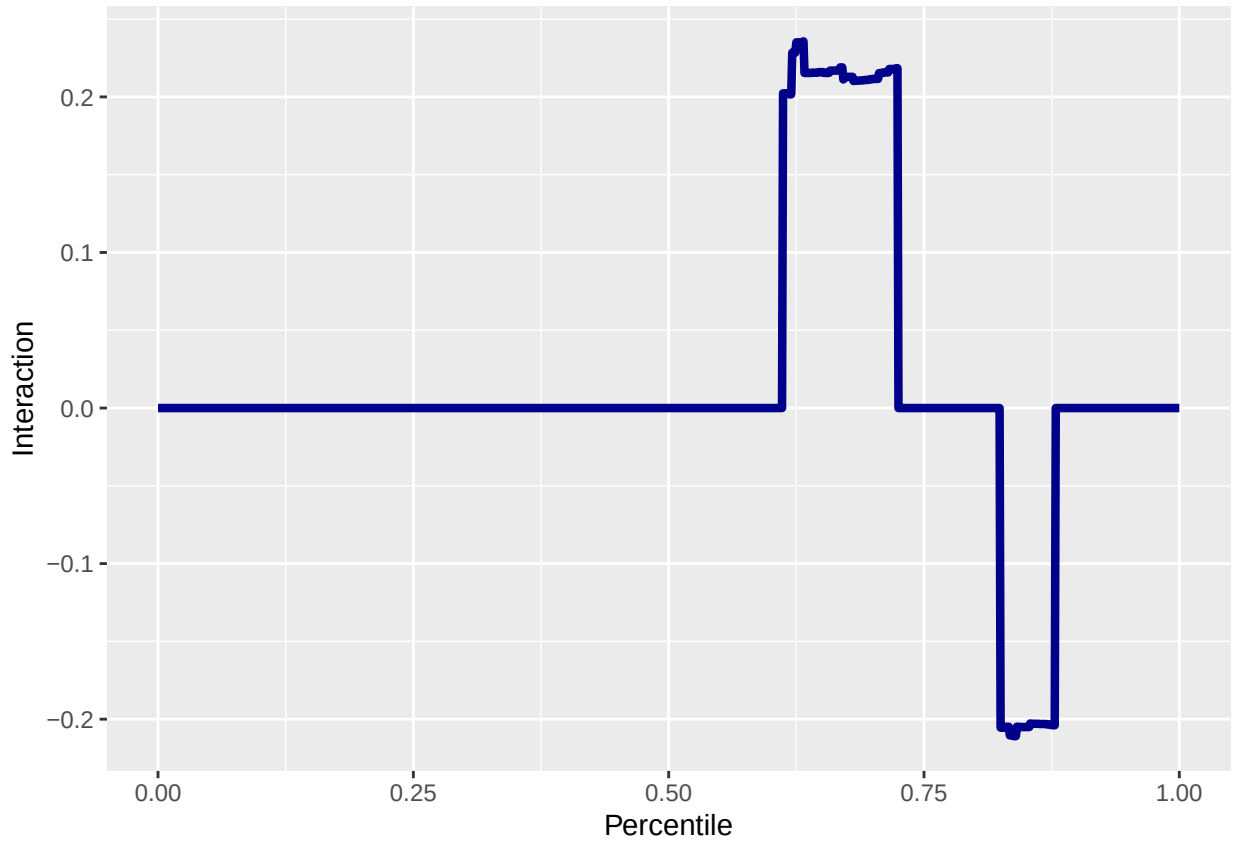
```
            data = af %>% filter(s08_girls_cnt == 0 &
                                 nonoutlier == 1 &
                                 s08_test_observed == 1))
```

```
# Getting mean values for column 1
col_1 = af %>%
  filter(treatment == 1 & nonoutlier == 1 & f07_observed == 1) %>%
  select(
    c(f07_heads_child_cnt, f07_girl_cnt, f07_age_cnt, f07_duration_village_cnt,
                  f07_farsi_cnt,  f07_tajik_cnt,  f07_farmer_cnt,
                  f07_age_head_cnt, f07_yrs_ed_head_cnt, f07_jeribs_cnt,
                  f07_num_sheep_cnt, f07_nearest_scl)
  ) %>%
  mutate(
  across(.cols = c(f07_heads_child_cnt, f07_girl_cnt, f07_age_cnt, f07_duration_village_cnt,
                  f07_farsi_cnt,  f07_tajik_cnt,  f07_farmer_cnt,
                  f07_age_head_cnt, f07_yrs_ed_head_cnt, f07_jeribs_cnt,
                  f07_num_sheep_cnt, f07_nearest_scl), mean, na.rm = TRUE)
  ) %>% head(., 1) %>% unlist(.)

# Getting mean values for column 2
col_2 = af %>%
  filter(treatment == 0 & nonoutlier == 1 & f07_observed == 1) %>%
  select(
    c(f07_heads_child_cnt, f07_girl_cnt, f07_age_cnt, f07_duration_village_cnt,
                  f07_farsi_cnt,  f07_tajik_cnt,  f07_farmer_cnt,
                  f07_age_head_cnt, f07_yrs_ed_head_cnt, f07_jeribs_cnt,
                  f07_num_sheep_cnt, f07_nearest_scl)
  ) %>%
  mutate(
  across(.cols = c(f07_heads_child_cnt, f07_girl_cnt, f07_age_cnt, f07_duration_village_cnt,
                  f07_farsi_cnt,  f07_tajik_cnt,  f07_farmer_cnt,
                  f07_age_head_cnt, f07_yrs_ed_head_cnt, f07_jeribs_cnt,
                  f07_num_sheep_cnt, f07_nearest_scl), mean, na.rm = TRUE)
  ) %>% head(., 1) %>% unlist(.)



# Create regression functions for column 3
reg_function_coef = function(x){
    reg = felm(x ~ treatment | 0 | 0 | clustercode,
               data = af %>% filter(nonoutlier == 1 & f07_observed == 1))
    coef = coef(reg)[2]
    return(coef)
}


# Regression function for SE
reg_function_se = function(x){
  reg = felm(x ~ treatment | 0 | 0 | clustercode,
             data = af %>% filter(nonoutlier == 1 & f07_observed == 1))
  se = sqrt(diag(vcov(reg)))[2]
  return(se)
```

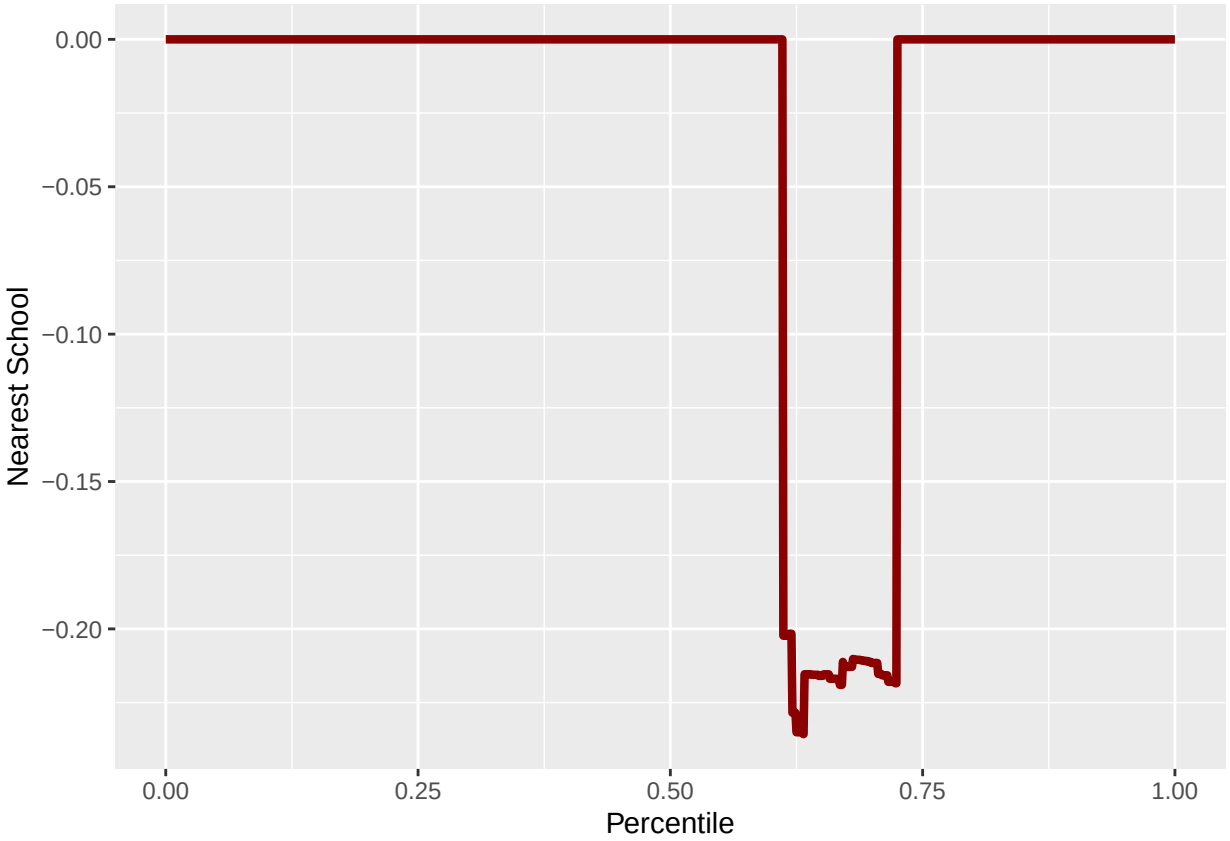
```
}



# Obtain column 3 coefficients
col_3_coefs = af %>%
  filter(nonoutlier == 1 & f07_observed == 1) %>%
  select(
    c(f07_heads_child_cnt, f07_girl_cnt, f07_age_cnt, f07_duration_village_cnt,
                  f07_farsi_cnt,  f07_tajik_cnt,  f07_farmer_cnt,
                  f07_age_head_cnt, f07_yrs_ed_head_cnt, f07_jeribs_cnt,
                  f07_num_sheep_cnt, f07_nearest_scl)
  ) %>%
  mutate(
  across(.cols = c(f07_heads_child_cnt, f07_girl_cnt, f07_age_cnt, f07_duration_village_cnt,
                  f07_farsi_cnt,  f07_tajik_cnt,  f07_farmer_cnt,
                  f07_age_head_cnt, f07_yrs_ed_head_cnt, f07_jeribs_cnt,
                  f07_num_sheep_cnt, f07_nearest_scl), reg_function_coef)
  ) %>% head(., 1) %>% unlist(.)

# Obtain column 3 SE
col_3_se = af %>%
  filter(nonoutlier == 1 & f07_observed == 1) %>%
  select(
    c(f07_heads_child_cnt, f07_girl_cnt, f07_age_cnt, f07_duration_village_cnt,
                  f07_farsi_cnt,  f07_tajik_cnt,  f07_farmer_cnt,
                  f07_age_head_cnt, f07_yrs_ed_head_cnt, f07_jeribs_cnt,
                  f07_num_sheep_cnt, f07_nearest_scl)
  ) %>%
  mutate(
  across(.cols = c(f07_heads_child_cnt, f07_girl_cnt, f07_age_cnt, f07_duration_village_cnt,
                  f07_farsi_cnt,  f07_tajik_cnt,  f07_farmer_cnt,
                  f07_age_head_cnt, f07_yrs_ed_head_cnt, f07_jeribs_cnt,
                  f07_num_sheep_cnt, f07_nearest_scl), reg_function_se)
  ) %>% head(., 1) %>% unlist(.)

# Regression for column 7
col_7 = felm(f07_formal_school ~ f07_heads_child_cnt +
               f07_girl_cnt + f07_age_cnt + f07_duration_village_cnt +
               f07_farsi_cnt + f07_tajik_cnt + f07_farmer_cnt +
               f07_age_head_cnt + f07_yrs_ed_head_cnt + f07_num_ppl_hh_cnt +
               f07_jeribs_cnt + f07_num_sheep_cnt + f07_nearest_scl
             | 0 | 0 | clustercode,
             data = af %>%
               filter(nonoutlier == 1 & f07_observed == 1 & treatment == 0))

# Regression for column 8
col_8 = felm(f07_both_norma_total ~ f07_heads_child_cnt +
               f07_girl_cnt + f07_age_cnt + f07_duration_village_cnt +
               f07_farsi_cnt + f07_tajik_cnt + f07_farmer_cnt +
               f07_age_head_cnt + f07_yrs_ed_head_cnt + f07_num_ppl_hh_cnt +
               f07_jeribs_cnt + f07_num_sheep_cnt + f07_nearest_scl
             | 0 | 0 | clustercode,
```

```
            data = af %>%
              filter(nonoutlier == 1 & f07_observed == 1 &
                       f07_test_observed == 1 & treatment == 0))
```

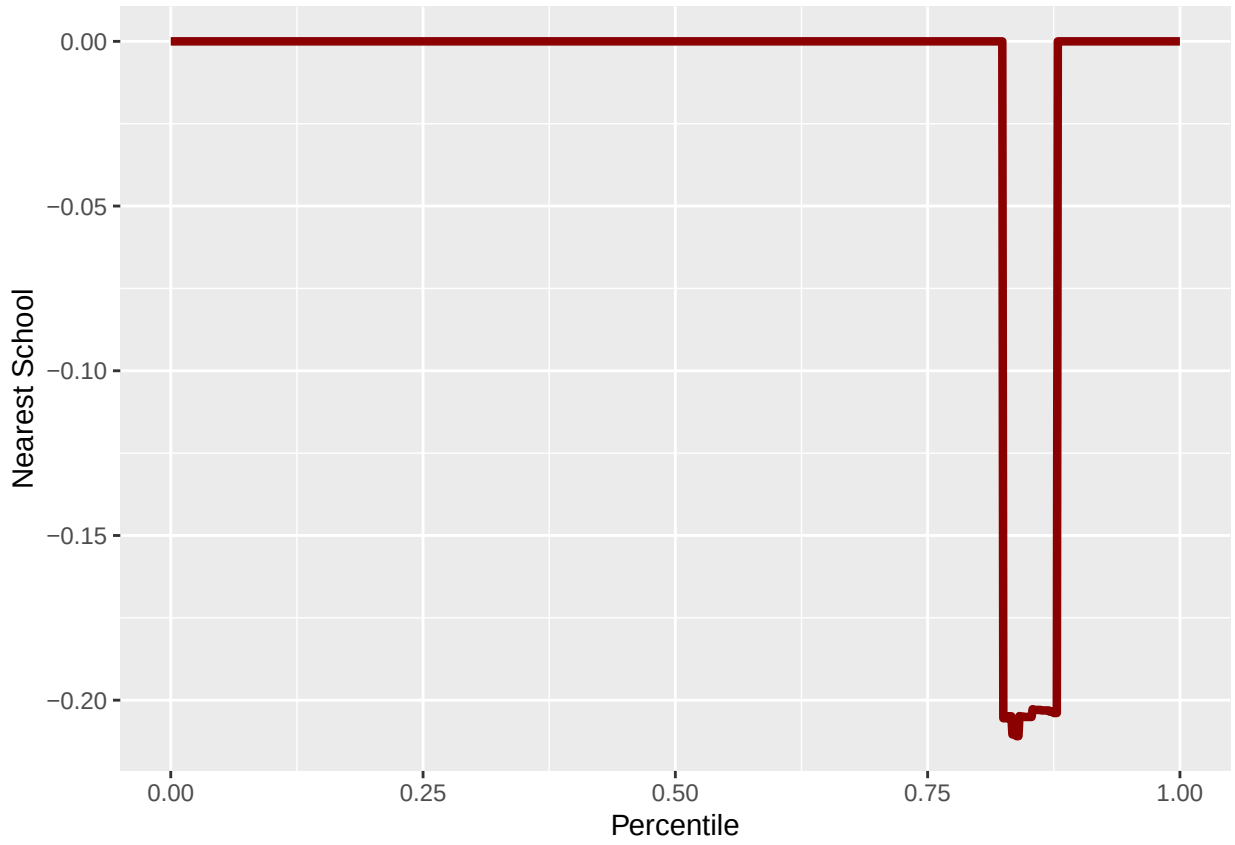
```
# Quant regression function to gather coefficients at different percentiles
quant_coef = function(x, inter = FALSE, is_girl = FALSE){
  nearest_scl = vector()
  interaction = vector()
  grl_count = vector()
  for (i in x) {
          qr <- rq(f07_formal_school ~ f07_nearest_scl*f07_girl_cnt +
                     f07_nearest_scl + f07_girl_cnt, data= af %>%
                     filter(
                       nonoutlier == 1 & f07_observed == 1 & treatment == 0
                       ),
                     tau = i)
          scl = coef(qr)[2]
          grl = coef(qr)[3]
          int = coef(qr)[4]
          nearest_scl = append(nearest_scl, scl)
          grl_count = append(grl_count, grl)
          interaction = append(interaction, int)
  }
  if (inter == TRUE){
    return(interaction)
  } else if (is_girl == TRUE){
    return(grl_count)
  } else{
    return(nearest_scl)
  }
}

# Gathering coefficients
interaction = quant_coef(seq(0, 1, .001), int = TRUE)
nearest_school = quant_coef(seq(0, 1, .001))
grl_count = quant_coef(seq(0, 1, .001), is_girl = TRUE)


# Creating coefficient dataframe
quant_df = tibble(
  scl_girl = interaction,
  nearest_scl = nearest_school,
  grl_count = grl_count,
  tau = seq(0, 1, .001)
) %>% mutate(difference = scl_girl + nearest_scl)


# Plotting coefficients
ggplot(quant_df, aes(x = tau)) +
  geom_line(aes(y = scl_girl), color = "dark blue", size = 1.5) +
  labs(x = "Percentile", y = "Interaction")
```



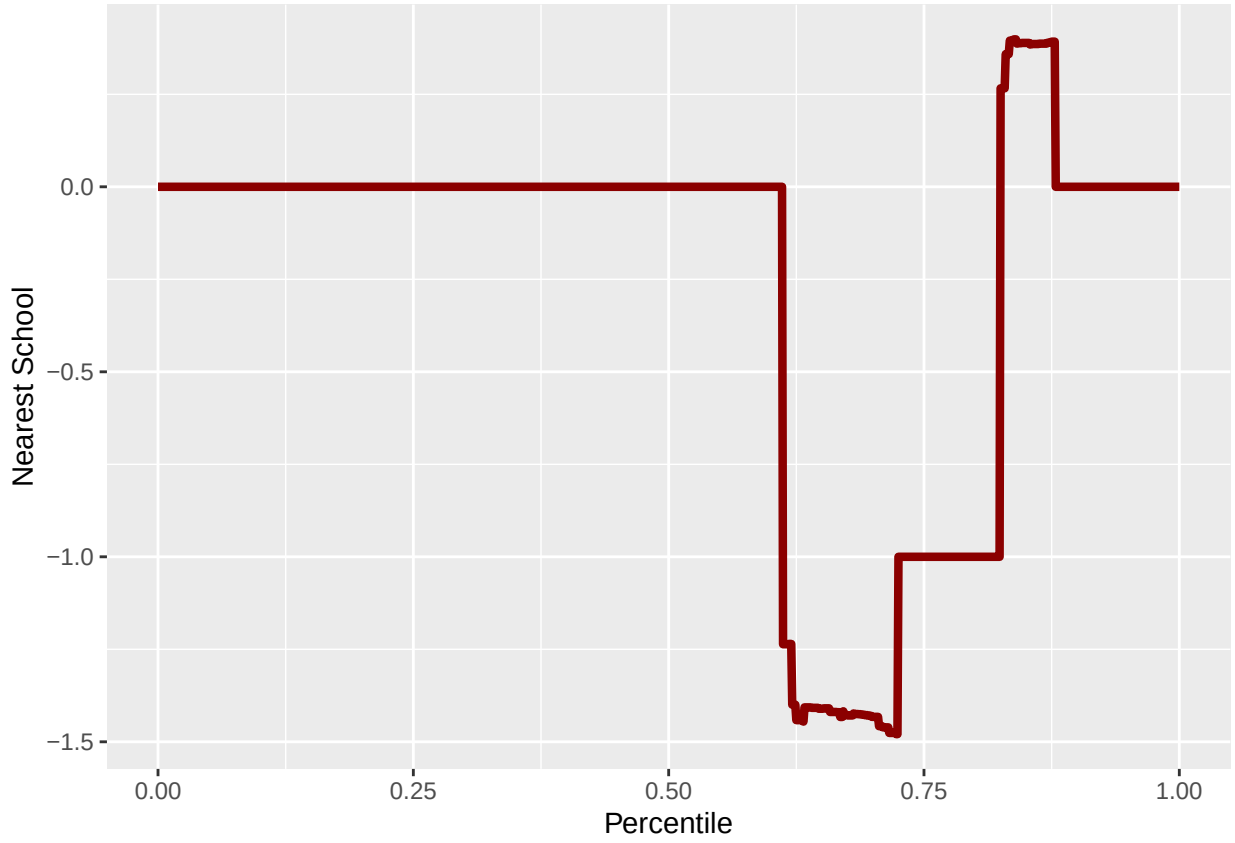

```
ggplot(quant_df, aes(x = tau)) +
  geom_line(aes(y = nearest_school), color = "dark red", size = 1.5) +
  labs(x = "Percentile", y = "Nearest School")
```




```
ggplot(quant_df, aes(x = tau)) +
  geom_line(aes(y = difference), color = "dark red", size = 1.5) +
  labs(x = "Percentile", y = "Nearest School")
```




```
ggplot(quant_df, aes(x = tau)) +
  geom_line(aes(y = grl_count), color = "dark red", size = 1.5) +
  labs(x = "Percentile", y = "Nearest School")
```



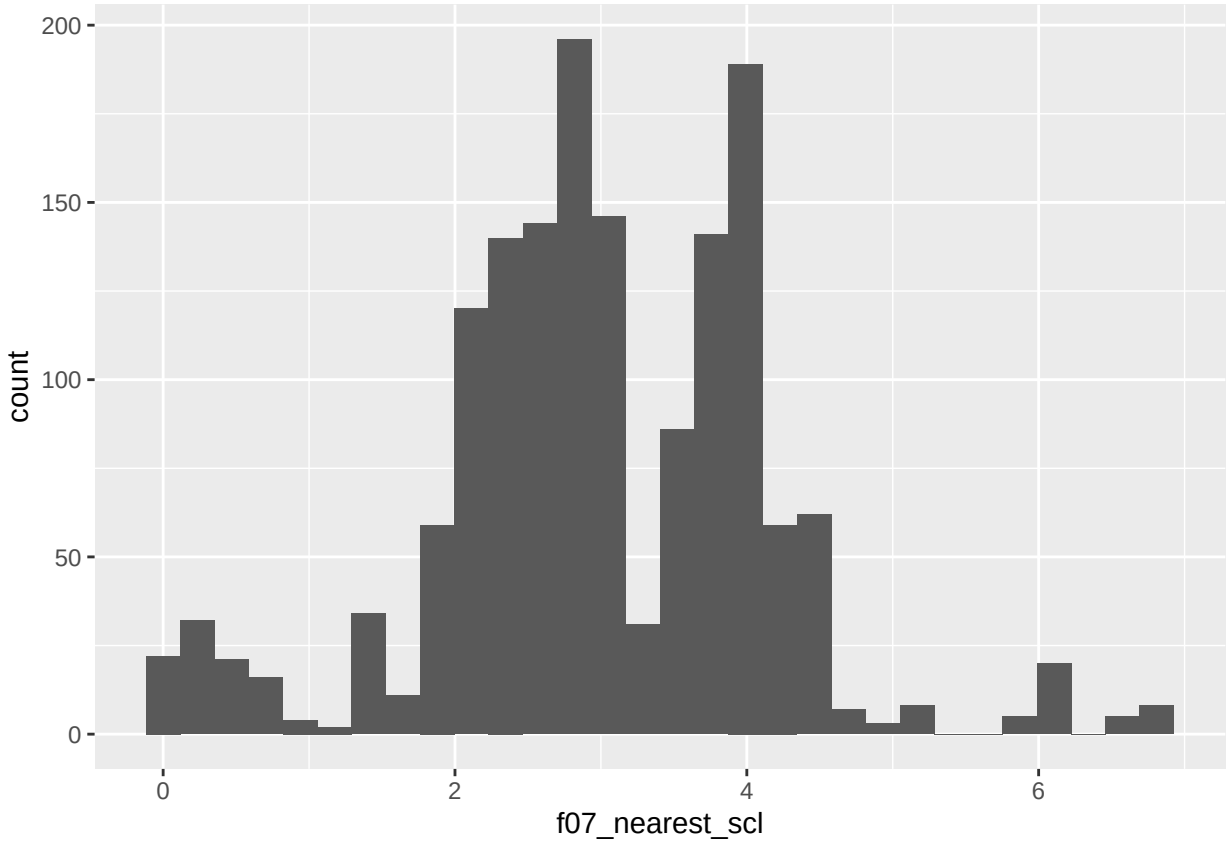

```
summary(af$f07_nearest_scl)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
## 0.01841 2.33871 2.91533 3.00122 3.85273 6.82849     233
```

```
ggplot(af, aes(x=f07_nearest_scl)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 233 rows containing non-finite values (stat_bin).
```




```
age_reg = felm(f07_formal_school ~ f07_age_cnt + f07_age_cnt*f07_nearest_scl +
                 f07_nearest_scl | 0 | 0 | clustercode,
               data = af %>% filter(
                 nonoutlier == 1 & f07_observed == 1 &
                   treatment == 0 & f07_girl_cnt == 1)
               )
```