Capstone Final

Yusei Hara

2020/11/14

Table of Contents

[ASER Pakistan 2016 2](#_Toc57024361)

[Preparation 2](#_Toc57024362)

[Data Installation 3](#_Toc57024363)

[Data Wrangling 4](#_Toc57024364)

[Preparation for Logistic Regression Analysis 8](#_Toc57024365)

[Notes on Data Wrangling 11](#_Toc57024366)

[Hunza. Generalized Linear Models 17](#_Toc57024367)

[Formula 1 17](#_Toc57024368)

[Model 1 [NofChild + HHOwnership] 17](#_Toc57024369)

[Model 1 [NofChild + HHOwnership] 23](#_Toc57024370)

[Model 2 [NofChild + ParentsBothEdu] 33](#_Toc57024371)

[Model 3 [NofChild + ParentsEitherEdu] 40](#_Toc57024372)

[Pak. Generalized Linear Model 44](#_Toc57024373)

[Model 1 [NofChild + HHOwnership] 44](#_Toc57024374)

[Model 2 [NofChild + ParentsBothEdu] 60](#_Toc57024375)

[Appendix 78](#_Toc57024376)

[Parents.Edu1 & Num.of.Child 78](#_Toc57024377)

[Parents.Edu2 & Num.of.Child 79](#_Toc57024378)

## ASER Pakistan 2016

In this piece of paper, a set of data obtained from Annual Status of Education Report (ASER) is explored. The raw data was downloaded from the link here. <https://palnetwork.org/aser-centre/>

### Preparation

#### Packages Used

library(tidyverse)  
library(ggplot2)  
library(ggthemes)  
library(ggrepel)  
library(gghighlight)  
library(stringr)  
library(dplyr)  
library(sf)  
library(scatterplot3d)  
library(car)  
library(ResourceSelection) # to excute Hosmer-Lemeshow test

## Warning: package 'ResourceSelection' was built under R version 4.0.3

library(equatiomatic) # to convert model to equation

## Warning: package 'equatiomatic' was built under R version 4.0.3

library(caret)  
require(ggiraph)

## Warning: package 'ggiraph' was built under R version 4.0.3

require(ggiraphExtra)

### Data Installation

#### ASER2016

provdist <- read.csv("aser/ASER2016ProvDist.csv")  
school <- read.csv("aser/ASER2016GSchool.csv")  
child <- read.csv("aser/ASER2016Child.csv")  
pschool <- read.csv("aser/ASER2016PvtSchool.csv")  
gschool <- read.csv("aser/ASER2016GSchool.csv")  
parent <- read.csv("aser/ASER2016Parent.csv")  
house <- read.csv("aser/ASER2016HouseholdIndicators.csv")

RegionName <- c("2" = "Panjab",   
 "3" = "Sindh",   
 "4" = "Balochistan",   
 "5" = "Khyber Pakhtunkhwa",   
 "6" = "Gilgit-Baltistan",   
 "7" = "Azad Jammu and Kashmir",   
 "8" = "Islamabad - ICT",   
 "9" = "Federally Administrated Tribal Areas")  
Gender <- c("0" = "Male",  
 "-1" = "Female")

#### Spatial Data

ica <- sf::st\_read("map/pak\_ica\_categories\_areas\_geonode\_apr2017.shp")

## Reading layer `pak\_ica\_categories\_areas\_geonode\_apr2017' from data source `C:\Program Files\R\capstone\map\pak\_ica\_categories\_areas\_geonode\_apr2017.shp' using driver `ESRI Shapefile'  
## Simple feature collection with 156 features and 8 fields  
## geometry type: MULTIPOLYGON  
## dimension: XY  
## bbox: xmin: 60.8786 ymin: 23.69468 xmax: 77.83397 ymax: 37.08942  
## geographic CRS: WGS 84

### Data Wrangling

ica\_df <- ica %>%   
 mutate(centroid = st\_centroid(geometry),  
 x = st\_coordinates(centroid)[,1],  
 y = st\_coordinates(centroid)[,2]) %>%   
 as.data.frame()

## Warning: Problem with `mutate()` input `centroid`.  
## x st\_centroid does not give correct centroids for longitude/latitude data  
## i Input `centroid` is `st\_centroid(geometry)`.

## Warning in st\_centroid.sfc(geometry): st\_centroid does not give correct  
## centroids for longitude/latitude data

ica\_df <- ica\_df %>% select(Province, Districts, x, y)  
ica\_df <- ica\_df %>% summarize(Province = tolower(Province), Districts = tolower(Districts), x = x, y = y)

#### Child and ProvDist data combination

child\_dname <- child %>% left\_join(provdist[-1])

## Joining, by = "DID"

child\_dname <- child\_dname %>% mutate(dname = tolower(DNAME))

ica\_df\_3 <- ica\_df %>% filter(Province == "sindh")  
  
ica\_df\_3$Districts <- ica\_df\_3$Districts %>%   
 str\_replace("ghotki", "gotki") %>%  
 str\_replace("mirpur khas", "mirpurkhas") %>%   
 str\_replace("malir karachi", "karachi-malir-rural") %>%   
 str\_replace("naushahro feroze", "nowshero feroze") %>%   
 str\_replace("kambar shahdad kot", "qambar shahdadkot") %>%   
 str\_replace("sujawal", "sajawal") %>%   
 str\_replace("shaheed benazir abad", "shaheed benazirabad") %>%   
 str\_replace("tando allahyar", "tando allah yar") %>%   
 as.vector()  
  
child\_dname\_3 <- child\_dname %>% filter(RNAME == "Sindh") %>% left\_join(ica\_df\_3, by = c("dname" = "Districts"))  
  
child\_dname\_3 %>% group\_by(dname) %>% summarize(n = sum(x))

## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 25 x 2  
## dname n  
## <chr> <dbl>  
## 1 badin 132177.  
## 2 dadu 130944.  
## 3 gotki 138955.  
## 4 hyderabad 126370.  
## 5 jacobabad 113124.  
## 6 jamshoro 95520.  
## 7 karachi-malir-rural 75832.  
## 8 karachi-west-rural NA   
## 9 kashmore 132646.  
## 10 khairpur 107834.  
## # ... with 15 more rows

ica\_df\_3

## Province Districts x y  
## 1 sindh badin 68.84219 24.72277  
## 2 sindh central karachi 67.05813 24.94777  
## 3 sindh dadu 67.49687 26.86763  
## 4 sindh east karachi 67.13429 24.93583  
## 5 sindh gotki 69.65163 27.82827  
## 6 sindh hyderabad 68.45633 25.34110  
## 7 sindh jacobabad 68.47705 28.20721  
## 8 sindh jamshoro 67.79244 25.73650  
## 9 sindh qambar shahdadkot 67.71121 27.63562  
## 10 sindh kashmore 69.23065 28.27468  
## 11 sindh khairpur 69.08038 26.82907  
## 12 sindh korangi karachi 67.14997 24.84413  
## 13 sindh larkana 68.19492 27.52356  
## 14 sindh karachi-malir-rural 67.28693 25.11444  
## 15 sindh matiari 68.45326 25.76496  
## 16 sindh mirpurkhas 69.16431 25.34044  
## 17 sindh nowshero feroze 68.12521 26.87884  
## 18 sindh sanghar 69.29576 25.95919  
## 19 sindh shaheed benazirabad 68.33324 26.36101  
## 20 sindh shikarpur 68.60669 27.94883  
## 21 sindh south karachi 66.87388 24.89373  
## 22 sindh sajawal 68.15257 24.24510  
## 23 sindh sukkur 69.17648 27.50249  
## 24 sindh tando allah yar 68.77546 25.46575  
## 25 sindh tando muhammad khan 68.50212 25.00184  
## 26 sindh tharparkar 70.17772 24.78158  
## 27 sindh thatta 67.76193 24.73303  
## 28 sindh umer kot 69.77834 25.38386  
## 29 sindh west karachi 67.01128 24.99852

remove(ica\_df\_2, ica\_df\_3, ica\_df\_4,  
 ica\_df\_5, ica\_df\_6, ica\_df\_7,  
 ica\_df\_8, ica\_df\_9)

remove(child\_dname\_2, child\_dname\_3, child\_dname\_4,  
 child\_dname\_5, child\_dname\_6, child\_dname\_7,  
 child\_dname\_8, child\_dname\_9)

### Preparation for Logistic Regression Analysis

#### Making Dataframe With Dummy Variables

child\_ica\_dummy <- child\_ica %>% filter(!is.na(C002), !is.na(C003), !is.na(PR004), !is.na(PR009), !is.na(H003))  
  
child\_ica\_dummy$Age <- child\_ica\_dummy$C001  
child\_ica\_dummy$Gender <- ifelse(child\_ica\_dummy$C002 == -1, 1, 0)  
child\_ica\_dummy$School.Enrollment <- ifelse(child\_ica\_dummy$C003 == 3, 1, 0) # currently-enrolled  
child\_ica\_dummy$C003\_1\_01 <- ifelse(child\_ica\_dummy$C003 == 1, 1, 0) # never-enrolled  
child\_ica\_dummy$C003\_2\_01 <- ifelse(child\_ica\_dummy$C003 == 2, 1, 0) # drop-out  
child\_ica\_dummy$PR004\_01 <- ifelse(child\_ica\_dummy$PR004 == -1, 1, 0)  
child\_ica\_dummy$PR009\_01 <- ifelse(child\_ica\_dummy$PR009 == -1, 1, 0)  
child\_ica\_dummy$Parents.Edu1 <- ifelse(  
 child\_ica\_dummy$PR004 == -1 | child\_ica\_dummy$PR009 == -1, 1, 0)  
  
  
child\_ica\_dummy$PR004\_only\_01 <- ifelse(  
 child\_ica\_dummy$PR004 == -1 & child\_ica\_dummy$PR009 == 0, 1, 0)  
child\_ica\_dummy$PR009\_only\_01 <- ifelse(  
 child\_ica\_dummy$PR004 == 0 & child\_ica\_dummy$PR009 == -1, 1, 0)  
child\_ica\_dummy$Parents.Edu2 <- ifelse(  
 child\_ica\_dummy$PR004 == -1 & child\_ica\_dummy$PR009 == -1, 1, 0)  
  
child\_ica\_dummy$Household.Ownership <- ifelse(  
 child\_ica\_dummy$H003 == -1, 1, 0)

Num.of.Child <- child\_ica\_dummy %>%   
 group\_by(HHID) %>%   
 summarize(Num.of.Child = length(unique(CID)))

## `summarise()` ungrouping output (override with `.groups` argument)

child\_ica\_dummy <- child\_ica\_dummy %>% left\_join(Num.of.Child)

## Joining, by = "HHID"

child\_ica\_dummy$H002\_1\_01 <- ifelse(child\_ica\_dummy$H002 == 1, 1, 0)  
child\_ica\_dummy$H002\_2\_01 <- ifelse(child\_ica\_dummy$H002 == 2, 1, 0)  
child\_ica\_dummy$H002\_3\_01 <- ifelse(child\_ica\_dummy$H002 == 3, 1, 0)

#### Regional Dummy Variables

child\_ica\_dummy <- child\_ica\_dummy  
child\_ica\_dummy$Panjab <- ifelse(child\_ica\_dummy$RID == 2, 1, 0)  
child\_ica\_dummy$Sindh <- ifelse(child\_ica\_dummy$RID == 3, 1, 0)  
child\_ica\_dummy$Balochistan <- ifelse(child\_ica\_dummy$RID == 4, 1, 0)  
child\_ica\_dummy$Khyber\_Pakhtunkhwa <- ifelse(child\_ica\_dummy$RID == 5, 1, 0)  
child\_ica\_dummy$Gilgit\_Baltistan <- ifelse(child\_ica\_dummy$RID == 6, 1, 0)  
child\_ica\_dummy$Azad\_Jammu\_and\_Kashmir <- ifelse(child\_ica\_dummy$RID == 7, 1, 0)  
child\_ica\_dummy$Islamabad\_ICT <- ifelse(child\_ica\_dummy$RID == 8, 1, 0)  
child\_ica\_dummy$Federally\_Administrated\_Tribal\_Areas <- ifelse(child\_ica\_dummy$RID == 9, 1, 0)

#### Factor Variables

child\_ica\_dummy$DID <- as.factor(child\_ica\_dummy$DID)  
child\_ica\_dummy$C002 <- as.factor(child\_ica\_dummy$C002)  
child\_ica\_dummy$C003 <- as.factor(child\_ica\_dummy$C003)  
# child\_ica\_dummy$School.Enrollment <- as.factor(child\_ica\_dummy$School.Enrollment)  
child\_ica\_dummy$Household.Ownership <- as.factor(child\_ica\_dummy$Household.Ownership)  
child\_ica\_dummy$Parents.Edu1 <- as.factor(child\_ica\_dummy$Parents.Edu1)  
child\_ica\_dummy$Parents.Edu2 <- as.factor(child\_ica\_dummy$Parents.Edu2)  
child\_ica\_dummy$Gender <- as.factor(child\_ica\_dummy$Gender)

### Notes on Data Wrangling

#### Eliminated NAs

child\_ica %>%   
 summarize(  
 School.Enrollment\_NA = sum(is.na(C003)),  
 Gender\_NA = sum(is.na(C002)),  
 Parents.Edu\_NA = sum(is.na(PR004))+sum(is.na(PR009)),  
 Household.Ownership\_NA = sum(is.na(H003))  
 )

## School.Enrollment\_NA Gender\_NA Parents.Edu\_NA Household.Ownership\_NA  
## 1 0 66 10775 4500

#### Eliminated Rows in Total

data.frame(original\_rows = nrow(child\_ica),  
 eliminated\_rows = nrow(child\_ica) - nrow(child\_ica\_dummy),  
 ratio = (nrow(child\_ica)-nrow(child\_ica\_dummy))/nrow(child\_ica))

## original\_rows eliminated\_rows ratio  
## 1 255196 13218 0.05179548

child\_ica %>%  
 summarize(  
 "Variables" = c(  
 "School.Enrollment\_NA",   
 "Gender\_NA",   
 "Parents.Edu\_NA",   
 "Household.Ownership\_NA",  
 "Total"),   
 "Deleted Rows" = c(  
 sum(is.na(C003)),   
 sum(is.na(C002)),   
 sum(is.na(PR004))+sum(is.na(PR009)),   
 sum(is.na(H003)),  
 nrow(child\_ica %>% filter(is.na(C003) | is.na(C002) | is.na(PR004) | is.na(PR009) | is.na(H003)))),  
 "Ratio" = c(  
 sum(is.na(C003))/nrow(child\_ica),  
 sum(is.na(C002))/nrow(child\_ica),   
 (sum(is.na(PR004))+sum(is.na(PR009)))/nrow(child\_ica),   
 sum(is.na(H003))/nrow(child\_ica),  
 nrow(child\_ica %>% filter(is.na(C003) | is.na(C002) | is.na(PR004) | is.na(PR009) | is.na(H003)))/nrow(child\_ica))  
 )

## Variables Deleted Rows Ratio  
## 1 School.Enrollment\_NA 0 0.0000000000  
## 2 Gender\_NA 66 0.0002586247  
## 3 Parents.Edu\_NA 10775 0.0422224486  
## 4 Household.Ownership\_NA 4500 0.0176335052  
## 5 Total 13218 0.0517954827

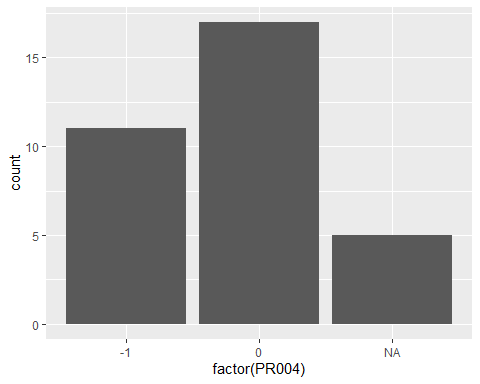
#### Eliminated NAs Hunza

child\_ica %>%   
 filter(DID == 266) %>%   
 summarize(  
 "Variables" = c(  
 "School.Enrollment\_NA",   
 "Gender\_NA",   
 "Parents.Edu\_NA",   
 "Household.Ownership\_NA",  
 "Total"),   
 "Deleted Rows" = c(  
 sum(is.na(C003)),   
 sum(is.na(C002)),   
 sum(is.na(PR004))+sum(is.na(PR009)),   
 sum(is.na(H003)),  
 nrow(child\_ica %>% filter(DID == 266, is.na(C003) | is.na(C002) | is.na(PR004) | is.na(PR009) | is.na(H003)))),  
 "Ratio" = c(  
 sum(is.na(C003))/nrow(child\_ica %>% filter(DID ==266)),  
 sum(is.na(C002))/nrow(child\_ica %>% filter(DID ==266)),   
 (sum(is.na(PR004))+sum(is.na(PR009)))/nrow(child\_ica %>% filter(DID ==266)),   
 sum(is.na(H003))/nrow(child\_ica %>% filter(DID ==266)),  
 nrow(child\_ica %>% filter(DID == 266, is.na(C003) | is.na(C002) | is.na(PR004) | is.na(PR009) | is.na(H003)))/nrow(child\_ica %>% filter(DID ==266)))  
 )

## Variables Deleted Rows Ratio  
## 1 School.Enrollment\_NA 0 0.00000000  
## 2 Gender\_NA 0 0.00000000  
## 3 Parents.Edu\_NA 33 0.02010969  
## 4 Household.Ownership\_NA 33 0.02010969  
## 5 Total 59 0.03595369

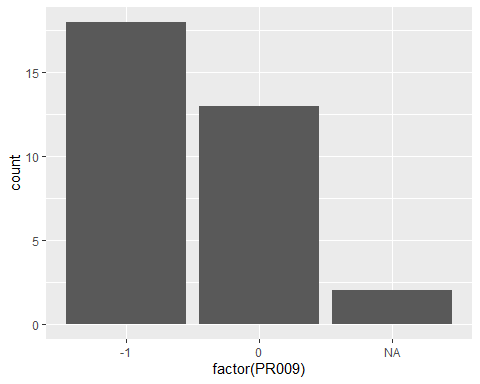
child\_ica %>%   
 filter(DID == 266, is.na(H003)) %>%   
 ggplot(aes(factor(PR004))) +  
 geom\_histogram(stat = "count")

## Warning: Ignoring unknown parameters: binwidth, bins, pad



child\_ica %>%   
 filter(DID == 266, is.na(H003)) %>%   
 ggplot(aes(factor(PR009))) +  
 geom\_histogram(stat = "count")

## Warning: Ignoring unknown parameters: binwidth, bins, pad



#### Eliminated Rows in Total Hunza

data.frame(original\_rows = nrow(child\_ica %>% filter(DID == 266)),  
 eliminated\_rows = nrow(child\_ica %>% filter(DID == 266)) - nrow(child\_ica\_dummy %>% filter(DID == 266)),  
 ratio = (nrow(child\_ica %>% filter(DID == 266) %>% filter(DID == 266))-nrow(child\_ica\_dummy %>% filter(DID == 266)))/nrow(child\_ica %>% filter(DID == 266)))

## original\_rows eliminated\_rows ratio  
## 1 1641 59 0.03595369

## Hunza. Generalized Linear Models

### Formula 1

### Model 1 [NofChild + HHOwnership]

#### Hunza. GLM Age >= 5 School.Enrollment ~ Age + Gender + Parents.Edu1 + Num.of.Child + Household.Ownership

glm\_child <- glm(School.Enrollment ~ Age + Gender + Parents.Edu1 + Num.of.Child + Household.Ownership, family = "binomial", data = child\_ica\_dummy %>% filter(C001 >= 5, !is.na(Household.Ownership), DID == 266))  
glm\_child %>% summary()

##   
## Call:  
## glm(formula = School.Enrollment ~ Age + Gender + Parents.Edu1 +   
## Num.of.Child + Household.Ownership, family = "binomial",   
## data = child\_ica\_dummy %>% filter(C001 >= 5, !is.na(Household.Ownership),   
## DID == 266))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.0718 0.1538 0.1881 0.2457 0.9822   
##   
## Coefficients:  
##

Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 3.31784 0.85973 3.859 0.000114 \*\*\*  
## Age -0.02366 0.04764 -0.497 0.619400   
## Gender1 0.26299 0.32483 0.810 0.418151   
## Parents.Edu11 0.68028 0.35535 1.914 0.055566 .   
## Num.of.Child -0.36511 0.11685 -3.125 0.001781 \*\*   
## Household.Ownership1 1.45449 0.38143 3.813 0.000137 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 376.54 on 1387 degrees of freedom  
## Residual deviance: 337.19 on 1382 degrees of freedom  
## AIC: 349.19  
##   
## Number of Fisher Scoring iterations: 7

# ggPredict(glm\_child, se = TRUE, colorAsFactor = TRUE, show.summary = TRUE, point = TRUE) +  
# theme(text = element\_text(family = "serif")) +  
# xlab("Number of Children in Each House Hold") +  
# ylab("Enrollment") +  
# labs(color = "House Hold Ownership", fill = "House Hold Ownership")

extract\_eq(glm\_child)

extract\_eq(glm\_child, use\_coefs = TRUE)

##### exponential transformation

exp(glm\_child$coefficients)

## (Intercept) Age Gender1   
## 27.600566 0.976615 1.300815   
## Parents.Edu11 Num.of.Child Household.Ownership1   
## 1.974439 0.694119 4.282316

##### confidence interval (intercept and coefficient)

confint(glm\_child, level = 0.95)

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 1.67909700 5.05580503  
## Age -0.11830273 0.06927947  
## Gender1 -0.37479224 0.90703643  
## Parents.Edu11 -0.02667129 1.37406680  
## Num.of.Child -0.59642519 -0.13697033  
## Household.Ownership1 0.67326418 2.17905463

##### exponential transformation of confidence interval (odds ratio of intercept and coefficient)

exp(confint(glm\_child, level = 0.95))

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 5.3607131 156.9308136  
## Age 0.8884271 1.0717357  
## Gender1 0.6874321 2.4769710  
## Parents.Edu11 0.9736812 3.9513876  
## Num.of.Child 0.5507770 0.8719961  
## Household.Ownership1 1.9606267 8.8379472

##### Hosmer-Lemeshow

hoslem.test(x = glm\_child$y, y = fitted(glm\_child))

##   
## Hosmer and Lemeshow goodness of fit (GOF) test  
##   
## data: glm\_child$y, fitted(glm\_child)  
## X-squared = 4.9975, df = 8, p-value = 0.7578

##### AIC

extractAIC(glm\_child)

## [1] 6.0000 349.1946

##### BIC

extractAIC(glm\_child, k = log(nrow(glm\_child$data)))

## [1] 6.0000 380.6083

##### effectiveness of explanatory variables

glm\_child\_null <- glm(School.Enrollment ~ 1, family = "binomial",   
 data = child\_ica\_dummy %>% filter(!is.na(Household.Ownership), C001 >= 5, DID == 266))  
anova(glm\_child\_null, glm\_child, test = "Chisq")

## Analysis of Deviance Table  
##   
## Model 1: School.Enrollment ~ 1  
## Model 2: School.Enrollment ~ Age + Gender + Parents.Edu1 + Num.of.Child +   
## Household.Ownership  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)   
## 1 1387 376.54   
## 2 1382 337.19 5 39.349 2.02e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##### variables selection

step(glm\_child\_null, direction = "both",  
 scope = (~ Age + Gender + Num.of.Child + Household.Ownership))

## Start: AIC=378.54  
## School.Enrollment ~ 1  
##   
## Df Deviance AIC  
## + Household.Ownership 1 357.40 361.40  
## + Num.of.Child 1 357.92 361.92  
## <none> 376.54 378.54  
## + Age 1 375.57 379.57  
## + Gender 1 376.15 380.15  
##   
## Step: AIC=361.4  
## School.Enrollment ~ Household.Ownership  
##   
## Df Deviance AIC  
## + Num.of.Child 1 342.04 348.04  
## <none> 357.40 361.40  
## + Age 1 356.13 362.13  
## + Gender 1 357.22 363.22  
## - Household.Ownership 1 376.54 378.54  
##   
## Step: AIC=348.04  
## School.Enrollment ~ Household.Ownership + Num.of.Child  
##   
## Df Deviance AIC  
## <none> 342.04 348.04  
## + Gender 1 341.38 349.38  
## + Age 1 341.47 349.47  
## - Num.of.Child 1 357.40 361.40  
## - Household.Ownership 1 357.92 361.92

##   
## Call: glm(formula = School.Enrollment ~ Household.Ownership + Num.of.Child,   
## family = "binomial", data = child\_ica\_dummy %>% filter(!is.na(Household.Ownership),   
## C001 >= 5, DID == 266))  
##   
## Coefficients:  
## (Intercept) Household.Ownership1 Num.of.Child   
## 3.7653 1.6398 -0.4373   
##   
## Degrees of Freedom: 1387 Total (i.e. Null); 1385 Residual  
## Null Deviance: 376.5   
## Residual Deviance: 342 AIC: 348

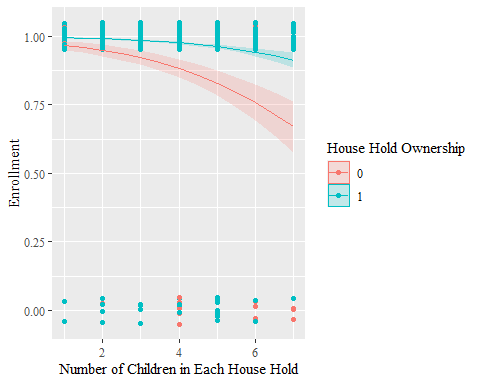
### Model 1 [NofChild + HHOwnership]

#### Hunza. GLM Age >= 5 School.Enrollment ~ Num.of.Child + Household.Ownership

glm\_child <- glm(School.Enrollment ~ Num.of.Child + Household.Ownership, family = "binomial", data = child\_ica\_dummy %>% filter(C001 >= 5, !is.na(Household.Ownership), DID == 266))  
  
ggPredict(glm\_child, se = TRUE, colorAsFactor = TRUE, show.summary = TRUE, point = TRUE) +  
 theme(text = element\_text(family = "serif")) +  
 xlab("Number of Children in Each House Hold") +  
 ylab("Enrollment") +  
 labs(color = "House Hold Ownership", fill = "House Hold Ownership")

##   
## Call:  
## glm(formula = School.Enrollment ~ Num.of.Child + Household.Ownership,   
## family = "binomial", data = child\_ica\_dummy %>% filter(C001 >=   
## 5, !is.na(Household.Ownership), DID == 266))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.1543 0.1464 0.1819 0.2259 0.8965   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 3.7653 0.5861 6.425 1.32e-10 \*\*\*  
## Num.of.Child -0.4373 0.1105 -3.958 7.54e-05 \*\*\*  
## Household.Ownership1 1.6398 0.3680 4.456 8.36e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 376.54 on 1387 degrees of freedom  
## Residual deviance: 342.04 on 1385 degrees of freedom  
## AIC: 348.04  
##   
## Number of Fisher Scoring iterations: 7

## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!



extract\_eq(glm\_child)

extract\_eq(glm\_child, use\_coefs = TRUE)

##### exponential transformation

exp(glm\_child$coefficients)

## (Intercept) Num.of.Child Household.Ownership1   
## 43.1780424 0.6457498 5.1540565

##### confidence interval (intercept and coefficient)

confint(glm\_child, level = 0.95)

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 2.6561231 4.9606809  
## Num.of.Child -0.6549398 -0.2204113  
## Household.Ownership1 0.8840013 2.3381830

##### exponential transformation of confidence interval (odds ratio of intercept and coefficient)

exp(confint(glm\_child, level = 0.95))

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 14.2409716 142.6909268  
## Num.of.Child 0.5194733 0.8021888  
## Household.Ownership1 2.4205657 10.3623908

##### Hosmer-Lemeshow

hoslem.test(x = glm\_child$y, y = fitted(glm\_child))

##   
## Hosmer and Lemeshow goodness of fit (GOF) test  
##   
## data: glm\_child$y, fitted(glm\_child)  
## X-squared = 13.475, df = 8, p-value = 0.09653

##### AIC

extractAIC(glm\_child)

## [1] 3.0000 348.0375

##### BIC

extractAIC(glm\_child, k = log(nrow(glm\_child$data)))

## [1] 3.0000 363.7443

##### effectiveness of explanatory variables

glm\_child\_null <- glm(School.Enrollment ~ 1, family = "binomial",   
 data = child\_ica\_dummy %>% filter(!is.na(Household.Ownership), C001 >= 5, DID == 266))  
anova(glm\_child\_null, glm\_child, test = "Chisq")

## Analysis of Deviance Table  
##   
## Model 1: School.Enrollment ~ 1  
## Model 2: School.Enrollment ~ Num.of.Child + Household.Ownership  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)   
## 1 1387 376.54   
## 2 1385 342.04 2 34.506 3.214e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

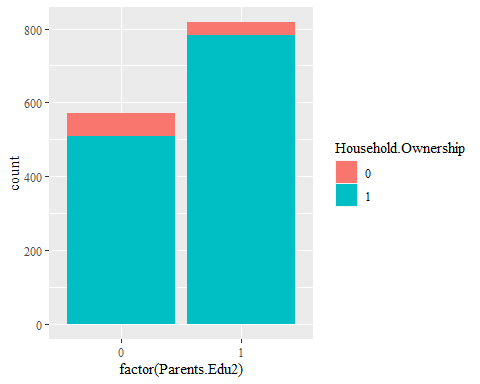
##### variables selection

step(glm\_child\_null, direction = "both",  
 scope = (~ Age + Gender + Num.of.Child + Household.Ownership))

## Start: AIC=378.54  
## School.Enrollment ~ 1  
##   
## Df Deviance AIC  
## + Household.Ownership 1 357.40 361.40  
## + Num.of.Child 1 357.92 361.92  
## <none> 376.54 378.54  
## + Age 1 375.57 379.57  
## + Gender 1 376.15 380.15  
##   
## Step: AIC=361.4  
## School.Enrollment ~ Household.Ownership  
##   
## Df Deviance AIC  
## + Num.of.Child 1 342.04 348.04  
## <none> 357.40 361.40  
## + Age 1 356.13 362.13  
## + Gender 1 357.22 363.22  
## - Household.Ownership 1 376.54 378.54  
##   
## Step: AIC=348.04  
## School.Enrollment ~ Household.Ownership + Num.of.Child  
##   
## Df Deviance AIC  
## <none> 342.04 348.04  
## + Gender 1 341.38 349.38  
## + Age 1 341.47 349.47  
## - Num.of.Child 1 357.40 361.40  
## - Household.Ownership 1 357.92 361.92

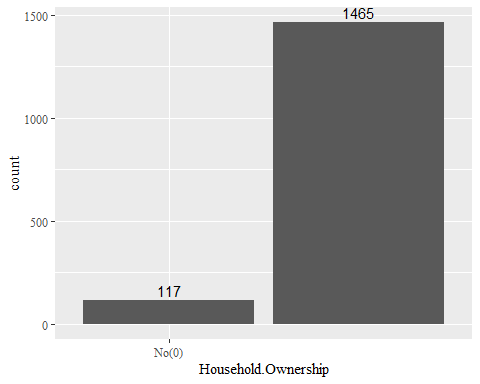
##   
## Call: glm(formula = School.Enrollment ~ Household.Ownership + Num.of.Child,   
## family = "binomial", data = child\_ica\_dummy %>% filter(!is.na(Household.Ownership),   
## C001 >= 5, DID == 266))  
##   
## Coefficients:  
## (Intercept) Household.Ownership1 Num.of.Child   
## 3.7653 1.6398 -0.4373   
##   
## Degrees of Freedom: 1387 Total (i.e. Null); 1385 Residual  
## Null Deviance: 376.5   
## Residual Deviance: 342 AIC: 348

child\_ica\_dummy %>%   
 filter(!is.na(Household.Ownership), Age >= 5, DID == 266) %>%   
 group\_by(Household.Ownership) %>%   
 ggplot(aes(factor(Parents.Edu2), fill = Household.Ownership)) +  
 geom\_bar(stat = "count") +  
 theme(text = element\_text(family = "serif"))

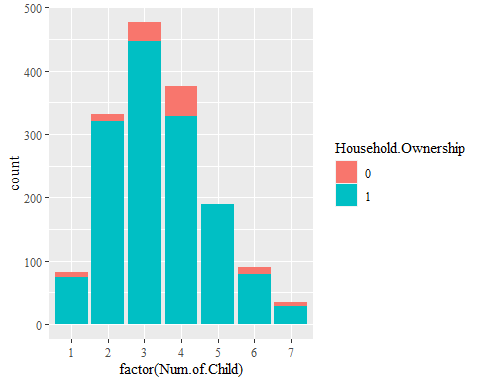


There seems multi-colinearity between Household Ownership and Both-Parents Education

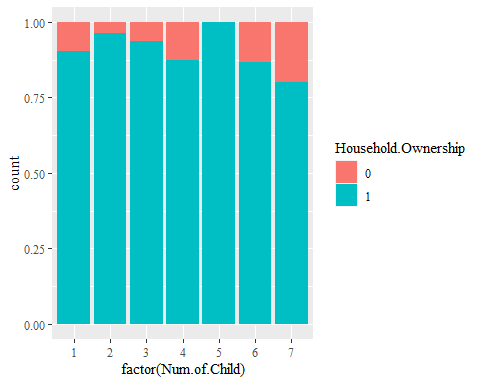
child\_ica\_dummy %>%   
 filter(DID == 266) %>%   
 ggplot(aes(Household.Ownership)) +  
 geom\_bar() +  
 geom\_text(aes(label = ..count..), stat = "count", vjust = -.3) +  
 scale\_x\_discrete(breaks = c(-1, 0, NA), labels = c("Yes(-1)", "No(0)", "NA")) +  
 theme(text = element\_text(family = "serif"))



child\_ica\_dummy %>%   
 filter(DID == 266) %>%   
 ggplot(aes(factor(Num.of.Child), fill = Household.Ownership)) +  
 geom\_bar() +  
 theme(text = element\_text(family = "serif"))



child\_ica\_dummy %>%   
 filter(DID == 266, !is.na(Household.Ownership)) %>%   
 ggplot(aes(factor(Num.of.Child), fill = Household.Ownership)) +  
 geom\_bar(stat = "count", position = "fill") +  
 theme(text = element\_text(family = "serif"))



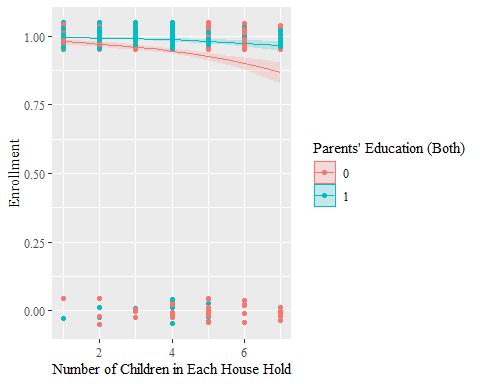
### Model 2 [NofChild + ParentsBothEdu]

#### Hunza. GLM Age >= 5 School.Enrollment ~ Num.of.Child + Parents.Edu2

glm\_child <- glm(School.Enrollment ~ Num.of.Child + Parents.Edu2, family = "binomial", data = child\_ica\_dummy %>% filter(Age >= 5, DID == 266))  
  
ggPredict(glm\_child, se = TRUE, colorAsFactor = TRUE, show.summary = TRUE, point = TRUE) +  
 theme(text = element\_text(family = "serif")) +  
 xlab("Number of Children in Each House Hold") +  
 ylab("Enrollment") +  
 labs(color = "Parents' Education (Both)", fill = "Parents' Education (Both)")

##   
## Call:  
## glm(formula = School.Enrollment ~ Num.of.Child + Parents.Edu2,   
## family = "binomial", data = child\_ica\_dummy %>% filter(Age >=   
## 5, DID == 266))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.2505 0.1401 0.1650 0.2849 0.5364   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 4.1725 0.5338 7.817 5.4e-15 \*\*\*  
## Num.of.Child -0.3295 0.1113 -2.961 0.003062 \*\*   
## Parents.Edu21 1.4349 0.3945 3.637 0.000276 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 376.54 on 1387 degrees of freedom  
## Residual deviance: 342.33 on 1385 degrees of freedom  
## AIC: 348.33  
##   
## Number of Fisher Scoring iterations: 7

## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!



extract\_eq(glm\_child)

extract\_eq(glm\_child, use\_coefs = TRUE)

##### exponential transformation

exp(glm\_child$coefficients)

## (Intercept) Num.of.Child Parents.Edu21   
## 64.8757234 0.7192855 4.1993047

##### confidence interval (intercept and coefficient)

confint(glm\_child, level = 0.95)

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 3.1726908 5.2699476  
## Num.of.Child -0.5498156 -0.1123412  
## Parents.Edu21 0.6990630 2.2636876

##### exponential transformation of confidence interval (odds ratio of intercept and coefficient)

exp(confint(glm\_child, level = 0.95))

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 23.8716309 194.4057829  
## Num.of.Child 0.5770562 0.8937393  
## Parents.Edu21 2.0118667 9.6184928

##### Hosmer-Lemeshow

hoslem.test(x = glm\_child$y, y = fitted(glm\_child))

##   
## Hosmer and Lemeshow goodness of fit (GOF) test  
##   
## data: glm\_child$y, fitted(glm\_child)  
## X-squared = 5.696, df = 8, p-value = 0.6812

##### AIC

extractAIC(glm\_child)

## [1] 3.0000 348.3284

##### BIC

extractAIC(glm\_child, k = log(nrow(glm\_child$data)))

## [1] 3.0000 364.0352

##### effectiveness of explanatory variables

glm\_child\_null <- glm(School.Enrollment~1, family = "binomial", data = child\_ica\_dummy %>% filter(Age >= 5, DID == 266))  
anova(glm\_child\_null, glm\_child, test = "Chisq")

## Analysis of Deviance Table  
##   
## Model 1: School.Enrollment ~ 1  
## Model 2: School.Enrollment ~ Num.of.Child + Parents.Edu2  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)   
## 1 1387 376.54   
## 2 1385 342.33 2 34.215 3.717e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##### variables selection

step(glm\_child\_null, direction = "both",   
 scope = (~ Age + Gender + Parents.Edu1 + Num.of.Child))

## Start: AIC=378.54  
## School.Enrollment ~ 1  
##   
## Df Deviance AIC  
## + Num.of.Child 1 357.92 361.92  
## + Parents.Edu1 1 361.88 365.88  
## <none> 376.54 378.54  
## + Age 1 375.57 379.57  
## + Gender 1 376.15 380.15  
##   
## Step: AIC=361.92  
## School.Enrollment ~ Num.of.Child  
##   
## Df Deviance AIC  
## + Parents.Edu1 1 350.58 356.58  
## <none> 357.92 361.92  
## + Gender 1 356.86 362.86  
## + Age 1 357.27 363.27  
## - Num.of.Child 1 376.54 378.54  
##   
## Step: AIC=356.58  
## School.Enrollment ~ Num.of.Child + Parents.Edu1  
##   
## Df Deviance AIC  
## <none> 350.58 356.58  
## + Gender 1 349.63 357.63  
## + Age 1 350.37 358.37  
## - Parents.Edu1 1 357.92 361.92  
## - Num.of.Child 1 361.88 365.88

##   
## Call: glm(formula = School.Enrollment ~ Num.of.Child + Parents.Edu1,   
## family = "binomial", data = child\_ica\_dummy %>% filter(Age >=   
## 5, DID == 266))  
##   
## Coefficients:  
## (Intercept) Num.of.Child Parents.Edu11   
## 4.3285 -0.3773 0.9195   
##   
## Degrees of Freedom: 1387 Total (i.e. Null); 1385 Residual  
## Null Deviance: 376.5   
## Residual Deviance: 350.6 AIC: 356.6

##### multicolinearity

vif(glm\_child)

## Num.of.Child Parents.Edu2   
## 1.072003 1.072003

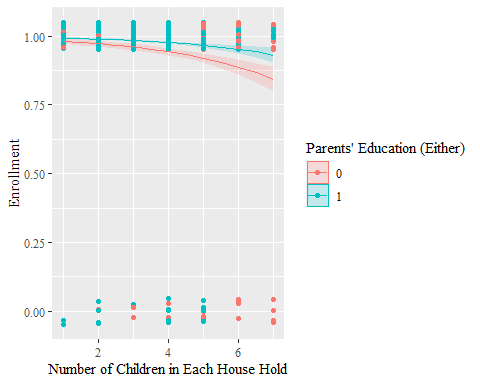
### Model 3 [NofChild + ParentsEitherEdu]

#### Hunza. GLM Age >= 5 School.Enrollment ~ Num.of.Child + Parents.Edu1

glm\_child <- glm(School.Enrollment ~ Num.of.Child + Parents.Edu1, family = "binomial", data = child\_ica\_dummy %>% filter(Age >= 5, DID == 266))  
  
ggPredict(glm\_child, se = TRUE, colorAsFactor = TRUE, show.summary = TRUE, point = TRUE) +  
 theme(text = element\_text(family = "serif")) +  
 xlab("Number of Children in Each House Hold") +  
 ylab("Enrollment") +  
 labs(color = "Parents' Education (Either)", fill = "Parents' Education (Either)")

##   
## Call:  
## glm(formula = School.Enrollment ~ Num.of.Child + Parents.Edu1,   
## family = "binomial", data = child\_ica\_dummy %>% filter(Age >=   
## 5, DID == 266))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.1236 0.1799 0.2168 0.2611 0.5827   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 4.3285 0.5830 7.425 1.13e-13 \*\*\*  
## Num.of.Child -0.3773 0.1123 -3.359 0.000783 \*\*\*  
## Parents.Edu11 0.9195 0.3340 2.753 0.005913 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 376.54 on 1387 degrees of freedom  
## Residual deviance: 350.58 on 1385 degrees of freedom  
## AIC: 356.58  
##   
## Number of Fisher Scoring iterations: 6

## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!



extract\_eq(glm\_child)

extract\_eq(glm\_child, use\_coefs = TRUE)

##### exponential transformation

exp(glm\_child$coefficients)

## (Intercept) Num.of.Child Parents.Edu11   
## 75.8306084 0.6857011 2.5079778

##### confidence interval (intercept and coefficient)

confint(glm\_child, level = 0.95)

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 3.2334116 5.5234326  
## Num.of.Child -0.5994177 -0.1576826  
## Parents.Edu11 0.2586663 1.5759324

##### exponential transformation of confidence interval (odds ratio of intercept and coefficient)

exp(confint(glm\_child, level = 0.95))

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 25.3660473 250.4934057  
## Num.of.Child 0.5491313 0.8541208  
## Parents.Edu11 1.2952016 4.8352479

##### Hosmer-Lemeshow

hoslem.test(x = glm\_child$y, y = fitted(glm\_child))

##   
## Hosmer and Lemeshow goodness of fit (GOF) test  
##   
## data: glm\_child$y, fitted(glm\_child)  
## X-squared = 17.158, df = 8, p-value = 0.02851

##### AIC

extractAIC(glm\_child)

## [1] 3.000 356.582

##### BIC

extractAIC(glm\_child, k = log(nrow(glm\_child$data)))

## [1] 3.0000 372.2889

## Pak. Generalized Linear Model

### Model 1 [NofChild + HHOwnership]

#### Each Dist. GLM Age >= 5 C003\_01 ~ NCH + H003

each\_dist <- sapply(as.numeric(as.character(unique(child\_ica\_dummy$DID))), function(id){  
 child\_ica\_dummy <- child\_ica\_dummy %>% filter(DID == id)  
 glm\_child <- glm(School.Enrollment ~ Num.of.Child + Household.Ownership,   
 family = "binomial", data = child\_ica\_dummy %>% filter(Age >= 5, !is.na(Household.Ownership)))  
 temp\_hoslem <- hoslem.test(x = glm\_child$y, y = fitted(glm\_child))  
 data.frame(DID = id, hoslem\_p\_value = temp\_hoslem$p.value)  
}) %>% t()

each\_dist <- each\_dist %>%   
 as.data.frame()   
each\_dist$DID <- each\_dist %>%   
 pull(DID) %>%   
 as.numeric()  
each\_dist$hoslem\_p\_value <- each\_dist %>%   
 pull(hoslem\_p\_value) %>%   
 as.numeric()  
each\_dist

## DID hoslem\_p\_value  
## 1 146 3.702152e-01  
## 2 147 8.402235e-01  
## 3 148 5.747210e-01  
## 4 149 9.583907e-01  
## 5 150 8.932666e-01  
## 6 151 8.977396e-01  
## 7 152 3.931301e-01  
## 8 153 1.868806e-01  
## 9 154 2.252387e-01  
## 10 155 9.988024e-01  
## 11 156 4.570744e-01  
## 12 157 1.274284e-01  
## 13 158 7.766738e-01  
## 14 159 9.786181e-01  
## 15 160 NaN  
## 16 161 6.622064e-01  
## 17 162 9.541911e-01  
## 18 163 3.094742e-01  
## 19 164 7.114949e-01  
## 20 165 1.554080e-02  
## 21 166 5.996638e-01  
## 22 167 8.584151e-01  
## 23 169 9.561827e-01  
## 24 170 2.845733e-01  
## 25 171 NaN  
## 26 172 7.646206e-01  
## 27 173 1.801413e-01  
## 28 174 4.621954e-01  
## 29 175 9.488777e-01  
## 30 176 6.698873e-01  
## 31 177 3.693581e-01  
## 32 178 6.375828e-01  
## 33 179 2.404270e-01  
## 34 180 7.712604e-02  
## 35 181 9.756719e-01  
## 36 182 8.916927e-01  
## 37 183 1.778700e-01  
## 38 184 NaN  
## 39 185 NaN  
## 40 186 7.997788e-01  
## 41 187 3.649090e-01  
## 42 188 3.434400e-01  
## 43 189 4.202915e-01  
## 44 190 8.283956e-01  
## 45 191 8.599984e-01  
## 46 192 8.128284e-01  
## 47 193 2.314288e-06  
## 48 194 3.126762e-01  
## 49 195 1.378695e-01  
## 50 196 4.316092e-01  
## 51 197 8.306028e-01  
## 52 198 2.599524e-01  
## 53 199 1.082096e-02  
## 54 200 1.518975e-02  
## 55 202 3.956771e-01  
## 56 203 1.349658e-06  
## 57 204 6.204874e-01  
## 58 315 8.202915e-01  
## 59 316 1.389904e-01  
## 60 320 5.912602e-05  
## 61 205 9.999470e-01  
## 62 206 9.963441e-01  
## 63 207 7.404630e-01  
## 64 208 9.846597e-01  
## 65 209 3.866598e-01  
## 66 210 1.507208e-01  
## 67 211 8.883225e-01  
## 68 212 9.725283e-01  
## 69 213 8.938643e-01  
## 70 214 8.987225e-01  
## 71 215 6.623776e-01  
## 72 216 9.976142e-01  
## 73 217 5.119611e-01  
## 74 218 2.221756e-01  
## 75 219 8.848497e-01  
## 76 220 3.592986e-02  
## 77 221 7.504679e-01  
## 78 222 9.552782e-02  
## 79 223 9.962632e-01  
## 80 224 9.536877e-01  
## 81 225 9.999157e-01  
## 82 226 3.045858e-01  
## 83 227 4.912747e-01  
## 84 228 9.983706e-01  
## 85 229 9.995215e-01  
## 86 230 9.992232e-01  
## 87 231 9.209521e-01  
## 88 232 3.822277e-01  
## 89 233 6.763309e-01  
## 90 234 1.920139e-01  
## 91 318 3.606647e-01  
## 92 319 5.949079e-01  
## 93 235 6.125810e-01  
## 94 236 4.774151e-01  
## 95 237 1.786874e-02  
## 96 238 6.661981e-01  
## 97 239 9.394117e-01  
## 98 240 2.511028e-01  
## 99 241 2.053414e-03  
## 100 242 7.527516e-01  
## 101 243 5.712028e-03  
## 102 244 5.000419e-01  
## 103 245 9.999208e-01  
## 104 246 5.653004e-01  
## 105 247 7.135440e-01  
## 106 248 9.079060e-03  
## 107 249 5.100203e-01  
## 108 250 8.443304e-02  
## 109 251 9.194270e-01  
## 110 252 9.796447e-01  
## 111 253 2.954387e-01  
## 112 254 2.235738e-01  
## 113 255 1.950090e-01  
## 114 256 4.500482e-01  
## 115 257 7.300007e-01  
## 116 258 4.399887e-04  
## 117 259 8.222252e-01  
## 118 260 5.329579e-01  
## 119 261 7.541846e-01  
## 120 262 5.458027e-01  
## 121 263 9.993226e-01  
## 122 264 3.744874e-01  
## 123 265 9.957911e-01  
## 124 266 9.652839e-02  
## 125 267 6.100159e-01  
## 126 268 6.908767e-01  
## 127 269 9.977134e-01  
## 128 270 4.451470e-01  
## 129 271 7.444678e-04  
## 130 272 8.059775e-01  
## 131 273 5.713590e-01  
## 132 274 8.981888e-01  
## 133 275 6.771278e-01  
## 134 276 9.998250e-01  
## 135 277 7.339147e-01  
## 136 278 1.568815e-02  
## 137 279 7.700977e-01  
## 138 280 5.990123e-01  
## 139 281 9.261617e-01  
## 140 282 9.572941e-01  
## 141 284 1.741378e-01  
## 142 287 4.787838e-01  
## 143 289 1.259827e-01  
## 144 290 1.823365e-01

#### Dists with P value <= 0.05

DID\_HosLem0.05\_model\_1 <- each\_dist %>%   
 filter(hoslem\_p\_value <= 0.05) %>%   
 pull(DID)

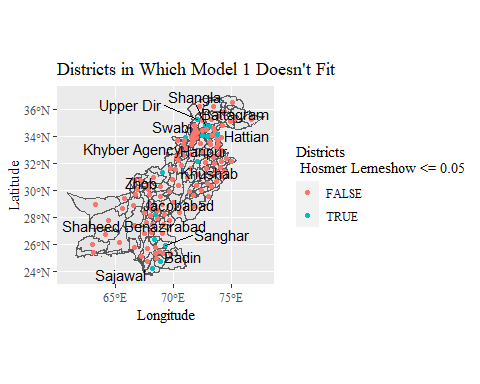
ica %>%   
 mutate(centroid = st\_centroid(geometry),  
 x = st\_coordinates(centroid)[,1],  
 y = st\_coordinates(centroid)[,2]) %>%   
 ggplot(ratio = 1) +  
 coord\_sf() +  
 geom\_sf(ratio = 1) +  
 geom\_point(data = child\_ica\_dummy %>%   
 filter(!is.na(x)) %>%   
 summarize(x = unique(x),  
 y = unique(y),  
 DID = unique(DID)),  
 aes(x, y, color = DID %in% DID\_HosLem0.05\_model\_1)) +  
 geom\_text\_repel(data = child\_ica\_dummy %>%   
 filter(DID %in% DID\_HosLem0.05\_model\_1, !is.na(x)) %>%   
 summarize(x = unique(x),  
 y = unique(y),  
 DID = unique(DID),  
 DNAME = unique(DNAME)),  
 aes(x, y, label = DNAME), force = 2) +  
 ggtitle("Districts in Which Model 1 Doesn't Fit") +  
 labs(color = "Districts \n Hosmer Lemeshow <= 0.05") +  
 theme(text = element\_text(family = "serif")) +  
 xlab("Longitude") +  
 ylab("Latitude")

## Warning: Problem with `mutate()` input `centroid`.  
## x st\_centroid does not give correct centroids for longitude/latitude data  
## i Input `centroid` is `st\_centroid(geometry)`.

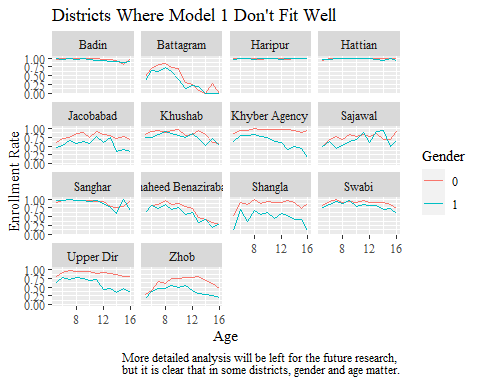
## Warning in st\_centroid.sfc(geometry): st\_centroid does not give correct  
## centroids for longitude/latitude data

## Warning: Ignoring unknown parameters: ratio

## Coordinate system already present. Adding new coordinate system, which will replace the existing one.



child\_ica\_dummy %>%   
 filter(Age >= 5, DID %in% DID\_HosLem0.05\_model\_1) %>%   
 group\_by(DID, Age, Gender) %>%   
 mutate(rate = mean(C003 == 3)) %>%   
 ggplot(aes(Age, rate, color = factor(Gender))) +  
 geom\_line() +  
 facet\_wrap(.~DNAME) +  
 xlab("Age") +  
 ylab("Enrollment Rate") +  
 labs(title = "Districts Where Model 1 Don't Fit Well",   
 caption = "More detailed analysis will be left for the future research, \n but it is clear that in some districts, gender and age matter.",  
 color = "Gender") +  
 theme(text = element\_text(family = "serif"))



#### Each Dist. GLM Age >= 5 School.Enrollment ~ Num.of.Child + Household.Ownership + [C002]

step\_each\_dist <- sapply(as.numeric(as.character(unique(child\_ica\_dummy$DID))), function(id){  
 glm\_child\_null <- glm(School.Enrollment ~ 1, family = "binomial",   
 data = child\_ica\_dummy %>% filter(!is.na(Household.Ownership), Age >= 5, DID == id))  
 step <- step(glm\_child\_null, direction = "both", scope = (~ Age + Gender + Num.of.Child + Household.Ownership))  
 data.frame(as.character(step$formula), id)  
}) %>% t()

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

step\_each\_dist[,1] <- step\_each\_dist[,1] %>%   
 as.character() %>%   
 str\_remove("c\\(\"~\",\\s\"School.Enrollment\",\\s\"") %>%   
 str\_remove("\"\\)")   
  
step\_each\_dist[,2] <- step\_each\_dist[,2] %>%   
 as.character() %>%   
 str\_remove("c\\(\\d{3},\\s\\d{3},\\s") %>%   
 str\_remove("\\)") %>%   
 as.numeric()  
  
  
dist\_gender <- step\_each\_dist[,1] %>%   
 str\_which("Gender")   
  
dists\_not\_gender <- step\_each\_dist[-dist\_gender,] %>%   
 .[,2] %>%   
 as.integer()

ica %>%   
 mutate(centroid = st\_centroid(geometry),  
 x = st\_coordinates(centroid)[,1],  
 y = st\_coordinates(centroid)[,2]) %>%   
 ggplot(ratio = 1) +  
 coord\_sf() +  
 geom\_sf(ratio = 1, aes(fill = Province)) +  
 geom\_point(data = child\_ica\_dummy %>%   
 filter(!is.na(x)) %>%   
 summarize(x = unique(x),  
 y = unique(y),  
 DID = unique(DID)),  
 aes(x, y, color = DID %in% dists\_not\_gender)) +  
 scale\_color\_manual(values = c("darkgrey", "white")) +  
 geom\_text\_repel(data = child\_ica\_dummy %>%   
 filter(DID %in% dists\_not\_gender, !is.na(x)) %>%   
 summarize(x = unique(x),  
 y = unique(y),  
 DID = unique(DID),  
 DNAME = unique(DNAME)),  
 aes(x, y, label = DNAME), family = "serif", force = 50, color = "black", label.padding = .1, box.padding = .5) +  
 ggtitle("") +  
 guides(color = FALSE) +  
 theme(text = element\_text(family = "serif"),title = element\_text(family = "serif"),aspect.ratio = 1) +  
 xlab("") +  
 ylab("")

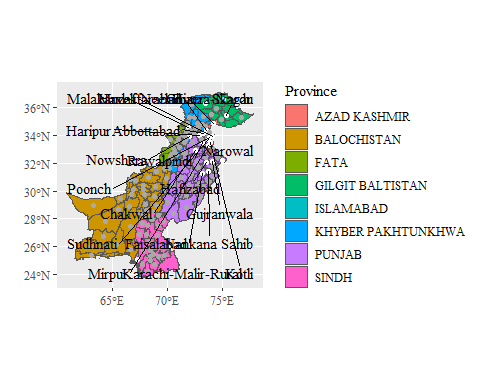
## Warning: Problem with `mutate()` input `centroid`.  
## x st\_centroid does not give correct centroids for longitude/latitude data  
## i Input `centroid` is `st\_centroid(geometry)`.

## Warning in st\_centroid.sfc(geometry): st\_centroid does not give correct  
## centroids for longitude/latitude data

## Warning: Ignoring unknown parameters: ratio

## Coordinate system already present. Adding new coordinate system, which will replace the existing one.

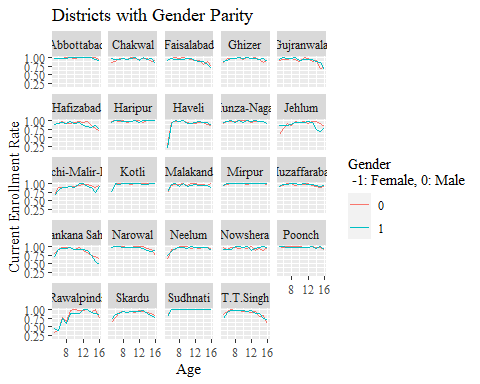
## Warning: Ignoring unknown parameters: label.padding



ggsave("map\_gender.png", unit = "cm", dpi = 300)

## Saving 12.7 x 10.2 cm image

child\_ica\_dummy %>%   
 filter(Age >= 5, DID %in% dists\_not\_gender, !is.na(Household.Ownership)) %>%   
 group\_by(DID, Age, Gender) %>%   
 mutate(rate = mean(C003 == 3)) %>%   
 ggplot(aes(Age, rate, color = factor(Gender))) +  
 geom\_line() +  
 facet\_wrap(.~DNAME) +  
 labs(title = "Districts with Gender Parity", color = "Gender \n -1: Female, 0: Male") +  
 xlab("Age") +  
 ylab("Current Enrollment Rate") +  
 theme(text = element\_text(family = "serif"))



ggsave("dists\_gender.png", unit = "cm", dpi = 300)

## Saving 12.7 x 10.2 cm image

### Model 2 [NofChild + ParentsBothEdu]

#### Whole Pak. GLM Age >= 5 School.Enrollment ~ Num.of.Child + Parents.Edu2 + relevel(DID, ref = “266”)

glm\_child <- glm(School.Enrollment ~ Num.of.Child + Parents.Edu2 + relevel(DID, ref = "266"), family = "binomial", data = child\_ica\_dummy %>% filter(Age >= 5))  
  
glm\_child %>% summary()

##   
## Call:  
## glm(formula = School.Enrollment ~ Num.of.Child + Parents.Edu2 +   
## relevel(DID, ref = "266"), family = "binomial", data = child\_ica\_dummy %>%   
## filter(Age >= 5))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.2006 0.1767 0.4684 0.6991 1.5750   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 3.520329 0.158012 22.279 < 2e-16 \*\*\*  
## Num.of.Child -0.108566 0.004001 -27.137 < 2e-16 \*\*\*  
## Parents.Edu21 0.673153 0.017490 38.488 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")146 -0.894525 0.199994 -4.473 7.72e-06 \*\*\*  
## relevel(DID, ref = "266")147 -1.486988 0.174559 -8.519 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")148 -0.927000 0.180863 -5.125 2.97e-07 \*\*\*  
## relevel(DID, ref = "266")149 -1.657619 0.170144 -9.742 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")150 -1.611451 0.169864 -9.487 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")151 -0.850554 0.187403 -4.539 5.66e-06 \*\*\*  
## relevel(DID, ref = "266")152 -0.736904 0.198088 -3.720 0.000199 \*\*\*  
## relevel(DID, ref = "266")153 -1.757663 0.171919 -10.224 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")154 -1.558946 0.174346 -8.942 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")155 -1.687751 0.168990 -9.987 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")156 -2.149207 0.183952 -11.684 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")157 -1.116234 0.183266 -6.091 1.12e-09 \*\*\*  
## relevel(DID, ref = "266")158 -1.164593 0.179292 -6.496 8.28e-11 \*\*\*  
## relevel(DID, ref = "266")159 -2.194987 0.173782 -12.631 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")160 -1.727416 0.171798 -10.055 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")161 -2.321023 0.164894 -14.076 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")162 -0.615197 0.207127 -2.970 0.002977 \*\*   
## relevel(DID, ref = "266")163 -1.041756 0.197676 -5.270 1.36e-07 \*\*\*  
## relevel(DID, ref = "266")164 -1.612564 0.183059 -8.809 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")165 -1.871483 0.172123 -10.873 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")166 -1.696227 0.172708 -9.821 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")167 -0.427180 0.203599 -2.098 0.035892 \*   
## relevel(DID, ref = "266")169 -2.707663 0.165687 -16.342 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")170 -1.261089 0.180463 -6.988 2.79e-12 \*\*\*  
## relevel(DID, ref = "266")171 -1.201221 0.182507 -6.582 4.65e-11 \*\*\*  
## relevel(DID, ref = "266")172 -1.395792 0.176964 -7.887 3.08e-15 \*\*\*  
## relevel(DID, ref = "266")173 -1.148310 0.178896 -6.419 1.37e-10 \*\*\*  
## relevel(DID, ref = "266")174 -1.728883 0.174139 -9.928 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")175 -2.323555 0.168330 -13.804 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")176 -0.402304 0.229868 -1.750 0.080093 .   
## relevel(DID, ref = "266")177 -1.774395 0.169762 -10.452 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")178 -0.095423 0.213153 -0.448 0.654389   
## relevel(DID, ref = "266")179 -1.167812 0.181549 -6.433 1.26e-10 \*\*\*  
## relevel(DID, ref = "266")180 -1.455521 0.175018 -8.316 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")181 -1.157964 0.183823 -6.299 2.99e-10 \*\*\*  
## relevel(DID, ref = "266")182 -2.330947 0.167045 -13.954 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")183 -1.580630 0.174456 -9.060 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")184 -2.166575 0.166939 -12.978 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")185 -0.252103 0.192314 -1.311 0.189895   
## relevel(DID, ref = "266")186 -2.397592 0.166642 -14.388 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")187 -2.180873 0.168413 -12.950 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")188 -1.960075 0.169116 -11.590 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")189 -1.158025 0.179453 -6.453 1.10e-10 \*\*\*  
## relevel(DID, ref = "266")190 -1.840270 0.175813 -10.467 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")191 -1.662123 0.176081 -9.440 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")192 -1.354327 0.178617 -7.582 3.39e-14 \*\*\*  
## relevel(DID, ref = "266")193 -0.494851 0.186674 -2.651 0.008028 \*\*   
## relevel(DID, ref = "266")194 -2.649108 0.166010 -15.958 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")195 -2.817870 0.167098 -16.864 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")196 -1.051247 0.193536 -5.432 5.58e-08 \*\*\*  
## relevel(DID, ref = "266")197 -2.354340 0.167589 -14.048 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")198 -2.650454 0.165891 -15.977 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")199 0.388040 0.215961 1.797 0.072366 .   
## relevel(DID, ref = "266")200 -2.383548 0.169121 -14.094 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")202 -2.265986 0.168120 -13.478 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")203 -2.452719 0.165223 -14.845 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")204 -1.694221 0.171660 -9.870 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")205 -2.503575 0.164981 -15.175 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")206 -2.494926 0.165114 -15.110 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")207 -3.065913 0.165283 -18.550 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")208 -2.350498 0.166279 -14.136 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")209 -1.686270 0.177545 -9.498 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")210 -2.668584 0.164427 -16.230 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")211 -1.809813 0.175482 -10.313 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")212 -2.428776 0.167017 -14.542 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")213 -2.528462 0.165453 -15.282 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")214 -2.862901 0.165609 -17.287 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")215 -1.900637 0.165303 -11.498 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")216 -2.674631 0.166984 -16.017 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")217 -2.631309 0.166486 -15.805 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")218 -2.396094 0.164395 -14.575 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")219 -2.573823 0.166915 -15.420 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")220 -2.933471 0.163097 -17.986 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")221 -3.385039 0.166391 -20.344 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")222 -2.649271 0.164190 -16.135 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")223 -2.343615 0.173022 -13.545 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")224 -2.950431 0.165513 -17.826 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")225 -2.770432 0.166010 -16.688 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")226 -2.457017 0.168486 -14.583 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")227 -2.510842 0.164794 -15.236 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")228 -3.442110 0.167163 -20.591 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")229 -2.802261 0.168027 -16.677 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")230 -2.958019 0.165646 -17.857 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")231 -2.547697 0.164957 -15.445 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")232 -1.313047 0.174697 -7.516 5.64e-14 \*\*\*  
## relevel(DID, ref = "266")233 -3.344977 0.168582 -19.842 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")234 -2.496392 0.171409 -14.564 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")235 -1.665329 0.170503 -9.767 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")236 -1.908919 0.172092 -11.092 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")237 -1.459497 0.176433 -8.272 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")238 1.205179 0.296470 4.065 4.80e-05 \*\*\*  
## relevel(DID, ref = "266")239 -1.137883 0.179953 -6.323 2.56e-10 \*\*\*  
## relevel(DID, ref = "266")240 -1.666088 0.172233 -9.673 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")241 -2.834317 0.166341 -17.039 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")242 -1.889061 0.171884 -10.990 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")243 0.310370 0.304382 1.020 0.307884   
## relevel(DID, ref = "266")244 -1.210256 0.182748 -6.623 3.53e-11 \*\*\*  
## relevel(DID, ref = "266")245 -0.327693 0.204173 -1.605 0.108499   
## relevel(DID, ref = "266")246 -2.366954 0.168223 -14.070 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")247 -2.212891 0.174037 -12.715 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")248 -1.682584 0.167708 -10.033 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")249 -0.636100 0.198727 -3.201 0.001370 \*\*   
## relevel(DID, ref = "266")250 -1.612341 0.173098 -9.315 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")251 -1.365377 0.177241 -7.704 1.32e-14 \*\*\*  
## relevel(DID, ref = "266")252 -1.537691 0.172906 -8.893 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")253 -1.326204 0.180424 -7.350 1.97e-13 \*\*\*  
## relevel(DID, ref = "266")254 -2.578461 0.182012 -14.166 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")255 -1.612984 0.177397 -9.092 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")256 -0.513078 0.184321 -2.784 0.005376 \*\*   
## relevel(DID, ref = "266")257 -0.537871 0.218214 -2.465 0.013706 \*   
## relevel(DID, ref = "266")258 -2.405340 0.165832 -14.505 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")259 -0.115540 0.221457 -0.522 0.601861   
## relevel(DID, ref = "266")260 -0.579921 0.184763 -3.139 0.001697 \*\*   
## relevel(DID, ref = "266")261 -3.076042 0.164442 -18.706 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")262 -1.115624 0.176774 -6.311 2.77e-10 \*\*\*  
## relevel(DID, ref = "266")263 -0.620619 0.187724 -3.306 0.000946 \*\*\*  
## relevel(DID, ref = "266")264 -0.828557 0.180846 -4.582 4.62e-06 \*\*\*  
## relevel(DID, ref = "266")265 -0.207445 0.206173 -1.006 0.314333   
## relevel(DID, ref = "266")267 -0.210390 0.202406 -1.039 0.298598   
## relevel(DID, ref = "266")268 -1.655485 0.181349 -9.129 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")269 0.944794 0.303461 3.113 0.001849 \*\*   
## relevel(DID, ref = "266")270 -0.223894 0.200294 -1.118 0.263641   
## relevel(DID, ref = "266")271 0.500943 0.258984 1.934 0.053081 .   
## relevel(DID, ref = "266")272 0.924923 0.296469 3.120 0.001810 \*\*   
## relevel(DID, ref = "266")273 -0.223172 0.213170 -1.047 0.295136   
## relevel(DID, ref = "266")274 0.209940 0.275292 0.763 0.445696   
## relevel(DID, ref = "266")275 -1.881364 0.172144 -10.929 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")276 1.248094 0.296282 4.213 2.53e-05 \*\*\*  
## relevel(DID, ref = "266")277 -1.056220 0.227904 -4.634 3.58e-06 \*\*\*  
## relevel(DID, ref = "266")278 -1.042344 0.168823 -6.174 6.65e-10 \*\*\*  
## relevel(DID, ref = "266")279 -2.591932 0.166779 -15.541 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")280 -1.311645 0.174437 -7.519 5.51e-14 \*\*\*  
## relevel(DID, ref = "266")281 -0.710741 0.181640 -3.913 9.12e-05 \*\*\*  
## relevel(DID, ref = "266")282 -1.278387 0.183050 -6.984 2.87e-12 \*\*\*  
## relevel(DID, ref = "266")284 -1.551173 0.172563 -8.989 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")287 -1.479697 0.177404 -8.341 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")289 -1.870175 0.167735 -11.150 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")290 -1.813008 0.174322 -10.400 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")315 -1.346731 0.192187 -7.007 2.43e-12 \*\*\*  
## relevel(DID, ref = "266")316 -2.168195 0.169502 -12.792 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")318 -2.756596 0.170023 -16.213 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")319 -3.225241 0.164767 -19.575 < 2e-16 \*\*\*  
## relevel(DID, ref = "266")320 -2.653372 0.170133 -15.596 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 209345 on 204954 degrees of freedom  
## Residual deviance: 180925 on 204809 degrees of freedom  
## AIC: 181217  
##   
## Number of Fisher Scoring iterations: 7

glm\_child %>%   
 summary() %>%   
 .$coefficient %>%   
 as.data.frame() %>%   
 select(`Pr(>|z|)`) %>%   
 filter(`Pr(>|z|)` > 0.05) %>%   
 row.names() %>%   
 str\_split("\\)") %>%   
 as.data.frame() %>%   
 gather() %>%   
 select(value) %>%   
 .[seq(2,24,2),] %>%   
 as.numeric()

## [1] 176 178 185 199 243 245 259 265 267 270 271 273

#### Each Dist. GLM Age >= 5 School.Enrollment ~ Num.of.Child + Parents.Edu2

each\_dist <- sapply(as.numeric(as.character(unique(child\_ica\_dummy$DID))), function(id){  
 child\_ica\_dummy <- child\_ica\_dummy %>% filter(DID == id)  
 glm\_child <- glm(School.Enrollment ~ Num.of.Child + Parents.Edu2,   
 family = "binomial", data = child\_ica\_dummy %>% filter(Age >= 5))  
 temp\_hoslem <- hoslem.test(x = glm\_child$y, y = fitted(glm\_child))  
 data.frame(DID = id, hoslem\_p\_value = temp\_hoslem$p.value)  
}) %>% t()

each\_dist <- each\_dist %>%   
 as.data.frame()   
each\_dist$DID <- each\_dist %>%   
 pull(DID) %>%   
 as.numeric()  
each\_dist$hoslem\_p\_value <- each\_dist %>%   
 pull(hoslem\_p\_value) %>%   
 as.numeric()  
each\_dist

## DID hoslem\_p\_value  
## 1 146 5.078609e-01  
## 2 147 9.705539e-01  
## 3 148 2.283193e-01  
## 4 149 2.557160e-01  
## 5 150 4.143391e-01  
## 6 151 5.765446e-01  
## 7 152 5.531868e-01  
## 8 153 2.932811e-01  
## 9 154 8.460617e-01  
## 10 155 9.997077e-01  
## 11 156 3.882100e-01  
## 12 157 2.044527e-01  
## 13 158 4.009867e-01  
## 14 159 9.504181e-01  
## 15 160 1.638910e-01  
## 16 161 1.762149e-01  
## 17 162 9.657634e-01  
## 18 163 7.729692e-02  
## 19 164 8.075623e-01  
## 20 165 6.205887e-03  
## 21 166 6.525883e-01  
## 22 167 8.659659e-01  
## 23 169 9.580809e-01  
## 24 170 2.583724e-01  
## 25 171 3.176283e-02  
## 26 172 7.732572e-02  
## 27 173 8.005250e-03  
## 28 174 1.300363e-01  
## 29 175 9.883908e-01  
## 30 176 9.322484e-01  
## 31 177 4.857311e-01  
## 32 178 6.281153e-01  
## 33 179 1.868784e-01  
## 34 180 6.227748e-01  
## 35 181 7.627900e-01  
## 36 182 9.244111e-01  
## 37 183 4.378609e-03  
## 38 184 7.527982e-01  
## 39 185 1.269173e-01  
## 40 186 6.861957e-01  
## 41 187 NaN  
## 42 188 6.798774e-01  
## 43 189 3.549348e-01  
## 44 190 4.558728e-01  
## 45 191 9.614523e-01  
## 46 192 7.342397e-01  
## 47 193 3.817516e-05  
## 48 194 6.191901e-01  
## 49 195 4.219941e-01  
## 50 196 4.013176e-03  
## 51 197 9.225834e-01  
## 52 198 4.027120e-01  
## 53 199 3.190054e-01  
## 54 200 3.144582e-01  
## 55 202 NaN  
## 56 203 5.788088e-07  
## 57 204 7.970467e-01  
## 58 315 4.681664e-01  
## 59 316 9.253282e-03  
## 60 320 5.909862e-02  
## 61 205 9.819220e-01  
## 62 206 9.861930e-01  
## 63 207 9.325272e-01  
## 64 208 9.355852e-01  
## 65 209 8.493397e-01  
## 66 210 2.599405e-01  
## 67 211 9.814336e-01  
## 68 212 9.588182e-01  
## 69 213 8.849036e-01  
## 70 214 8.231951e-01  
## 71 215 6.468684e-01  
## 72 216 9.396100e-01  
## 73 217 8.282422e-01  
## 74 218 5.019456e-01  
## 75 219 9.248346e-01  
## 76 220 1.034852e-01  
## 77 221 9.170773e-01  
## 78 222 2.889351e-01  
## 79 223 8.395059e-01  
## 80 224 8.478827e-01  
## 81 225 9.966807e-01  
## 82 226 1.922047e-01  
## 83 227 4.715612e-01  
## 84 228 9.728302e-01  
## 85 229 1.000000e+00  
## 86 230 9.506553e-01  
## 87 231 9.992710e-01  
## 88 232 3.865323e-03  
## 89 233 7.629312e-01  
## 90 234 2.029032e-01  
## 91 318 8.656534e-01  
## 92 319 2.192380e-01  
## 93 235 1.815625e-01  
## 94 236 1.521464e-01  
## 95 237 1.865052e-01  
## 96 238 6.236704e-01  
## 97 239 1.795182e-01  
## 98 240 1.282194e-01  
## 99 241 5.726324e-05  
## 100 242 8.018237e-01  
## 101 243 9.499583e-01  
## 102 244 2.720158e-01  
## 103 245 9.952252e-01  
## 104 246 5.237763e-03  
## 105 247 5.387150e-01  
## 106 248 3.927352e-03  
## 107 249 5.023224e-01  
## 108 250 9.189316e-01  
## 109 251 9.401226e-01  
## 110 252 6.661516e-01  
## 111 253 9.387411e-01  
## 112 254 4.631390e-01  
## 113 255 1.986730e-01  
## 114 256 6.809872e-01  
## 115 257 1.684169e-03  
## 116 258 6.945218e-04  
## 117 259 8.446685e-01  
## 118 260 5.108894e-02  
## 119 261 8.481242e-01  
## 120 262 3.988636e-01  
## 121 263 9.127782e-01  
## 122 264 8.141610e-01  
## 123 265 9.733987e-01  
## 124 266 6.812400e-01  
## 125 267 1.644173e-01  
## 126 268 6.197229e-02  
## 127 269 9.914447e-01  
## 128 270 9.511241e-01  
## 129 271 9.567134e-01  
## 130 272 8.877388e-01  
## 131 273 3.964173e-01  
## 132 274 7.670534e-01  
## 133 275 1.565314e-02  
## 134 276 9.920801e-01  
## 135 277 NaN  
## 136 278 9.150707e-03  
## 137 279 6.343923e-01  
## 138 280 4.948066e-01  
## 139 281 9.726423e-01  
## 140 282 8.801284e-01  
## 141 284 7.330808e-01  
## 142 287 2.616193e-01  
## 143 289 5.450667e-02  
## 144 290 9.875596e-01

#### Dists with P value <= 0.05

each\_dist %>%   
 filter(hoslem\_p\_value <= 0.05)

## DID hoslem\_p\_value  
## 1 165 6.205887e-03  
## 2 171 3.176283e-02  
## 3 173 8.005250e-03  
## 4 183 4.378609e-03  
## 5 193 3.817516e-05  
## 6 196 4.013176e-03  
## 7 203 5.788088e-07  
## 8 316 9.253282e-03  
## 9 232 3.865323e-03  
## 10 241 5.726324e-05  
## 11 246 5.237763e-03  
## 12 248 3.927352e-03  
## 13 257 1.684169e-03  
## 14 258 6.945218e-04  
## 15 275 1.565314e-02  
## 16 278 9.150707e-03

dists\_not\_fit <- each\_dist %>%   
 filter(hoslem\_p\_value <= 0.05) %>%   
 pull(DID)

#### Not dists\_not\_fit. GLM Age >= 5 School.Enrollment ~ Num.of.Child + Parents.Edu2

glm\_child <- glm(School.Enrollment ~ Num.of.Child + Parents.Edu2, family = "binomial", data = child\_ica\_dummy %>% filter(Age >= 5, !DID %in% dists\_not\_fit))  
  
glm\_child %>% summary()

##   
## Call:  
## glm(formula = School.Enrollment ~ Num.of.Child + Parents.Edu2,   
## family = "binomial", data = child\_ica\_dummy %>% filter(Age >=   
## 5, !DID %in% dists\_not\_fit))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.3536 0.3823 0.6989 0.7799 1.1638   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.535753 0.016581 92.62 <2e-16 \*\*\*  
## Num.of.Child -0.125298 0.003818 -32.82 <2e-16 \*\*\*  
## Parents.Edu21 1.294502 0.017084 75.77 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 185933 on 180691 degrees of freedom  
## Residual deviance: 176980 on 180689 degrees of freedom  
## AIC: 176986  
##   
## Number of Fisher Scoring iterations: 5

##### Hosmer-Lemeshow

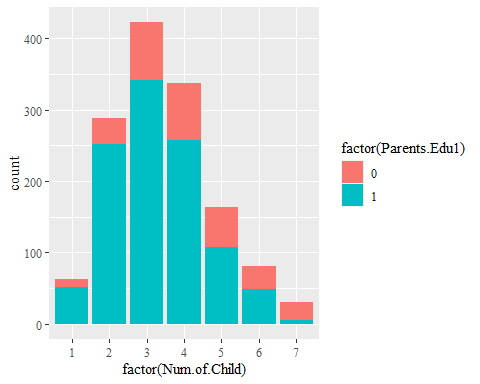
hoslem.test(x = glm\_child$y, y = fitted(glm\_child))

##   
## Hosmer and Lemeshow goodness of fit (GOF) test  
##   
## data: glm\_child$y, fitted(glm\_child)  
## X-squared = 51.996, df = 8, p-value = 1.686e-08

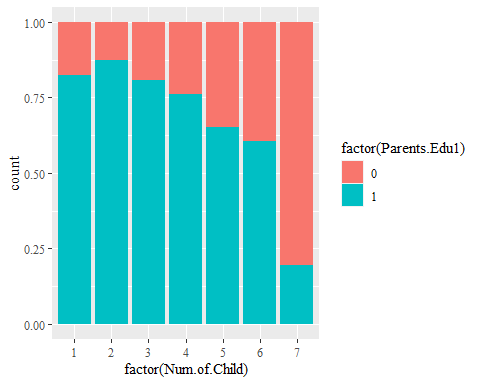
## Appendix

### Parents.Edu1 & Num.of.Child

child\_ica\_dummy %>%   
 filter(DID == 266, Age >= 5) %>%   
 ggplot(aes(factor(Num.of.Child), fill = factor(Parents.Edu1))) +  
 geom\_bar(stat = "count", position = "stack") +  
 theme(text = element\_text(family = "serif"))

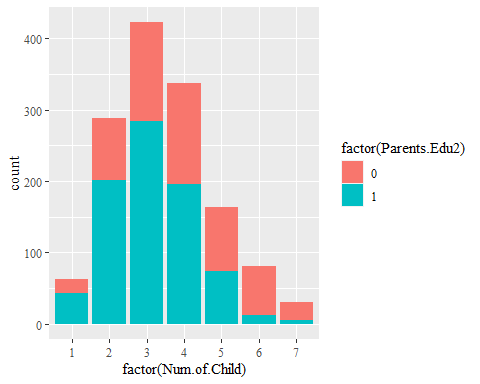


child\_ica\_dummy %>%   
 filter(DID == 266, Age >= 5) %>%   
 ggplot(aes(factor(Num.of.Child), fill = factor(Parents.Edu1))) +  
 geom\_bar(stat = "count", position = "fill") +  
 theme(text = element\_text(family = "serif"))

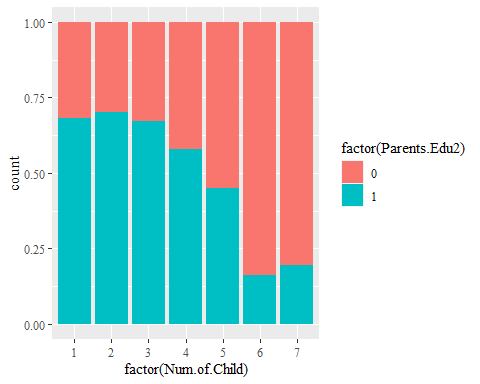


### Parents.Edu2 & Num.of.Child

child\_ica\_dummy %>%   
 filter(DID == 266, Age >= 5) %>%   
 ggplot(aes(factor(Num.of.Child), fill = factor(Parents.Edu2))) +  
 geom\_bar(stat = "count", position = "stack") +  
 theme(text = element\_text(family = "serif"))



child\_ica\_dummy %>%   
 filter(DID == 266, Age >= 5) %>%   
 ggplot(aes(factor(Num.of.Child), fill = factor(Parents.Edu2))) +  
 geom\_bar(stat = "count", position = "fill") +  
 theme(text = element\_text(family = "serif"))



child\_ica\_dummy %>%   
 filter(DID == 266, Age >= 5) %>%   
 ggplot(aes(Household.Ownership, fill = factor(Parents.Edu1))) +  
 geom\_bar(stat = "count", position = position\_stack())

