Group Project - Draft Proposal

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

2023-03-13

Contents

Title: The Influence of Parenthood and Age on Adults Pursuing Higher Education **1**

Purpose and Hypotheses **1**

Data **2**

Outcome Variable **2**

Explanatory Variables **2**

Primary Hypothesis 2

Secondary Hypothesis 2

Exploratory Visualization **3**

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

Purpose and Hypotheses

The aim of this project is to investigate the impact of parenthood on the likelihood of adults obtaining higher education, with a specific focus on parents with children and the number of children they have. 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 counterparts. They tended to finance their degrees through loans and personal savings rather than scholarships" (Kulp, p.255)." Given these findings, we are interested in exploring this relationship.

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.

The **secondary hypothesis**:

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

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. The data is available at this URL: https://highered.ipums.org/highered-action/variables/group?id=demog.

Outcome Variable

The outcome variable determines whether or not the adult is pursuing or has pursued higher education in the last 6 years. It is binary; 0 indicates false and 1 indicates true.

We will use logistic regression to model this outcome.

Explanatory Variables

For the primary explanatory variable, we will use the variable CHTOT; the number of children, which is categorical with three levels: 1 - one child, 3 - two or more children, 98 - no child

For the secondary explanatory variables,

-AGE; the age of the parent, which is numerical from 23 to 75.

-ACFPT; current enrollment student status (1 - part-time student, 2- full-time student, 3- not enrolled in a degree program but taking courses, 98 - logical skip)

The explanatory variables shown in our regression models:

Primary Hypothesis

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.

The test of our hypothesis will be:

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

$H_A : \beta_1 \neq \beta_2 \neq \beta_3 \neq \beta_4 \neq 0$

$logit(degreewithinlast6years) = \beta_0 + \beta_1 CHTOT + \beta_2 * 1(nochild) + \beta_3 * 1(onechild) + \beta_4 * 1(2 + children)$

Secondary Hypothesis

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

The test of our hypothesis will be:

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

$H_A : \beta_1 \neq \beta_2 \neq \beta_3 \neq \beta_4 \neq 0$

$logit(degreewithinlast6years) = \beta_0 + \beta_1 CHTOT + \beta_2 * 1(nochild) * AGE + \beta_3 * 1(onechild) * AGE + \beta_4 * 1(2 + children) * AGE$

Exploratory Visualization

```
library(ipumsr)
library(Stat2Data)
library(ggplot2)
library(dplyr)
```

```
##
## 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)
```

```
# 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 installe

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 appropriately. Use command `ipums_conditions()` for more
## details.
```

```
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")
                        )
```

```
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
```

```r
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
```

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

```
## Waiting for profiling to be done...
```

```r
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
```

```r
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
    )
```

```
## Don't know how to automatically pick scale for object of type
## <haven_labelled/vctrs_vctr/integer>. Defaulting to continuous.
```