

SDS291 Group Project - Final Proposal

Tseegi Nyamdorj, Tina Chang, Jinwen Xiang, My My Tran

2023-05-08

Contents

Title: The Influence of Parenthood and Age on Adults Pursuing Higher Education	1
Abstract	1
Introduction	2
Methods	2
Data	2
Variables	2
Analysis	3
Exploratory Visualization	8
Discussion	9
References	10

Title: The Influence of Parenthood and Age on Adults Pursuing Higher Education

Abstract

The purpose of our analysis is to explore factors that influence the likelihood of an adult pursuing higher education. Using data from the National Surveys of College Graduates (NSCG), Recent College Graduates (NSRCG), and Doctorate Recipients (SDR) available in the IPUMS database, we explore the likelihood of an adult to pursue higher education is based on the number of children the adult has, as well as their age. We will compare three different logistic regression models with enrolled/graduated higher education in the last five years as the response variable. Through our research, we have concluded that a person's age and the number of children are significant enough to contribute to predicting the chances of an adult pursuing higher education.

Introduction

The aim of this project is to investigate the likelihood of adults obtaining higher education, with a specific focus on adults with children and the number of children they have, and the age of the adult in the United States. Previous research suggests that having children may affect a parent's pursuit of higher education. For instance, Kulp (2016) found that "mothers with doctorates tended to take longer to complete the degree and graduated at an older age than their non-parent counterparts¹. We find this topic to be particularly interesting because of the Ada Comstock program at Smith College, allowing non-traditional aged students to pursue higher education and understanding the significance that age and number of children have on parent adults in higher education.

Our primary hypothesis is that adults with no children or fewer children are more likely to pursue higher education in their adult years than adults with 3 or more children. Our secondary hypothesis is that older adults are less likely to have attended higher education within the last 5 years than younger adults. We will be comparing logistic models to see which variables, between age and number of children, better predict the likelihood of adults pursuing higher education, meaning they have completed their degree or are part-time or full-time students within the years 2006-2010.

Additionally, previous studies have shown that student parents tend to be older² and are more likely to face intense economic challenges³ while enrolled in college than their peers without children, however there is no substantive research on the number of children dependent on a student parent. Regardless, through our analysis, we hope to offer more clarity in the effects that each variable has on adults pursuing higher education.

Methods

Data

This study will use data from the National Surveys of College Graduates (NSCG), Recent College Graduates (NSRCG), and Doctorate Recipients (SDR) available in the IPUMS database. The NSCG is a survey of individuals residing in the United States who hold a bachelor's degree or higher, while the SDR collects data from individuals who received a doctorate degree from a US institution. The NSRCG collects data from individuals who obtained a science or engineering degree from a US institution within the last two years. All three surveys collect data from individuals under the age of 76. To increase our sample size, we will be using data from the year 2010, and our total sample size is 108,337 adults. The data is available at this URL: <https://highered.ipums.org/highered-action/variables/group?id=demog>.

Variables

The variables in our analysis include binary response variable of an adult graduating/studying in a higher education in the last five years with primary categorical explanatory variable of number of children, and secondary numerical explanatory variable of adult's age.

Outcome variables: Binary variable indicating whether or not the adult is pursuing or has pursued (graduated) from higher education as a part-time/full-time student in the last 6 years (binary: 0 indicates false

¹Catalano, Amy J; Radin, Susan T. International Journal of Doctoral Studies; Santa Rosa. Parents Pursuing a Doctorate of Education: A Mixed Methods Examination of How Parents Manage the Roles of Student and Parent. Vol. 16, (2021): 253-272. DOI:10.28945/4741

²Institute for Women's Policy Research (IWPR). 2018. Institute for Women's Policy Research (IWPR) analysis of data from the U.S. Department of Education, National Center for Education Statistics, 2015–16 National Postsecondary Student Aid Study (NPSAS:16)

³Noll, Liz & Reichlin, Lindsey & Gault, Barbara. (2017). College Students with Children: National and Regional Profiles. 10.13140/RG.2.2.21338.75205.

and 1 indicates true). For analysis, we separated it into two binary values by whether their year of most recent degree falls under the last 6 years and/or is currently a full-time or part-time student.

Explanatory variables: number of children (categorical), age of parent (numerical) Primary explanatory variable is the number of children, which is categorical with three levels of factored labels of: “No Child,” “One Child,” and “2 or More Children.” Secondary explanatory variable is the age of the parent, which is numerical from 23 to 75.

Analysis

We performed two model comparisons between 3 logistic regression models to analyze our primary and secondary hypothesis. The response variable is the binary variable for whether the adult has pursued higher education in the last 6 years.

The **primary hypothesis** is that adults with no children or fewer children are more likely to pursue higher education in their adult years than adults with 3 or more children.

Reduced: $\text{logit}(\text{degreewithinlast6years}) = \beta_0 + \beta_1 * 1(\text{nochild}) + \beta_2 * 1(\text{onechild}) + \beta_3 * 1(2 + \text{children})$

Full: $\text{logit}(\text{degreewithinlast6years}) = \beta_0 + \beta_1 * 1(\text{nochild}) + \beta_2 * 1(\text{onechild}) + \beta_3 * 1(2 + \text{children}) + \beta_4 * \text{AGE}$

The test of our hypothesis will be: $H_0 : \beta_3 \text{ and } \beta_4 = 0$ $H_A : \beta_3 \text{ or } \beta_4 \neq 0$

Secondary hypothesis: Younger adults are more likely to have attended higher education within the last six years than older adults.

Reduced: $\text{logit}(\text{degreewithinlast6years}) = \beta_0 + \beta_1 * \text{AGE}$

Full: $\text{logit}(\text{degreewithinlast6years}) = \beta_0 + \beta_1 * 1(\text{nochild}) + \beta_2 * 1(\text{onechild}) + \beta_3 * 1(2 + \text{children}) + \beta_4 * \text{AGE}$

The test of our hypothesis will be: $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ $H_A : \beta_1 \text{ or } \beta_2 \text{ or } \beta_3 \text{ or } \beta_4 \neq 0$

Results: When comparing logistic regression models, we use the chi-squared test. We use this to see if between the two nested models are inconsistent with differences that are expected under the null hypothesis (from class slides on Comparing Logistic Regression Models).

The following are our two model comparisons for our hypotheses:

Primary hypothesis: Does the number of children that an adult has contribute to predicting the chances of an adult pursuing or graduating higher education in the last five years?

In the reduced model, `numberChild_6years_simple`, the intercept (having no children) is -0.717, the number of children (1) is -0.617 and number of children (3+) is -1.055.

In the full model, `model_6years`, the intercept (having no children) is 5.922, age is -0.166, the number of children (1) is -0.50, the number of children (3+) is -0.937.

```
library(ipumsr)
```

```
## Warning: package 'ipumsr' was built under R version 4.2.3
```

```
library(Stat2Data)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```

##  

## Attaching package: 'dplyr'  

##  

## The following objects are masked from 'package:stats':  

##  

##     filter, lag  

##  

## The following objects are masked from 'package:base':  

##  

##     intersect, setdiff, setequal, union  

library(moderndive)  

## Warning: package 'moderndive' was built under R version 4.2.3  

# NOTE: To load data, you must download both the extract's data and the DDI  

# and also set the working directory to the folder with these files (or change the path below).  

if (!require("ipumsr")) stop("Reading IPUMS data into R requires the ipumsr package. It can be installed  

ddi <- read_ipums_ddi("highered_00002.xml")  

data <- read_ipums_micro(ddi)  

## Use of data from IPUMS Higher Ed is subject to conditions including that users should cite the data as  

data <- data[, !names(data) %in% c("WEIGHT", "SAMPLE", "SURID")]  

data_6years <- data %>%
  mutate(last06 = case_when(MR03Y5 >= 2006 | ACFPT == 1 | ACFPT == 2 ~ 1,
                            MR03Y5 < 2006 | ACFPT == 98 | ACFPT == 3 ~ 0)) %>%
  mutate(chU6 = case_when(CH6IN == 0 | CH6IN == 98 ~ 0,
                         CH6IN == 1 ~ 1)) %>%
  mutate(GENDER = GENDER -1)  

data_6years$CHTOT <- factor(data_6years$CHTOT,
                             levels = c(98, 1, 3),
                             labels = c("No Child", "One Child", "2 or More Children")
                           )  

numberChild_6years_simple <- glm( last06 ~ CHTOT, data = data_6years, family = binomial)
summary(numberChild_6years_simple)  

##  

## Call:  

## glm(formula = last06 ~ CHTOT, family = binomial, data = data_6years)  

##  

## Deviance Residuals:  

##      Min        1Q    Median        3Q       Max  

## -0.8918 -0.8918 -0.6839   1.4930   1.9640  

##

```

```

## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -0.716914  0.008534 -84.01  <2e-16 ***
## CHTOTOne Child      -0.616752  0.019989 -30.85  <2e-16 ***
## CHTOT2 or More Children -1.054675  0.019118 -55.17  <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: 124443  on 108336  degrees of freedom
## Residual deviance: 120627  on 108334  degrees of freedom
## AIC: 120633
##
## Number of Fisher Scoring iterations: 4

model_6years <- glm( last06 ~ AGE + CHTOT, data = data_6years, family = binomial)
summary(model_6years)

```

```

##
## Call:
## glm(formula = last06 ~ AGE + CHTOT, family = binomial, data = data_6years)
##
## Deviance Residuals:
##      Min        1Q     Median        3Q       Max
## -2.1083  -0.4939  -0.2039   0.4789   3.8610
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)           5.922421  0.041927 141.26  <2e-16 ***
## AGE                  -0.165849  0.001135 -146.09  <2e-16 ***
## CHTOTOne Child      -0.501470  0.025220 -19.88  <2e-16 ***
## CHTOT2 or More Children -0.936770  0.022766 -41.15  <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: 124443  on 108336  degrees of freedom
## Residual deviance: 71926  on 108333  degrees of freedom
## AIC: 71934
##
## Number of Fisher Scoring iterations: 6

```

```
anova(numberChild_6years_simple, model_6years, test="Chisq")
```

```

## Analysis of Deviance Table
##
## Model 1: last06 ~ CHTOT
## Model 2: last06 ~ AGE + CHTOT
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      108334    120627
## 2      108333    71926  1      48701 < 2.2e-16 ***

```

```

## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Using ANOVA to perform a G-test, ($p < 0.05$) meaning we reject the null hypothesis and prefer the full model. We conclude that the number of children does contribute to predicting the chances of an adult pursuing or graduating higher education in the last five years.

Secondary hypothesis: Does the age of an adult contribute to their chances of pursuing or graduating higher education in the last five years?

In the reduced model, age_6years_simple, the intercept is 6.031 and the age is -0.179.

In the full model, model_6years, the intercept (having no children) is 5.922, age is -0.166, the number of children (1) is -0.50, the number of children (3+) is -0.937.

```

age_6years_simple <- glm( last06 ~ AGE, data = data_6years, family = binomial)
summary(age_6years_simple)

```

```

##
## Call:
## glm(formula = last06 ~ AGE, family = binomial, data = data_6years)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0280  -0.5007  -0.1761   0.5232   3.8411
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 6.031035  0.043254 139.4   <2e-16 ***
## AGE        -0.178765  0.001179 -151.7   <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: 124443 on 108336 degrees of freedom
## Residual deviance: 73725 on 108335 degrees of freedom
## AIC: 73729
##
## Number of Fisher Scoring iterations: 6

```

```

model_6years <- glm( last06 ~ AGE + CHTOT, data = data_6years, family = binomial)
summary(model_6years)

```

```

##
## Call:
## glm(formula = last06 ~ AGE + CHTOT, family = binomial, data = data_6years)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1083  -0.4939  -0.2039   0.4789   3.8610
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)

```

```

## (Intercept)      5.922421  0.041927 141.26 <2e-16 ***
## AGE             -0.165849  0.001135 -146.09 <2e-16 ***
## CHTOTOne Child  -0.501470  0.025220 -19.88 <2e-16 ***
## CHTOT2 or More Children -0.936770  0.022766 -41.15 <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: 124443  on 108336  degrees of freedom
## Residual deviance: 71926  on 108333  degrees of freedom
## AIC: 71934
##
## Number of Fisher Scoring iterations: 6

anova(age_6years_simple, model_6years, test="Chisq")

```

```

## Analysis of Deviance Table
##
## Model 1: last06 ~ AGE
## Model 2: last06 ~ AGE + CHTOT
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1     108335    73725
## 2     108333    71926  2    1798.9 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Using ANOVA to perform a G-test, ($p < 0.05$) meaning we reject the null hypothesis and prefer the full model. We conclude that age does contribute to predicting the chances of an adult pursuing or graduating higher education in the last five years for adults, and adults who have children.

Coefficient Interpretation:

For reduced model *last06 AGE*:

The intercept represents that when age is 0, the log-odds of the adult pursuing or has pursued (graduated) from higher education as a part-time/full-time student in the last 6 years is 6.03.

The AGE coefficient represents that the log-odds of the adult pursuing or has pursued (graduated) from higher education as a part-time/full-time student in the last 6 years decreases by 0.18 when the age variable increases by 1.

For reduced model *last06 CHTOT*:

The intercept represents that when CHTOT is 0, the log-odds of an adult pursuing or has pursued (graduated) from higher education as a part-time/full-time student in the last 6 years is -0.72.

The CHTOTOne Child represents that compared to the no child level, the log-odds of the adult pursuing or has pursued (graduated) from higher education as a part-time/full-time student in the last 6 years with one child decreases by 0.62 compared to no children.

The CHTOT2 or More Children represents that compared to no child level, the log of the odds of the adult pursuing or has pursued (graduated) from higher education as a part-time/full-time student in the last 6 years with two or more children decreases by 1.05 compared to no children.

For full model *last06 AGE + CHTOT*:

The intercept represents that when a person is 0 years old with no children, the log-odds of the person pursuing or has pursued (graduated) from higher education as a part-time/full-time student in the last 6 years is 5.92.

The AGE coefficient represents that for each additional age, the log-odds of the adult pursuing or has pursued (graduated) from higher education as a part-time/full-time student in the last 6 years decreases by 0.17.

The CHTOTOne Child represents that compared to the no child level and with age 0, the log-odds of the person pursuing or has pursued (graduated) from higher education as a part-time/full-time student in the last 6 years with one child is 0.5 lower compared to no children.

The CHTOTOne Child represents that compared to the no child level and with age 0, the log-odds of the person pursuing or has pursued (graduated) from higher education as a part-time/full-time student in the last 6 years with one child is 0.94 lower compared to no children.

Exploratory Visualization

```
probabilities <- model_6years %>%
  predict(data_6years, type = "response")
predicted.classes <- ifelse(probabilities > 0.5, 0, 1)
mean(predicted.classes == data_6years$last06)

## [1] 0.1335739

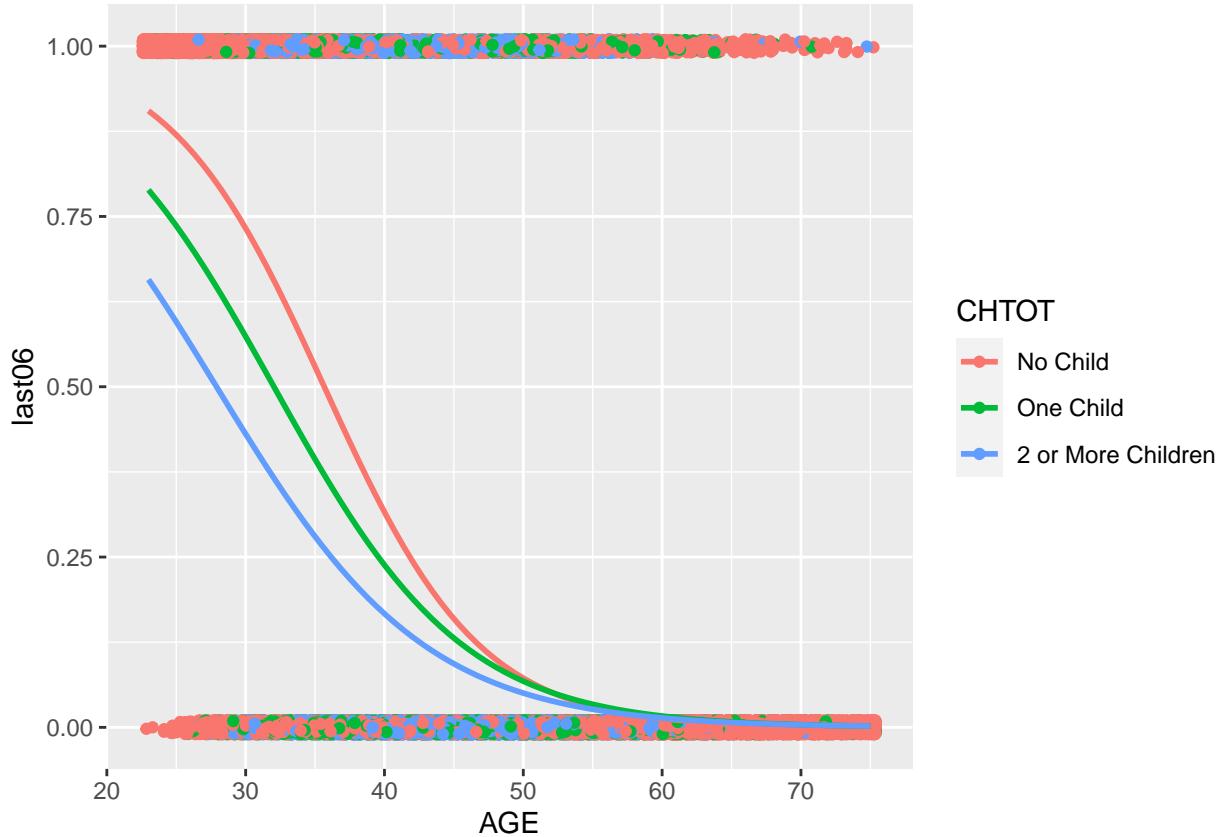
new_data1 <- exp(cbind(OR = coef(model_6years), confint(model_6years)))

## Waiting for profiling to be done...

head(new_data1)

##                                     OR      2.5 %     97.5 %
## (Intercept) 373.3143376 343.9672806 405.4106085
## AGE          0.8471740  0.8452821  0.8490522
## CHTOTOne Child 0.6056400  0.5764058  0.6363020
## CHTOT2 or More Children 0.3918915  0.3747675  0.4097512

data_6years %>%
  ggplot(aes(x=AGE, y=last06, color= CHTOT)) +
  geom_point(
    position=position_jitter(height=.01)
  ) +
  geom_smooth(
    method=glm,
    method.args=list(family=binomial),
    se=FALSE,
    formula = y~x
  )
```



Discussion

In the study, we want to investigate how the number of children or the age of the adult would influence the likelihood of adults pursuing higher education in the United States. After analysis, we conclude that both factors (age, number of children) contribute significantly to predicting the chances of adults obtaining higher education in the last five years. To streamline the analysis process, we limited our data to the last five years, and the data we used only includes adults who are younger than 76, meaning we cannot generalize our results to the whole US adult population. Although we conclude that both of these factors have significant contributions to our model, we cannot measure the exact strength of the two variables when doing model comparisons between logistic regression models.

To improve our study, we could include more data. For instance, instead of only investigating data in the last five years, we can expand our data set to last ten-year data. Also, since the number of children CHTOT variable is categorical with three levels (no children, 1 child, or 3+ children), this variable limited the model we could use. In our study, we used a logistic regression model, but if the CHTOT variable was numerical with a range of ages, we could possibly fit linear regression models to see the relationship between age and the number of children, which could provide more insight into their significance. Additionally, since our data was mainly categorical, it would have been interesting to see if there were significant contributions between other variables such as minority status, and income/salary.

In summary, though there are some limitations, the study provides important insight into the relationship between the number of children, the age of the adult, and the likelihood of adults pursuing higher education.

References

Catalano, Amy J; Radin, Susan T. International Journal of Doctoral Studies; Santa Rosa. Parents Pursuing a Doctorate of Education: A Mixed Methods Examination of How Parents Manage the Roles of Student and Parent. Vol. 16, (2021): 253-272. DOI:10.28945/4741.

Institute for Women's Policy Research (IWPR). 2018. Institute for Women's Policy Research (IWPR) analysis of data from the U.S. Department of Education, National Center for Education Statistics, 2015–16 National Postsecondary Student Aid Study (NPSAS:16)

Noll, Liz & Reichlin, Lindsey & Gault, Barbara. (2017). College Students with Children: National and Regional Profiles. 10.13140/RG.2.2.21338.75205.