Analisis preliminar

Valentina Moreno y Laura Barragan

24/5/2021

## Pregunta de investigación

¿En qué medida los diferentes factores socioeconómicos influyeron en que niños entre los 5 y 17 años trabajaran en Bogotá durante el año 2011?

## Identificación de las variables relevantes

### Variable dependiente

Como variable dependiente se consideraron a los niños entre 5 y 17 años de edad que en las últimas cuatro semanas hicieron alguna diligencia para conseguir trabajo o instalar un negocio, utilizando la variable P6280 de la Enti\_niños, que tiene 2 categorías de respuesta:

1. si
2. No

### Variables independientes principales

La variable independiente de mayor interés para esta investigación acerca del trabajo infantil es la pobreza, para tener una idea de el nivel de pobreza se tiene en cuenta el nivel de ingresos, utilizando como proxy la variable P6047 de la encuesta Enti\_adultos, que pregunta *¿disminuyeron los ingresos del hogar en los últimos doce meses?* y tiene 2 categorias de respuesta 1.Si y 2.No

### Variable de control

Otras variables independientes que se van a incluir en este estudio para controlar por factores relevantes para el análisis son: la edad, el sexo, número de personas en el hogar y finalmente la asistencia de los niños a una institución de educacion formal.

Las variables extraidas de la base de datos a analizar son:

* P220: Sexo

1. Hombre
2. Mujer

* P6037: Escriba la edad que registraron en la geih
* P6008: Total de personas en el hogar
* P6170: ¿usted asiste actualmente a la escuela, colegio o universidad, recibiendo clases de educación formal?

1. Si
2. No

### Ponderaciones

Teniendo en cuenta que la *Encuesta Nacional de Trabajo Infantil (ENTI 2011)* se hizo a traves de un muestreo aleatorio se incluye la variable denominada *Factor de expansión ( Fex\_c)*, la cual se interpreta como la cantidad de personas en la población, que representa una persona en la muestra.

## Datos

library(tidyverse)

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

## v ggplot2 3.3.3 v purrr 0.3.4  
## v tibble 3.0.5 v dplyr 1.0.3  
## v tidyr 1.1.2 v stringr 1.4.0  
## v readr 1.4.0 v forcats 0.5.1

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

library(haven)  
library(knitr)  
library(naniar)

## Warning: package 'naniar' was built under R version 4.0.5

library(stargazer)

##   
## Please cite as:

## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.

## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer

library(GGally)

## Warning: package 'GGally' was built under R version 4.0.5

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

library(lindia)

## Warning: package 'lindia' was built under R version 4.0.5

library(ggeffects)

## Warning: package 'ggeffects' was built under R version 4.0.5

library(modelsummary)

## Warning: package 'modelsummary' was built under R version 4.0.5

library(broom)  
library(car)

## Warning: package 'car' was built under R version 4.0.5

## Loading required package: carData

##   
## Attaching package: 'car'

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

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

library(sandwich)

## Warning: package 'sandwich' was built under R version 4.0.5

### Cargar las bases de datos

rm(list = ls())  
Base\_ninos <- read\_sav("Datos/Enti\_ninos.sav")  
Base\_adultos <- read\_sav("Datos/Enti\_adultos.sav")

### Adecuación de la base de datos para el análisis

Base\_ninos <- Base\_ninos %>%  
rename(Nino\_trabajo = "P6280", Nino\_sexo = "P220", Nino\_edad = "P6037", Nino\_personas\_hogar = "P6008", Nino\_educ = "P6170", Nino\_factor\_exp = Fex\_c)  
  
Ninos\_interes <- Base\_ninos %>%  
 select(starts\_with("Nino"), Directorio, Secuencia\_p, Orden)  
  
Base\_adultos <- Base\_adultos %>%  
rename(Adult\_disminucion\_ing = "P6047", Adult\_factor\_exp = Fex\_c)  
  
Adult\_interes <- Base\_adultos %>%  
 select(starts\_with("Adult"), Directorio, Secuencia\_p, Orden)  
  
Base\_completa <- left\_join(Ninos\_interes, Adult\_interes, by = c("Directorio", "Secuencia\_p", "Orden"))  
  
Base\_completa <- Base\_completa %>%  
 relocate("Directorio", "Secuencia\_p", "Orden","Nino\_trabajo", "Adult\_disminucion\_ing", "Nino\_personas\_hogar", "Nino\_sexo", "Nino\_edad", "Nino\_educ", "Adult\_factor\_exp", "Nino\_factor\_exp")

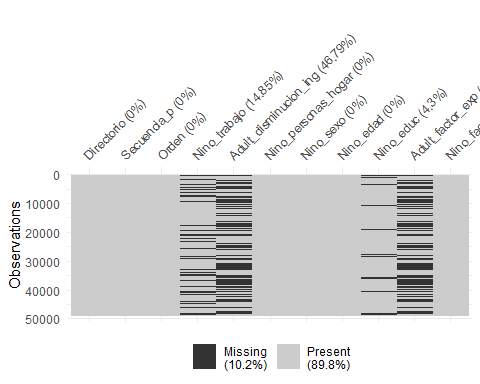
### Manejo de valores faltantes

En cuanto al manejo de datos faltantes para la variable llamada *Adult\_disminucion\_ing* que da cuenta de la disminucion de los ingresos en el hogar durante los ultimos 12 meses, en principio tenia dos categorias de respuesta, 1 para si y 2 para no; esta variable tenia un porcentaje de 46,79% de datos faltantes, sin embargo, luego de el analisis realizado, decidimos reemplazar esos valores faltantes por una tercera categoria de respuesta representada por el numero 3, esta nueva categoria de respuesta representara aquellas familias que no tuvieron ingresos durante los ultimos 12 meses.

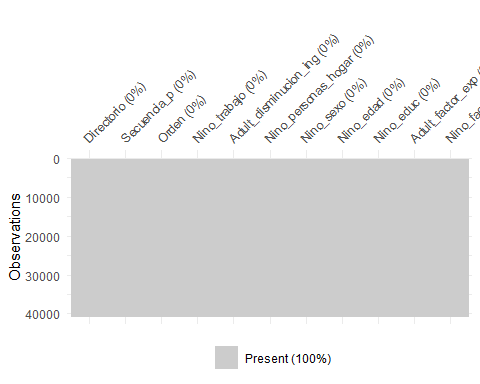
Por otro lado, en cuanto al *factor de expansion de la base de datos de los adultos*, que debe entenderse como la cantidad de personas en la población que representa una persona en la muestra tenia un porcentaje de datos faltantes del 46,79%. Dado el comportamiento de las observaciones de esta variable, se decidio rellenar esos valores faltantes con valores completos que estuvieran antes o despúes de los faltantes.

Finalmente, luego de modificar los valores faltantes de esas dos variables, la base de datos total quedo con un porcentaje de valores faltantes de 1,7%, por lo cual decidimos eliminar los valores faltantes de todas las variables, teniendo en cuenta que el procentaje de datos que quedaba disponible para realizar los análisis era del 98,3%.

faltantes\_antes <- vis\_miss(Base\_completa)  
faltantes\_antes



Base\_completa <- replace\_na(data = Base\_completa, replace = list(Adult\_disminucion\_ing = 3))  
Base\_completa <- fill(Base\_completa, Adult\_factor\_exp)  
Base\_completa <-na.omit(Base\_completa)  
  
faltantes\_final <- vis\_miss(Base\_completa)  
faltantes\_final



### Reconocimiento de las variables como numero o factor en R

class(Base\_completa$Directorio)

## [1] "numeric"

class(Base\_completa$Secuencia\_p)

## [1] "numeric"

class(Base\_completa$Orden)

## [1] "numeric"

class(Base\_completa$Nino\_trabajo)

## [1] "haven\_labelled" "vctrs\_vctr" "double"

class(Base\_completa$Nino\_jefe\_hogar)

## [1] "NULL"

class(Base\_completa$Nino\_sexo)

## [1] "haven\_labelled" "vctrs\_vctr" "double"

class(Base\_completa$Nino\_edad)

## [1] "numeric"

class(Base\_completa$Nino\_personas\_hogar)

## [1] "numeric"

class(Base\_completa$Nino\_educ)

## [1] "haven\_labelled" "vctrs\_vctr" "double"

class(Base\_completa$Nino\_factor\_exp)

## [1] "numeric"

class(Base\_completa$Adult\_factor\_exp)

## [1] "numeric"

class(Base\_completa$Adult\_disminucion\_ing)

## [1] "haven\_labelled" "vctrs\_vctr" "double"

Base\_completa <- Base\_completa %>%   
 mutate(across(.cols = c(4, 5, 7, 8, 10), as.factor))%>%   
 mutate(across(.cols = c(6, 9), as.numeric))

Base\_completa <- Base\_completa %>%   
 mutate(Nino\_sexo = recode(Nino\_sexo, "1=0; 2=1"))%>%  
 mutate(Nino\_educ = recode(Nino\_educ, "2=0; 1=1"))%>%  
 mutate(Nino\_trabajo = recode(Nino\_trabajo, "2=0; 1=1"))  
  
class(Base\_completa$Nino\_sexo)

## [1] "factor"

class(Base\_completa$Nino\_educ)

## [1] "numeric"

class(Base\_completa$Nino\_trabajo)

## [1] "factor"

## Plan de análisis

El objetivo central del siguiente documento es determinar en qué medida los diferentes factores socioeconómicos influyen en el trabajo infantil de los niños entre los 5 y 17 de edad en Bogotá durante el año 2011, para ello se realizará una descripción estadística de las variables y posteriormente se llevará a cabo una regresión lineal, utilizando como variable dependiente a los niños entre 5 y 17 años de edad que en las últimas cuatro semanas hicieron alguna diligencia para conseguir un trabajo o instalar un negocio, la variable explicativa principal es la disminución de ingresos en el último año como aproximación al nivel de pobreza del hogar; además, se agregaron 4 variables de control, el sexo, la edad, el número de personas en el hogar y la educacion de los niños para evitar un sesgo por variable omitida.

La primera regresión que se analizará tendrá en cuenta únicamente el efecto de la disminución de los ingresos en el hogar durante los últimos 12 meses sobre los niños que durante el último mes intentaron conseguir trabajo o instalar un negocio. La ecuación a estimar es la siguiente:

Donde *Y* representa la variable dependiente que viene dada por la busqueda de trabajo de los niños en las ultimas 4 semanas, *I* es la variable independiente principal que muestra la disminución de los ingresos en el hogar durante el último año y *e* es el término del error.

modelo\_1 <- lm(as.integer(Nino\_trabajo) ~ Adult\_disminucion\_ing, data = Base\_completa)  
summary(modelo\_1)

##   
## Call:  
## lm(formula = as.integer(Nino\_trabajo) ~ Adult\_disminucion\_ing,   
## data = Base\_completa)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.01385 -0.01385 -0.01212 -0.01208 0.98792   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.012e+00 1.348e-03 750.998 <2e-16 \*\*\*  
## Adult\_disminucion\_ing2 -4.289e-05 1.636e-03 -0.026 0.979   
## Adult\_disminucion\_ing3 1.725e-03 1.579e-03 1.093 0.274   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1129 on 40682 degrees of freedom  
## Multiple R-squared: 6.009e-05, Adjusted R-squared: 1.093e-05   
## F-statistic: 1.222 on 2 and 40682 DF, p-value: 0.2946

confint(modelo\_1)

## 2.5 % 97.5 %  
## (Intercept) 1.009478826 1.014761870  
## Adult\_disminucion\_ing2 -0.003249028 0.003163247  
## Adult\_disminucion\_ing3 -0.001369032 0.004819175

stargazer(modelo\_1, title = "Tabla 1", type = "text", out = "modelo1.txt")

##   
## Tabla 1  
## ==================================================  
## Dependent variable:   
## ---------------------------  
## as.integer(Nino\_trabajo)   
## --------------------------------------------------  
## Adult\_disminucion\_ing2 -0.00004   
## (0.002)   
##   
## Adult\_disminucion\_ing3 0.002   
## (0.002)   
##   
## Constant 1.012\*\*\*   
## (0.001)   
##   
## --------------------------------------------------  
## Observations 40,685   
## R2 0.0001   
## Adjusted R2 0.00001   
## Residual Std. Error 0.113 (df = 40682)   
## F Statistic 1.222 (df = 2; 40682)   
## ==================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Con la información de la Tabla 1 podemos inferir que, en primer lugar, no tiene ningún sentido interpretar el coeficiente teniendo en cuenta que la variable dependiente es binaria y la variable explicativa tiene mas de 2 categorias de respuesta, por lo tanto esta constante no se tiene en cuenta.

El coeficiente de la variable Adult\_disminucion\_ing2 indica que en los hogares en los que no diminuyo el ingreso en los ultimos 12 meses tienen 0,00004 menos probabilidad en comparación con los hogares en donde si disminuyo el ingreso en ese periodo de tiempo, de que los niños buscaran trabajo durante las últimas 4 semanas.

Además, el coeficiente de la variable Adult\_disminucion\_ing3 indica que en los hogares que no tuvieron ingresos en los ultimos 12 meses tienen 0,002 más probabilidad en comparación con los hogares en donde si disminuyo el ingreso en ese periodo de tiempo, de que los niños buscaran trabajo durante las últimas 4 semanas.

Por otra parte, podemos evidenciar que ninguno de los coeficientes de la variable Adult\_disminucion\_ing no son estadísticamente significativos, esto con un nivel de confianza del 95% y un valor p (0.979 y 0.274) mayor a 0,1 de significancia. Que no sean estadísticamente significativos indica que no se puede rechazar la hipótesis nula de que los coeficiente son iguales a 0, en favor de la hipótesis alternativa de que los coeficientes son diferente a 0.

Esto tambien se puede confirmar analizando los intervalos de confianza de las variables, ya que para todas las categorias de respuesta de la variable Adult\_disminucion\_ing el intervalo incluye el cero, por lo que no son estadisticamente significativos.

Adicionalmente, el (0,0001) indica que este modelo explica tan solo el 0.01% de la variación en Ninos\_trabajo.

La segunda regresión que se analizará tendrá en cuenta el efecto de la disminución de los ingresos durante los últimos 12 meses, el numero de personas en el hogar, el sexo, la edad que registraron en la encuesta y la asistencia del infante una institución de educación formal sobre la busqueda de trabajo por parte de los niños durante el último mes. La ecuación es la siguiente:

$Y = \_0 + \_1I + \_2P + \_3J + \_3S + \_4E + \_5A + e $

Donde las variables *Y, I y e* tienen el mismo significado que en la ecuacion anterior; P muestra el número de personas que viven en el hogar; J es una variable dicotómica que toma el valor de 1 cuando el niño tiene entre 5 y 17 años y vive con otros adultos y 0 cuando el adolescente tiene entre 12 y 17 años y es jefe del hogar; S es una variable dicotómica que toma el valor de 1 cuando el sexo es femenino y 0 cuando es masculino; E toma valores entre 5 y 17 y representa la edad que registraron los niños en el momento de la encuesta y finalmente, A es una variable dicotómica que toma el valor de 1 cuando el infante asiste a una institución de educación formal y 0 en el el caso contrario.

modelo\_2 <- lm(as.integer(Nino\_trabajo) ~ Adult\_disminucion\_ing + Nino\_personas\_hogar + Nino\_sexo + Nino\_edad + Nino\_educ, data = Base\_completa)  
summary(modelo\_2)

##   
## Call:  
## lm(formula = as.integer(Nino\_trabajo) ~ Adult\_disminucion\_ing +   
## Nino\_personas\_hogar + Nino\_sexo + Nino\_edad + Nino\_educ,   
## data = Base\_completa)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.12728 -0.00832 -0.00248 0.00057 1.00372   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.0609920 0.0636663 16.665 < 2e-16 \*\*\*  
## Adult\_disminucion\_ing2 -0.0002328 0.0015968 -0.146 0.88409   
## Adult\_disminucion\_ing3 -0.0003831 0.0015617 -0.245 0.80622   
## Nino\_personas\_hogar -0.0005704 0.0002789 -2.045 0.04087 \*   
## Nino\_sexo1 -0.0032958 0.0010936 -3.014 0.00258 \*\*   
## Nino\_edad5 -0.0021116 0.0636286 -0.033 0.97353   
## Nino\_edad6 -0.0003447 0.0636223 -0.005 0.99568   
## Nino\_edad7 -0.0004553 0.0636218 -0.007 0.99429   
## Nino\_edad8 0.0000941 0.0636220 0.001 0.99882   
## Nino\_edad9 0.0015975 0.0636218 0.025 0.97997   
## Nino\_edad10 0.0013619 0.0636209 0.021 0.98292   
## Nino\_edad11 0.0026716 0.0636205 0.042 0.96650   
## Nino\_edad12 0.0044303 0.0636218 0.070 0.94448   
## Nino\_edad13 0.0054915 0.0636229 0.086 0.93122   
## Nino\_edad14 0.0093539 0.0636244 0.147 0.88312   
## Nino\_edad15 0.0236836 0.0636255 0.372 0.70972   
## Nino\_edad16 0.0433587 0.0636278 0.681 0.49560   
## Nino\_edad17 0.0676569 0.0636338 1.063 0.28769   
## Nino\_educ -0.0567861 0.0025191 -22.542 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1101 on 40666 degrees of freedom  
## Multiple R-squared: 0.04802, Adjusted R-squared: 0.0476   
## F-statistic: 114 on 18 and 40666 DF, p-value: < 2.2e-16

stargazer(modelo\_2, title = "Tabla 2", type = "text", out = "modelo2.txt")

##   
## Tabla 2  
## ==================================================  
## Dependent variable:   
## ---------------------------  
## as.integer(Nino\_trabajo)   
## --------------------------------------------------  
## Adult\_disminucion\_ing2 -0.0002   
## (0.002)   
##   
## Adult\_disminucion\_ing3 -0.0004   
## (0.002)   
##   
## Nino\_personas\_hogar -0.001\*\*   
## (0.0003)   
##   
## Nino\_sexo1 -0.003\*\*\*   
## (0.001)   
##   
## Nino\_edad5 -0.002   
## (0.064)   
##   
## Nino\_edad6 -0.0003   
## (0.064)   
##   
## Nino\_edad7 -0.0005   
## (0.064)   
##   
## Nino\_edad8 0.0001   
## (0.064)   
##   
## Nino\_edad9 0.002   
## (0.064)   
##   
## Nino\_edad10 0.001   
## (0.064)   
##   
## Nino\_edad11 0.003   
## (0.064)   
##   
## Nino\_edad12 0.004   
## (0.064)   
##   
## Nino\_edad13 0.005   
## (0.064)   
##   
## Nino\_edad14 0.009   
## (0.064)   
##   
## Nino\_edad15 0.024   
## (0.064)   
##   
## Nino\_edad16 0.043   
## (0.064)   
##   
## Nino\_edad17 0.068   
## (0.064)   
##   
## Nino\_educ -0.057\*\*\*   
## (0.003)   
##   
## Constant 1.061\*\*\*   
## (0.064)   
##   
## --------------------------------------------------  
## Observations 40,685   
## R2 0.048   
## Adjusted R2 0.048   
## Residual Std. Error 0.110 (df = 40666)   
## F Statistic 113.952\*\*\* (df = 18; 40666)  
## ==================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Al igual que en el modelo anterior, no se analizara el valor del intercepto ya que no tiene ningun sentido interpretarlo.

El coeficiente de la variable Adult\_disminucion\_ing2 indica que en los hogares en los que no diminuyó el ingreso en los ultimos 12 meses tienen 0.000232 menos probabilidad en comparación con los hogares en donde si disminuyo el ingreso en ese periodo de tiempo, de que los niños buscaran trabajo durante las últimas 4 semanas, ceteris paribus.

Además, el coeficiente de la variable Adult\_disminucion\_ing3 cambio de sentido, indicando que en los hogares que no tuvieron ingresos en los ultimos 12 meses tienen 0.0003831 menos probabilidad en comparación con los hogares en donde si disminuyo el ingreso en ese periodo de tiempo, de que los niños buscaran trabajo durante las últimas 4 semanas, controlando por todo lo demás.

El coeficiente de la variable Nino\_personas\_hogar es -0.0005704, este coeficiente indica que si el número de personas que viven en el hogar aumenta en una unidad, la probabilidad de que los niños buscaran trabajo en el último mes disminuiría en 0.0005704 unidades, dejando constante las demás variables.

El coeficiente de la variable Nino\_sexo indica que, dejando todo lo demás constante, las mujeres tienen 0.003295 menos probabilidad que los hombres de haber buscado trabajo en las ultimas cuatro semanas.

El coeficiente de la variable Nino\_educ es -0.0567861, este coeficiente indica que los niños que reciben una educacion formal tienen 0.0567861 menos pobabilidad con respecto a los que no asisten al colegio o a la universidad de haber buscado trabajo en el último mes, ceteris paribus.

El coeficiente de la variable Nino\_edad es negativa para las edades 5, 6 y 7, pero es positiva para las edades de 8 a 17 años, indicando que los niños que tienen entre 5 y 7 años y asisten a una institución de educación formal tienen menos probabilidad en comparación con los que no asisten de haber buscado trabajo durante las últimas 4 semanas, manteniendo todo lo demás constante. Por su parte, los niños que tienen entre 8 y 17 años y asisten a una institución de educación formal tienen más probabilidad en comparación con los que no asisten de haber buscado trabajo durante las últimas 4 semanas, con todo lo demás constante.

De acuerdo con el valor p, al igual que en el modelo anterior,ninguno de los coeficientes de la variable Adult\_disminucion\_ing son estadísticamente significativos, al igual que ninguno de los intervalos de edades, ya que los valores p son mayores al 0,1, en otras palabras, no se puede rechazar la hipótesis nula de que los coeficiente son iguales a 0, en favor de la hipótesis alternativa de que los coeficientes son diferente a 0.

En contraste, los asteriscos al lado del valor p indican que los coeficientes del las variables de educación, sexo y número de personas en el hogar son estadisticamente significativos. La primera variable a un nivel de confianza del 99.9%, pues el valor p es menor a 0,001, el sexo y el numero de personas en el hogar a un nivel de confianza del 95% y del 90% respectivamente.

El (0.048) aumentó como es natural cuando se agregan variables en una regresión. En este caso, es mas acertado enfocarnos en el Ajustado, el cual ajusta el valor del para quitar el efecto del aumento natural que se da simplemente por agregar más variables al modelo. En este caso el Ajustado aumentó de 0,0001 a 0.048, lo cual indica que con las 4 variables adicionales, este nuevo modelo explica el 4.8% de la variación en Ninos\_trabajo.

Finalmente, se realizó una regresion incluyendo la interacción entre las variables *Nino\_sexo* y *Nino\_educ*, para evaluar en que proporción el efecto de ser mujer en *Y*, está mediado por la asistencia del niño a una institución de educación formal.

modelo\_3 <- lm(as.integer(Nino\_trabajo) ~ Adult\_disminucion\_ing + Nino\_personas\_hogar + Nino\_sexo + Nino\_edad + Nino\_educ + Nino\_sexo:Nino\_educ, data = Base\_completa)  
  
summary(modelo\_3)

##   
## Call:  
## lm(formula = as.integer(Nino\_trabajo) ~ Adult\_disminucion\_ing +   
## Nino\_personas\_hogar + Nino\_sexo + Nino\_edad + Nino\_educ +   
## Nino\_sexo:Nino\_educ, data = Base\_completa)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.14320 -0.00879 -0.00212 0.00054 1.00299   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.0765558 0.0636835 16.905 < 2e-16 \*\*\*  
## Adult\_disminucion\_ing2 -0.0001962 0.0015960 -0.123 0.9022   
## Adult\_disminucion\_ing3 -0.0004318 0.0015610 -0.277 0.7820   
## Nino\_personas\_hogar -0.0005637 0.0002788 -2.022 0.0432 \*   
## Nino\_sexo1 -0.0327277 0.0047828 -6.843 7.87e-12 \*\*\*  
## Nino\_edad5 -0.0025279 0.0635982 -0.040 0.9683   
## Nino\_edad6 -0.0006326 0.0635919 -0.010 0.9921   
## Nino\_edad7 -0.0008784 0.0635914 -0.014 0.9890   
## Nino\_edad8 -0.0002651 0.0635916 -0.004 0.9967   
## Nino\_edad9 0.0012484 0.0635914 0.020 0.9843   
## Nino\_edad10 0.0009882 0.0635905 0.016 0.9876   
## Nino\_edad11 0.0023874 0.0635900 0.038 0.9701   
## Nino\_edad12 0.0040604 0.0635914 0.064 0.9491   
## Nino\_edad13 0.0051724 0.0635925 0.081 0.9352   
## Nino\_edad14 0.0089367 0.0635940 0.141 0.8882   
## Nino\_edad15 0.0235605 0.0635951 0.370 0.7110   
## Nino\_edad16 0.0430476 0.0635974 0.677 0.4985   
## Nino\_edad17 0.0679717 0.0636033 1.069 0.2852   
## Nino\_educ -0.0728886 0.0035818 -20.350 < 2e-16 \*\*\*  
## Nino\_sexo1:Nino\_educ 0.0310444 0.0049113 6.321 2.63e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1101 on 40665 degrees of freedom  
## Multiple R-squared: 0.04895, Adjusted R-squared: 0.04851   
## F-statistic: 110.2 on 19 and 40665 DF, p-value: < 2.2e-16

stargazer(modelo\_3, title = "Tabla 3", type = "text", out = "modelo3.txt")

##   
## Tabla 3  
## ==================================================  
## Dependent variable:   
## ---------------------------  
## as.integer(Nino\_trabajo)   
## --------------------------------------------------  
## Adult\_disminucion\_ing2 -0.0002   
## (0.002)   
##   
## Adult\_disminucion\_ing3 -0.0004   
## (0.002)   
##   
## Nino\_personas\_hogar -0.001\*\*   
## (0.0003)   
##   
## Nino\_sexo1 -0.033\*\*\*   
## (0.005)   
##   
## Nino\_edad5 -0.003   
## (0.064)   
##   
## Nino\_edad6 -0.001   
## (0.064)   
##   
## Nino\_edad7 -0.001   
## (0.064)   
##   
## Nino\_edad8 -0.0003   
## (0.064)   
##   
## Nino\_edad9 0.001   
## (0.064)   
##   
## Nino\_edad10 0.001   
## (0.064)   
##   
## Nino\_edad11 0.002   
## (0.064)   
##   
## Nino\_edad12 0.004   
## (0.064)   
##   
## Nino\_edad13 0.005   
## (0.064)   
##   
## Nino\_edad14 0.009   
## (0.064)   
##   
## Nino\_edad15 0.024   
## (0.064)   
##   
## Nino\_edad16 0.043   
## (0.064)   
##   
## Nino\_edad17 0.068   
## (0.064)   
##   
## Nino\_educ -0.073\*\*\*   
## (0.004)   
##   
## Nino\_sexo1:Nino\_educ 0.031\*\*\*   
## (0.005)   
##   
## Constant 1.077\*\*\*   
## (0.064)   
##   
## --------------------------------------------------  
## Observations 40,685   
## R2 0.049   
## Adjusted R2 0.049   
## Residual Std. Error 0.110 (df = 40665)   
## F Statistic 110.161\*\*\* (df = 19; 40665)  
## ==================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Al igual que en el modelo anterior, no se analizara el valor del intercepto ya que no tiene ningun sentido interpretarlo.

Los coeficientes de las variables Adult\_disminucion\_ing siguen y Nino\_edad siguen sin ser significativos y con los mismos signos, excepto para la edad de 8 años, que ahora tiene signo negativo, la probabilidad de afectar la variable dependiente cambió en una proporcion mínima.

De igual forma, los coeficientes de la variable de sexo, número de personas en el hogar, edad y educacion de los niños varian muy poco y continuan teniendo el mismo nivel de significancia y el mismo sentido.

En este modelo es importante analizar el coeficiente , que nos muestra la interacción entre el sexo y la asistencia de los niños a una institución de educación formal. De acuerdo con los resultados de esta regresión las mujeres tienen, en promedio, 0.032727 menos probabilidad que los hombres de haber buscado trabajo en el último mes. La interacción Nino\_sexo:Nino\_educ indica que para las mujeres en que asisten a una institución de educacion formal es 0.031044 más probable que para los hombres que asisten al colegio o la universidad haber buscado trabajo durante el último mes, además este coeficiente es estadísticamente significativo, con un nivel de confianza del 99.9%, pues el valor p es menor a 0,001.

Nuevamente el (0.048) se mantuvo constante. Como se habia mencionado anteriormente, es mas acertado enfocarnos en el Ajustado, que en este caso aumentó de 0.00001093 a 0.048 a 0.04851, lo cual indica que incluyendo la interaccion entre el sexo y la educacion de los niños, el nuevo modelo explica el 4.85% de la variación en Ninos\_trabajo, un 0.051% más en comparación con el modelo 2.

### Revicion de los principales supestos de los modelos estimados

Los 5 supuestos claves de la regresión son:

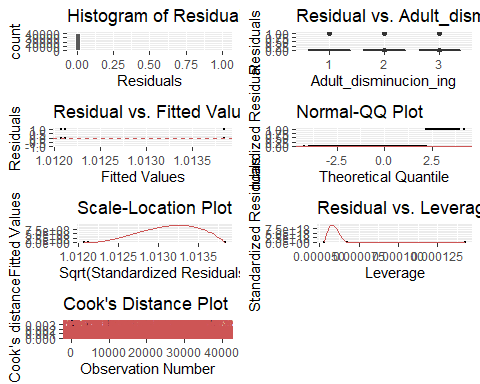
1. *Normalidad*: Como se puede evidenciar en las gráficas llamadas Normal\_QQ Plot, en ambos modelos los residuos estan aproximadamente sobre la línea, excepto por los valores ajustados más altos por lo cual es posible concluir que existe un comportamiento asimetrico en de los residuos, es necesario realizar pruebas estadísticas de normalidad, transformaciones de las variables involucradas, o utilizar modelos más complejos que consideren la falta de normalidad (GLM).
2. *Independencia*: A través de la prueba de Durbin\_Watson se observó que para el primer modelo el valor del estadistico DW es 1.905978 y para el modelo 2 es 1.905629, es decir, que no existen problemas de autocorrelaión para ningún modelo.
3. *Linealidad*: La linealidad se evalúa en la gráfica de residuos versus valores ajustados (fitted). En esta gráfica no debería identificarse una comportamiento sistemático que sugiera que los residuos varían según los valores ajustados. En ambos casos, ese supuesto no pareciera cumplirse, teniendo en cuenta que hay un comportamiento sistemático esto sugiere que la relación no es líneal y que debe emplearse algún polinomio en el modelo, o que hay alguna variable omitida en el modelo.
4. *Homocedasticidad*: La homocedasticidad se evalúa en el gráfico Scale Location. Allí no se debería observar ningún comportamiento sistemático en la raíz cuadrada de los residuos estandarizados para los diferentes valores ajustados. En ambos casos, este supuesto parece cumplirse excepto para los valores ajustados bajos.
5. *Multicolinealidad*: La multicolinealidad se puede evaluar calculando el factor de inflación de la varianza, si la raíz cuadrada del factor de inflación de la varianza es mayor a 2, esto sugiere que hay un problema de multicolinealidad. Para el caso del modelo 1 no es necesario aplicar esta prueba, ya que es un modelo de regresion lineal simple. En el modelo 2 se puede evidenciar que la raiz cuadrada del factor de inflación de la varianza siempre es menor a 2, por lo tanto no hay multicolinealidad.

* Modelo 1

gg\_diagnose(modelo\_1)

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## `geom\_smooth()` using formula 'y ~ x'  
## `geom\_smooth()` using formula 'y ~ x'



durbinWatsonTest(modelo\_1)

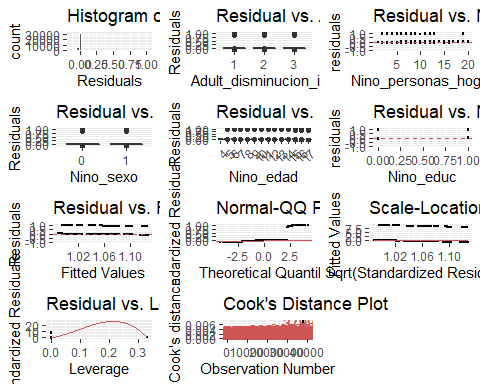
## lag Autocorrelation D-W Statistic p-value  
## 1 0.04701096 1.905978 0  
## Alternative hypothesis: rho != 0

* Modelo 2

gg\_diagnose(modelo\_2)

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## `geom\_smooth()` using formula 'y ~ x'  
## `geom\_smooth()` using formula 'y ~ x'



durbinWatsonTest(modelo\_2)

## lag Autocorrelation D-W Statistic p-value  
## 1 0.04718524 1.905629 0  
## Alternative hypothesis: rho != 0

vif(modelo\_2)

## GVIF Df GVIF^(1/(2\*Df))  
## Adult\_disminucion\_ing 1.060049 2 1.014686  
## Nino\_personas\_hogar 1.042379 1 1.020970  
## Nino\_sexo 1.002441 1 1.001220  
## Nino\_edad 1.084564 13 1.003127  
## Nino\_educ 1.056322 1 1.027775

sqrt(vif(modelo\_2))

## GVIF Df GVIF^(1/(2\*Df))  
## Adult\_disminucion\_ing 1.029587 1.414214 1.007316  
## Nino\_personas\_hogar 1.020970 1.000000 1.010430  
## Nino\_sexo 1.001220 1.000000 1.000610  
## Nino\_edad 1.041424 3.605551 1.001562  
## Nino\_educ 1.027775 1.000000 1.013792

## Resumen de los resultados

Para lograr visualizar de manera mas clara las diferencias entre los tres modelos que se analizaron, se realizo la Tabla 4, donde se muestran los cambios en los coeficientes para cada ecuación estimada.

stargazer(modelo\_1, modelo\_2, modelo\_3, title = "Tabla 4", type = "text", out = "modelos.txt")

##   
## Tabla 4  
## ====================================================================================================  
## Dependent variable:   
## -----------------------------------------------------------------------------  
## as.integer(Nino\_trabajo)   
## (1) (2) (3)   
## ----------------------------------------------------------------------------------------------------  
## Adult\_disminucion\_ing2 -0.00004 -0.0002 -0.0002   
## (0.002) (0.002) (0.002)   
##   
## Adult\_disminucion\_ing3 0.002 -0.0004 -0.0004   
## (0.002) (0.002) (0.002)   
##   
## Nino\_personas\_hogar -0.001\*\* -0.001\*\*   
## (0.0003) (0.0003)   
##   
## Nino\_sexo1 -0.003\*\*\* -0.033\*\*\*   
## (0.001) (0.005)   
##   
## Nino\_edad5 -0.002 -0.003   
## (0.064) (0.064)   
##   
## Nino\_edad6 -0.0003 -0.001   
## (0.064) (0.064)   
##   
## Nino\_edad7 -0.0005 -0.001   
## (0.064) (0.064)   
##   
## Nino\_edad8 0.0001 -0.0003   
## (0.064) (0.064)   
##   
## Nino\_edad9 0.002 0.001   
## (0.064) (0.064)   
##   
## Nino\_edad10 0.001 0.001   
## (0.064) (0.064)   
##   
## Nino\_edad11 0.003 0.002   
## (0.064) (0.064)   
##   
## Nino\_edad12 0.004 0.004   
## (0.064) (0.064)   
##   
## Nino\_edad13 0.005 0.005   
## (0.064) (0.064)   
##   
## Nino\_edad14 0.009 0.009   
## (0.064) (0.064)   
##   
## Nino\_edad15 0.024 0.024   
## (0.064) (0.064)   
##   
## Nino\_edad16 0.043 0.043   
## (0.064) (0.064)   
##   
## Nino\_edad17 0.068 0.068   
## (0.064) (0.064)   
##   
## Nino\_educ -0.057\*\*\* -0.073\*\*\*   
## (0.003) (0.004)   
##   
## Nino\_sexo1:Nino\_educ 0.031\*\*\*   
## (0.005)   
##   
## Constant 1.012\*\*\* 1.061\*\*\* 1.077\*\*\*   
## (0.001) (0.064) (0.064)   
##   
## ----------------------------------------------------------------------------------------------------  
## Observations 40,685 40,685 40,685   
## R2 0.0001 0.048 0.049   
## Adjusted R2 0.00001 0.048 0.049   
## Residual Std. Error 0.113 (df = 40682) 0.110 (df = 40666) 0.110 (df = 40665)   
## F Statistic 1.222 (df = 2; 40682) 113.952\*\*\* (df = 18; 40666) 110.161\*\*\* (df = 19; 40665)  
## ====================================================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

De los resultados obtenidos en las regresiones anteriormente realizadas podemos concluir principalmente que el modelo explica en muy baja proporción la variabilidad de los datos, ya que el modelo explica tan solo el 3% la variación de nuestra variable dependiente que es si los niños buscaron trabajo en las 4 semanas anteriores a la encuesta o no. Lo que nos muestra que la intuición que se tuvo en la elección de las variables independientes que son factores socio-económicos, para la creación del modelo no fue la más acertada, pues realmente no nos permite identificar con certeza las causales del trabajo en niños de 5 a 17 años en Colombia para el 2011.

Con la Tabla 4 podemos ver que la diferencia entre los valores de los coeficientes obtenidos en los 3 modelos estimados no es muy grande, incluso el modelo 1 que solo utiliza la variable dependiente y la independiente principal se asemeja mucho en el intercepto y en dichas variables a los resultados obtenidos en modelo 2 y 3 para estas mismas variables. La mayor diferencia que se da es en la variable que representa el sexo de los niños encuestados en los modelos 2 y 3, también se evidencia una diferencia en la variable de educación de los niños encuestados igualmente en los modelos 2 y 3.