Preliminary Report

Dylan Denner & Jackson Tucker

4/19/2021

## Preperation for Presentation

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3 v purrr 0.3.4  
## v tibble 3.0.5 v stringr 1.4.0  
## v tidyr 1.1.3 v forcats 0.5.1  
## v readr 1.4.0

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

##   
## Attaching package: 'gridExtra'

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

##   
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':  
##   
## select

## Loading required package: carData

##   
## Attaching package: 'car'

## The following object is masked from 'package:purrr':  
##   
## some

## The following object is masked from 'package:dplyr':  
##   
## recode

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

## Loading required package: splines

## Loading required package: foreach

##   
## Attaching package: 'foreach'

## The following objects are masked from 'package:purrr':  
##   
## accumulate, when

## Loaded gam 1.20

## Loading required package: nlme

##   
## Attaching package: 'nlme'

## The following object is masked from 'package:dplyr':  
##   
## collapse

## This is mgcv 1.8-34. For overview type 'help("mgcv-package")'.

##   
## Attaching package: 'mgcv'

## The following objects are masked from 'package:gam':  
##   
## gam, gam.control, gam.fit, s

library(tidycensus)  
  
poverty\_by\_sex\_edu\_attainment <- get\_acs(  
 geography = "school district (unified)",   
 variables = c(total\_population = "B17003\_001",  
 below\_poverty\_num = "B17003\_002",  
 below\_poverty\_num\_men = "B17003\_003",  
 M\_below\_poverty\_less\_than\_HS = "B17003\_004",  
 M\_below\_poverty\_HS = "B17003\_005",  
 M\_below\_poverty\_some\_college = "B17003\_006",  
 M\_below\_poverty\_bachelor\_plus = "B17003\_007",  
 below\_poverty\_num\_women = "B17003\_008",  
 W\_below\_poverty\_less\_than\_HS = "B17003\_009",  
 W\_below\_poverty\_HS = "B17003\_010",  
 W\_below\_poverty\_some\_college = "B17003\_011",  
 W\_below\_poverty\_bachelor\_plus = "B17003\_012",  
 above\_poverty\_num = "B17003\_013",  
 above\_poverty\_num\_men = "B17003\_014",  
 M\_above\_poverty\_less\_than\_HS = "B17003\_015",  
 M\_above\_poverty\_HS = "B17003\_016",  
 M\_above\_poverty\_some\_college = "B17003\_017",  
 M\_above\_poverty\_bachelor\_plus = "B17003\_018",  
 above\_poverty\_num\_women = "B17003\_019",  
 W\_above\_poverty\_less\_than\_HS = "B17003\_020",  
 W\_above\_poverty\_HS = "B17003\_021",  
 W\_above\_poverty\_some\_college = "B17003\_022",  
 W\_above\_poverty\_bachelor\_plus = "B17003\_023"),  
 state = "OH",   
 year = 2019)

## Getting data from the 2015-2019 5-year ACS

head(poverty\_by\_sex\_edu\_attainment)

## # A tibble: 6 x 5  
## GEOID NAME variable estimate moe  
## <chr> <chr> <chr> <dbl> <dbl>  
## 1 3900094 Monroe Local School District,~ total\_population 9268 352  
## 2 3900094 Monroe Local School District,~ below\_poverty\_num 230 133  
## 3 3900094 Monroe Local School District,~ below\_poverty\_num\_men 60 69  
## 4 3900094 Monroe Local School District,~ M\_below\_poverty\_less\_th~ 10 15  
## 5 3900094 Monroe Local School District,~ M\_below\_poverty\_HS 50 67  
## 6 3900094 Monroe Local School District,~ M\_below\_poverty\_some\_co~ 0 18

dim(poverty\_by\_sex\_edu\_attainment)

## [1] 14099 5

length(unique(poverty\_by\_sex\_edu\_attainment$NAME))

## [1] 613

poverty\_by\_sex\_edu\_attainment[,-c(5)]

## # A tibble: 14,099 x 4  
## GEOID NAME variable estimate  
## <chr> <chr> <chr> <dbl>  
## 1 3900094 Monroe Local School District, O~ total\_population 9268  
## 2 3900094 Monroe Local School District, O~ below\_poverty\_num 230  
## 3 3900094 Monroe Local School District, O~ below\_poverty\_num\_men 60  
## 4 3900094 Monroe Local School District, O~ M\_below\_poverty\_less\_than\_~ 10  
## 5 3900094 Monroe Local School District, O~ M\_below\_poverty\_HS 50  
## 6 3900094 Monroe Local School District, O~ M\_below\_poverty\_some\_colle~ 0  
## 7 3900094 Monroe Local School District, O~ M\_below\_poverty\_bachelor\_p~ 0  
## 8 3900094 Monroe Local School District, O~ below\_poverty\_num\_women 170  
## 9 3900094 Monroe Local School District, O~ W\_below\_poverty\_less\_than\_~ 10  
## 10 3900094 Monroe Local School District, O~ W\_below\_poverty\_HS 120  
## # ... with 14,089 more rows

poverty\_by\_sex\_edu\_attainment <- spread(poverty\_by\_sex\_edu\_attainment, variable, estimate)  
  
head(poverty\_by\_sex\_edu\_attainment)

## # A tibble: 6 x 26  
## GEOID NAME moe above\_poverty\_n~ above\_poverty\_nu~ above\_poverty\_nu~  
## <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 39000~ Monroe Loca~ 15 NA NA NA  
## 2 39000~ Monroe Loca~ 18 NA NA NA  
## 3 39000~ Monroe Loca~ 47 NA NA NA  
## 4 39000~ Monroe Loca~ 67 NA NA NA  
## 5 39000~ Monroe Loca~ 69 NA NA NA  
## 6 39000~ Monroe Loca~ 87 NA NA NA  
## # ... with 20 more variables: below\_poverty\_num <dbl>,  
## # below\_poverty\_num\_men <dbl>, below\_poverty\_num\_women <dbl>,  
## # M\_above\_poverty\_bachelor\_plus <dbl>, M\_above\_poverty\_HS <dbl>,  
## # M\_above\_poverty\_less\_than\_HS <dbl>, M\_above\_poverty\_some\_college <dbl>,  
## # M\_below\_poverty\_bachelor\_plus <dbl>, M\_below\_poverty\_HS <dbl>,  
## # M\_below\_poverty\_less\_than\_HS <dbl>, M\_below\_poverty\_some\_college <dbl>,  
## # total\_population <dbl>, W\_above\_poverty\_bachelor\_plus <dbl>,  
## # W\_above\_poverty\_HS <dbl>, W\_above\_poverty\_less\_than\_HS <dbl>,  
## # W\_above\_poverty\_some\_college <dbl>, W\_below\_poverty\_bachelor\_plus <dbl>,  
## # W\_below\_poverty\_HS <dbl>, W\_below\_poverty\_less\_than\_HS <dbl>,  
## # W\_below\_poverty\_some\_college <dbl>

dim(poverty\_by\_sex\_edu\_attainment)

## [1] 13399 26

mod\_poverty\_by\_sex\_edu\_attainment <- poverty\_by\_sex\_edu\_attainment %>% mutate(  
 "tot\_less\_than\_HS" = W\_below\_poverty\_less\_than\_HS + M\_below\_poverty\_less\_than\_HS +   
 W\_above\_poverty\_less\_than\_HS + M\_above\_poverty\_less\_than\_HS,  
 "tot\_HS" = W\_below\_poverty\_HS + M\_below\_poverty\_HS +   
 W\_above\_poverty\_HS + M\_above\_poverty\_HS,  
 "tot\_some\_college" = W\_below\_poverty\_some\_college + M\_below\_poverty\_some\_college +   
 W\_above\_poverty\_some\_college + M\_above\_poverty\_some\_college,  
 "tot\_bachelor\_plus" = W\_below\_poverty\_bachelor\_plus + M\_below\_poverty\_bachelor\_plus +   
 W\_above\_poverty\_bachelor\_plus + M\_above\_poverty\_bachelor\_plus,  
 "tot\_HS\_PLUS" = tot\_HS + tot\_some\_college + tot\_bachelor\_plus,  
 "poverty\_percentage" = below\_poverty\_num/total\_population,  
 "HS\_PLUS\_percentage" = tot\_HS\_PLUS/total\_population)

mod\_poverty\_by\_sex\_edu\_attainment <- dplyr::select(  
 mod\_poverty\_by\_sex\_edu\_attainment,  
 NAME,  
 total\_population,  
 below\_poverty\_num,  
 above\_poverty\_num,  
 poverty\_percentage,  
 tot\_less\_than\_HS,  
 tot\_HS,  
 tot\_some\_college,  
 tot\_bachelor\_plus,  
 HS\_PLUS\_percentage  
)  
  
dim(mod\_poverty\_by\_sex\_edu\_attainment)

## [1] 13399 10

file\_path <- "../DATA/BUILDING\_OVERVIEW\_1819.xlsx"  
overview\_data <- read\_excel(file\_path, sheet = "BUILDING\_OVERVIEW")  
  
file\_path <- "../DATA/BUILDING\_DISCIPLINE\_1819.xlsx"  
discipline\_data <- read\_excel(file\_path, sheet = "DISCIPLINE")

overview\_data <- dplyr::select(  
 overview\_data,  
 "Building IRN",  
 "Building Name",  
 "District Name",  
 "County",  
 "Region",  
 "Enrollment 2018-2019",  
 "Attendance Rate 2018-2019",  
 "Chronic Absenteeism Percent 2018-2019"  
)  
  
overview\_data[,6:8] <- sapply(overview\_data[,6:8],as.numeric)

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

discipline\_data[ discipline\_data == "<10" ] <- "0"  
discipline\_data[,8:26] <- sapply(discipline\_data[,8:26],as.numeric)

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

discipline\_data <- discipline\_data[,-26]  
discipline\_data <- subset(discipline\_data, discipline\_data$`Discpline Reason Description` == "Disobedient/Disruptive Behavior")

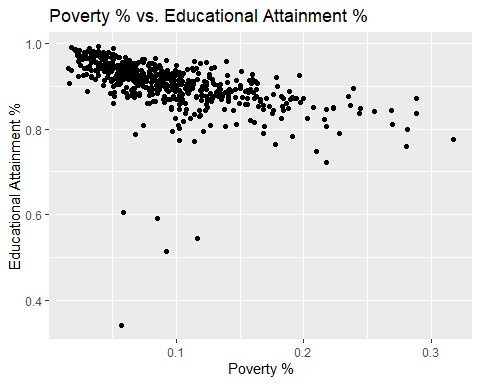
discipline\_data <- dplyr::select(  
 discipline\_data,  
 "Building IRN",  
 "Students Disciplined - Expulsions",  
 "Students Disciplined - Out-of-School Suspensions",  
 "Students Disciplined - In-School Suspensions"  
)

discipline\_data <- discipline\_data %>% mutate(  
 "total\_students\_discipline" = (rowSums(discipline\_data[,2:4])))  
  
joined\_df <- left\_join(overview\_data, discipline\_data, by = "Building IRN")

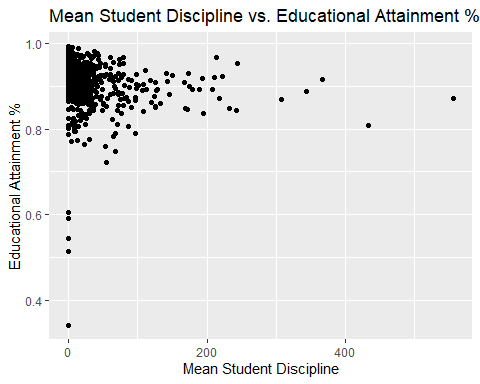
grouped\_df <- joined\_df %>%  
 group\_by(`District Name`, County, Region) %>%  
 summarise(mean\_enrollment = mean(`Enrollment 2018-2019`, na.rm = TRUE),  
 mean\_attendence = mean(`Attendance Rate 2018-2019`, na.rm = TRUE),  
 mean\_chronic\_absenteesim = mean(`Chronic Absenteeism Percent 2018-2019`, na.rm = TRUE),  
 mean\_expulsions = mean(`Students Disciplined - Expulsions`, na.rm = TRUE),  
 mean\_out\_of\_school\_suspensions = mean(`Students Disciplined - Out-of-School Suspensions`, na.rm = TRUE),  
 mean\_in\_school\_suspensions = mean(`Students Disciplined - In-School Suspensions`, na.rm = TRUE),  
 mean\_total\_students\_discipline = mean(`total\_students\_discipline`, na.rm = TRUE))

## `summarise()` has grouped output by 'District Name', 'County'. You can override using the `.groups` argument.

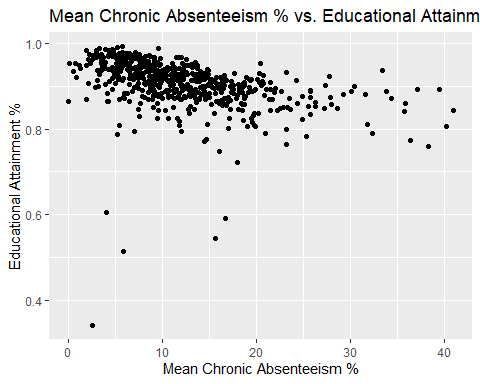
file\_path <- "../DATA/model\_data\_3\_30.csv"  
model\_data <- read.csv(file\_path)  
  
ggplot(model\_data, aes(x=poverty\_percentage, y=HS\_PLUS\_percentage)) + geom\_point() +   
 ggtitle("Poverty % vs. Educational Attainment %") +  
 xlab("Poverty %") +   
 ylab("Educational Attainment %")



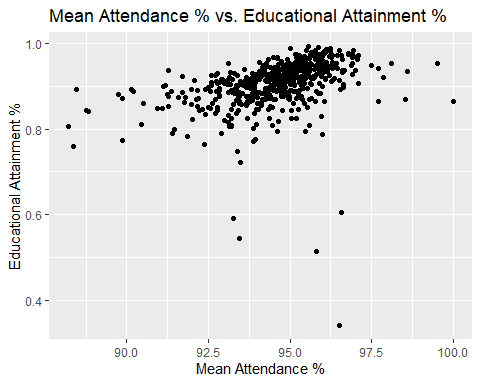
ggplot(model\_data, aes(x=mean\_total\_students\_discipline, y=HS\_PLUS\_percentage)) + geom\_point() +   
 ggtitle("Mean Student Discipline vs. Educational Attainment %") +  
 xlab("Mean Student Discipline") +   
 ylab("Educational Attainment %")



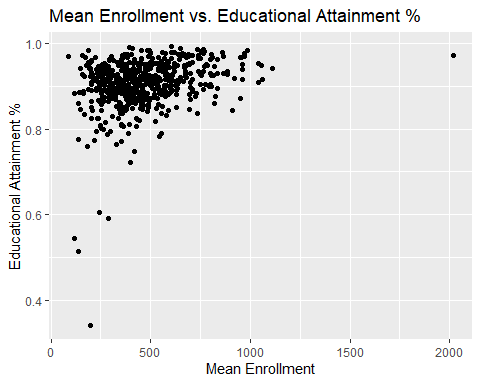
ggplot(model\_data, aes(x=mean\_chronic\_absenteeism, y=HS\_PLUS\_percentage)) + geom\_point() +   
 ggtitle("Mean Chronic Absenteeism % vs. Educational Attainment %") +  
 xlab("Mean Chronic Absenteeism %") +   
 ylab("Educational Attainment %")



ggplot(model\_data, aes(x=mean\_attendance, y=HS\_PLUS\_percentage)) + geom\_point() +   
 ggtitle("Mean Attendance % vs. Educational Attainment %") +  
 xlab("Mean Attendance %") +   
 ylab("Educational Attainment %")



ggplot(model\_data, aes(x=mean\_enrollment, y=HS\_PLUS\_percentage)) + geom\_point() +   
 ggtitle("Mean Enrollment vs. Educational Attainment %") +  
 xlab("Mean Enrollment") +   
 ylab("Educational Attainment %")



POVERTY\_lm <- lm(log(HS\_PLUS\_percentage) ~ poverty\_percentage, data = model\_data)  
DISCIPLINE\_lm <- lm(log(HS\_PLUS\_percentage) ~ mean\_total\_students\_discipline, data = model\_data)  
CHRONIC\_lm <- lm(log(HS\_PLUS\_percentage) ~ mean\_chronic\_absenteeism, data = model\_data)  
ATTENDANCE\_lm <- lm(log(HS\_PLUS\_percentage) ~ mean\_attendance, data = model\_data)  
ENROLLMENT\_lm <- lm(log(HS\_PLUS\_percentage) ~ mean\_enrollment, data = model\_data)  
  
 mod\_model\_data <- subset(model\_data, model\_data$HS\_PLUS\_percentage > 0.65)  
  
POVERTY\_lm\_mod <- lm(log(HS\_PLUS\_percentage) ~ poverty\_percentage, data = mod\_model\_data)  
DISCIPLINE\_lm\_mod <- lm(log(HS\_PLUS\_percentage) ~ mean\_total\_students\_discipline, data = mod\_model\_data)  
CHRONIC\_lm\_mod <- lm(log(HS\_PLUS\_percentage) ~ mean\_chronic\_absenteeism, data = mod\_model\_data)  
ATTENDANCE\_lm\_mod <- lm(log(HS\_PLUS\_percentage) ~ mean\_attendance, data = mod\_model\_data)  
ENROLLMENT\_lm\_mod <- lm(log(HS\_PLUS\_percentage) ~ mean\_enrollment, data = mod\_model\_data)

DISCIPLINE\_sresid <- studres(DISCIPLINE\_lm)  
shapiro.test(DISCIPLINE\_sresid)

##   
## Shapiro-Wilk normality test  
##   
## data: DISCIPLINE\_sresid  
## W = 0.58504, p-value < 2.2e-16

DISCIPLINE\_sresid\_mod <- studres(DISCIPLINE\_lm\_mod)  
shapiro.test(DISCIPLINE\_sresid\_mod)

##   
## Shapiro-Wilk normality test  
##   
## data: DISCIPLINE\_sresid\_mod  
## W = 0.94473, p-value = 5.046e-14

#heterodastic  
ncvTest(POVERTY\_lm)

## Non-constant Variance Score Test   
## Variance formula: ~ fitted.values   
## Chisquare = 9.368701, Df = 1, p = 0.0022072

ncvTest(POVERTY\_lm\_mod)

## Non-constant Variance Score Test   
## Variance formula: ~ fitted.values   
## Chisquare = 41.32826, Df = 1, p = 1.287e-10

#homoscedastic  
ncvTest(DISCIPLINE\_lm)

## Non-constant Variance Score Test   
## Variance formula: ~ fitted.values   
## Chisquare = 29.73787, Df = 1, p = 4.9459e-08

ncvTest(DISCIPLINE\_lm\_mod)

## Non-constant Variance Score Test   
## Variance formula: ~ fitted.values   
## Chisquare = 0.2833977, Df = 1, p = 0.59448

#heteroscedastic  
ncvTest(CHRONIC\_lm)

## Non-constant Variance Score Test   
## Variance formula: ~ fitted.values   
## Chisquare = 74.68988, Df = 1, p = < 2.22e-16

ncvTest(CHRONIC\_lm\_mod)

## Non-constant Variance Score Test   
## Variance formula: ~ fitted.values   
## Chisquare = 11.74019, Df = 1, p = 0.00061165

#homoscedastic  
ncvTest(ATTENDANCE\_lm)

## Non-constant Variance Score Test   
## Variance formula: ~ fitted.values   
## Chisquare = 68.58723, Df = 1, p = < 2.22e-16

ncvTest(ATTENDANCE\_lm\_mod)

## Non-constant Variance Score Test   
## Variance formula: ~ fitted.values   
## Chisquare = 1.00596, Df = 1, p = 0.31587

#heteroscedastic  
ncvTest(ENROLLMENT\_lm)

## Non-constant Variance Score Test   
## Variance formula: ~ fitted.values   
## Chisquare = 208.538, Df = 1, p = < 2.22e-16

ncvTest(ENROLLMENT\_lm\_mod)

## Non-constant Variance Score Test   
## Variance formula: ~ fitted.values   
## Chisquare = 15.26271, Df = 1, p = 9.3545e-05

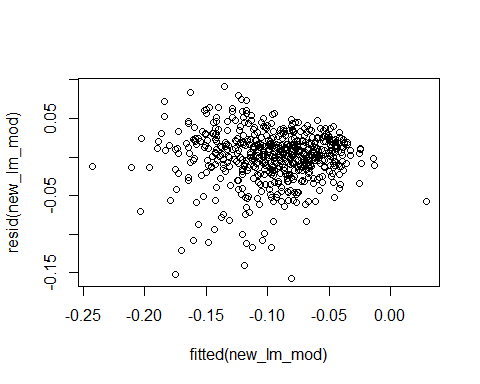
raw\_lm <- lm(log(HS\_PLUS\_percentage) ~ poverty\_percentage\* mean\_chronic\_absenteeism + mean\_enrollment, data = model\_data)  
summary(raw\_lm)

##   
## Call:  
## lm(formula = log(HS\_PLUS\_percentage) ~ poverty\_percentage \* mean\_chronic\_absenteeism +   
## mean\_enrollment, data = model\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.98522 -0.01167 0.00930 0.02412 0.10475   
##   
## Coefficients:  
## Estimate Std. Error t value  
## (Intercept) -6.546e-02 1.264e-02 -5.177  
## poverty\_percentage -6.787e-01 1.137e-01 -5.968  
## mean\_chronic\_absenteeism -1.782e-03 8.838e-04 -2.016  
## mean\_enrollment 8.128e-05 1.389e-05 5.851  
## poverty\_percentage:mean\_chronic\_absenteeism 8.737e-03 5.867e-03 1.489  
## Pr(>|t|)   
## (Intercept) 3.10e-07 \*\*\*  
## poverty\_percentage 4.16e-09 \*\*\*  
## mean\_chronic\_absenteeism 0.0443 \*   
## mean\_enrollment 8.09e-09 \*\*\*  
## poverty\_percentage:mean\_chronic\_absenteeism 0.1370   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.06551 on 590 degrees of freedom  
## Multiple R-squared: 0.2545, Adjusted R-squared: 0.2494   
## F-statistic: 50.35 on 4 and 590 DF, p-value: < 2.2e-16

new\_lm <- lm(log(HS\_PLUS\_percentage) ~ poverty\_percentage\* mean\_chronic\_absenteeism + mean\_enrollment, data = mod\_model\_data)  
  
summary(new\_lm)

##   
## Call:  
## lm(formula = log(HS\_PLUS\_percentage) ~ poverty\_percentage \* mean\_chronic\_absenteeism +   
## mean\_enrollment, data = mod\_model\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.157212 -0.013543 0.004167 0.019618 0.091506   
##   
## Coefficients:  
## Estimate Std. Error t value  
## (Intercept) -3.505e-02 6.587e-03 -5.321  
## poverty\_percentage -7.287e-01 5.885e-02 -12.382  
## mean\_chronic\_absenteeism -2.534e-03 4.592e-04 -5.519  
## mean\_enrollment 4.703e-05 7.238e-06 6.498  
## poverty\_percentage:mean\_chronic\_absenteeism 1.161e-02 3.039e-03 3.819  
## Pr(>|t|)   
## (Intercept) 1.47e-07 \*\*\*  
## poverty\_percentage < 2e-16 \*\*\*  
## mean\_chronic\_absenteeism 5.14e-08 \*\*\*  
## mean\_enrollment 1.74e-10 \*\*\*  
## poverty\_percentage:mean\_chronic\_absenteeism 0.000148 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.03386 on 585 degrees of freedom  
## Multiple R-squared: 0.5557, Adjusted R-squared: 0.5526   
## F-statistic: 182.9 on 4 and 585 DF, p-value: < 2.2e-16

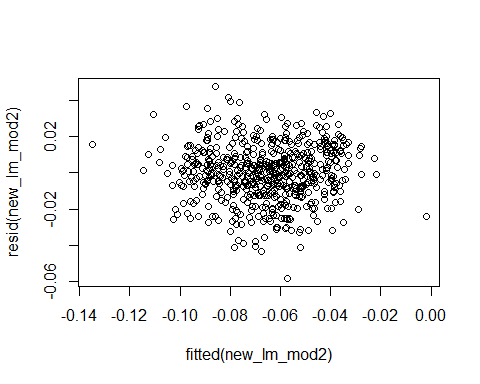
new\_lm\_mod <- lm(log(HS\_PLUS\_percentage) ~ poverty\_percentage\* mean\_chronic\_absenteeism + mean\_enrollment, data = mod\_model\_data)  
  
plot(fitted(new\_lm\_mod), resid(new\_lm\_mod))



new\_lm\_mod2 <- lm((((HS\_PLUS\_percentage ^ 7) - 1) / 7) ~ poverty\_percentage\* mean\_chronic\_absenteeism + mean\_enrollment, data = mod\_model\_data)  
summary(new\_lm\_mod2)

##   
## Call:  
## lm(formula = (((HS\_PLUS\_percentage^7) - 1)/7) ~ poverty\_percentage \*   
## mean\_chronic\_absenteeism + mean\_enrollment, data = mod\_model\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.058550 -0.008652 0.000090 0.010474 0.048104   
##   
## Coefficients:  
## Estimate Std. Error t value  
## (Intercept) -3.030e-02 3.015e-03 -10.050  
## poverty\_percentage -4.029e-01 2.694e-02 -14.958  
## mean\_chronic\_absenteeism -1.891e-03 2.102e-04 -8.998  
## mean\_enrollment 2.390e-05 3.313e-06 7.214  
## poverty\_percentage:mean\_chronic\_absenteeism 1.022e-02 1.391e-03 7.343  
## Pr(>|t|)   
## (Intercept) < 2e-16 \*\*\*  
## poverty\_percentage < 2e-16 \*\*\*  
## mean\_chronic\_absenteeism < 2e-16 \*\*\*  
## mean\_enrollment 1.69e-12 \*\*\*  
## poverty\_percentage:mean\_chronic\_absenteeism 7.06e-13 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.0155 on 585 degrees of freedom  
## Multiple R-squared: 0.5914, Adjusted R-squared: 0.5886   
## F-statistic: 211.7 on 4 and 585 DF, p-value: < 2.2e-16

plot(fitted(new\_lm\_mod2), resid(new\_lm\_mod2))



new\_sresid\_mod2 <- studres(new\_lm\_mod2)  
new\_sresid\_mod <- studres(new\_lm\_mod)  
shapiro.test(new\_sresid\_mod2)

##   
## Shapiro-Wilk normality test  
##   
## data: new\_sresid\_mod2  
## W = 0.99568, p-value = 0.1021

shapiro.test(new\_sresid\_mod)

##   
## Shapiro-Wilk normality test  
##   
## data: new\_sresid\_mod  
## W = 0.9374, p-value = 4.921e-15

attach(mod\_model\_data)  
smooth\_spline\_poverty\_cv <- smooth.spline(poverty\_percentage, HS\_PLUS\_percentage, cv=TRUE)  
  
smooth\_spline\_absenteeism\_cv <- smooth.spline(mean\_chronic\_absenteeism, HS\_PLUS\_percentage, cv=TRUE)  
  
smooth\_spline\_attendance\_cv <- smooth.spline(mean\_attendance, HS\_PLUS\_percentage, cv=TRUE)  
  
smooth\_spline\_enrollment\_cv <- smooth.spline(mean\_enrollment, HS\_PLUS\_percentage, cv=TRUE)  
  
smooth\_spline\_discipline\_cv <- smooth.spline(mean\_total\_students\_discipline, HS\_PLUS\_percentage, cv=TRUE)  
  
smooth\_spline\_poverty\_cv$df

## [1] 12.39204

smooth\_spline\_absenteeism\_cv$df

## [1] 4.644571

smooth\_spline\_attendance\_cv$df

## [1] 6.114501

smooth\_spline\_enrollment\_cv$df

## [1] 4.820783

smooth\_spline\_discipline\_cv$df

## [1] 4.455439

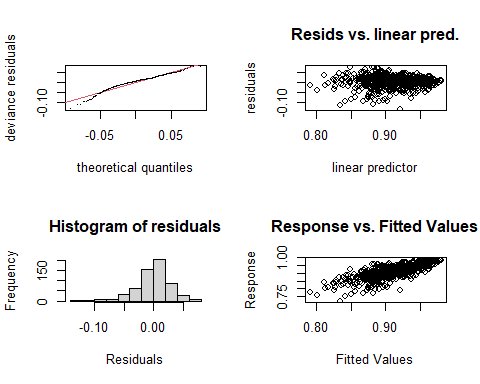
gam\_model <- gam(HS\_PLUS\_percentage ~ s(poverty\_percentage) + s(mean\_chronic\_absenteeism) + s(mean\_attendance) + s(mean\_enrollment) + s(mean\_total\_students\_discipline), data = mod\_model\_data)  
summary(gam\_model)

##   
## Family: gaussian   
## Link function: identity   
##   
## Formula:  
## HS\_PLUS\_percentage ~ s(poverty\_percentage) + s(mean\_chronic\_absenteeism) +   
## s(mean\_attendance) + s(mean\_enrollment) + s(mean\_total\_students\_discipline)  
##   
## Parametric coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.9098 0.0012 757.9 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Approximate significance of smooth terms:  
## edf Ref.df F p-value   
## s(poverty\_percentage) 6.291 7.451 30.421 <2e-16 \*\*\*  
## s(mean\_chronic\_absenteeism) 2.265 2.813 2.036 0.114   
## s(mean\_attendance) 2.603 3.323 1.120 0.513   
## s(mean\_enrollment) 4.155 5.071 8.882 <2e-16 \*\*\*  
## s(mean\_total\_students\_discipline) 1.457 1.786 0.301 0.744   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## R-sq.(adj) = 0.58 Deviance explained = 59.2%  
## GCV = 0.00087666 Scale est. = 0.00085026 n = 590

gam\_model$coefficients

## (Intercept) s(poverty\_percentage).1   
## 9.098155e-01 9.446393e-03   
## s(poverty\_percentage).2 s(poverty\_percentage).3   
## -1.138911e-02 -7.940860e-03   
## s(poverty\_percentage).4 s(poverty\_percentage).5   
## 1.576313e-02 3.947602e-03   
## s(poverty\_percentage).6 s(poverty\_percentage).7   
## -1.112359e-02 -1.206357e-04   
## s(poverty\_percentage).8 s(poverty\_percentage).9   
## 3.644813e-02 -3.172598e-02   
## s(mean\_chronic\_absenteeism).1 s(mean\_chronic\_absenteeism).2   
## 2.037495e-03 -1.281603e-03   
## s(mean\_chronic\_absenteeism).3 s(mean\_chronic\_absenteeism).4   
## -6.602388e-04 6.510306e-04   
## s(mean\_chronic\_absenteeism).5 s(mean\_chronic\_absenteeism).6   
## -7.261620e-05 -8.057436e-04   
## s(mean\_chronic\_absenteeism).7 s(mean\_chronic\_absenteeism).8   
## 3.435326e-04 -5.495997e-03   
## s(mean\_chronic\_absenteeism).9 s(mean\_attendance).1   
## -1.243316e-02 3.834296e-03   
## s(mean\_attendance).2 s(mean\_attendance).3   
## -3.804934e-03 3.885479e-05   
## s(mean\_attendance).4 s(mean\_attendance).5   
## 1.606776e-03 4.975597e-04   
## s(mean\_attendance).6 s(mean\_attendance).7   
## 1.604821e-03 2.999232e-04   
## s(mean\_attendance).8 s(mean\_attendance).9   
## 1.866085e-02 -7.320744e-03   
## s(mean\_enrollment).1 s(mean\_enrollment).2   
## -7.448337e-04 8.808405e-03   
## s(mean\_enrollment).3 s(mean\_enrollment).4   
## -2.372885e-03 1.787789e-03   
## s(mean\_enrollment).5 s(mean\_enrollment).6   
## 8.513674e-04 7.144315e-04   
## s(mean\_enrollment).7 s(mean\_enrollment).8   
## 3.675601e-04 1.062097e-03   
## s(mean\_enrollment).9 s(mean\_total\_students\_discipline).1   
## 7.686718e-04 4.129903e-04   
## s(mean\_total\_students\_discipline).2 s(mean\_total\_students\_discipline).3   
## -2.175084e-03 1.053924e-03   
## s(mean\_total\_students\_discipline).4 s(mean\_total\_students\_discipline).5   
## -1.526401e-03 -7.071597e-04   
## s(mean\_total\_students\_discipline).6 s(mean\_total\_students\_discipline).7   
## 1.383736e-03 -1.304242e-03   
## s(mean\_total\_students\_discipline).8 s(mean\_total\_students\_discipline).9   
## -5.433946e-03 3.190974e-04

gam.check(gam\_model)



##   
## Method: GCV Optimizer: magic  
## Smoothing parameter selection converged after 5 iterations.  
## The RMS GCV score gradient at convergence was 8.983539e-08 .  
## The Hessian was not positive definite.  
## Model rank = 46 / 46   
##   
## Basis dimension (k) checking results. Low p-value (k-index<1) may  
## indicate that k is too low, especially if edf is close to k'.  
##   
## k' edf k-index p-value  
## s(poverty\_percentage) 9.00 6.29 1.00 0.44  
## s(mean\_chronic\_absenteeism) 9.00 2.27 0.98 0.34  
## s(mean\_attendance) 9.00 2.60 1.02 0.61  
## s(mean\_enrollment) 9.00 4.15 1.02 0.74  
## s(mean\_total\_students\_discipline) 9.00 1.46 1.03 0.68

plot(gam\_model, se=TRUE, col="blue")

