1. Introduction
   1. Abstract
   2. Background
2. Hypothesis and referencing Deming’s paper on headstart
   1. What are the **short and long term effects** of **head start program on youth** based on a **child's ethnicity**.
3. Methodology
   1. Data processing
      1. Defining what headstart is represented in the dataset
         1. Most likely determining whether or not an individual participated in any head start program, rather than a set of years. Need to generate a new variable.
      2. Defining what is “short and long term effects” based on dataset
         1. Education Enrollment
            1. High school
            2. College
         2. Family income
            1. We can do a before/after. Taking an individual, is their income higher than their peers holding all other factors fixed?
      3. Defining the “ethnicities” that we are classifying
         1. Hispanic
         2. black
         3. white
      4. Head Start
         1. If ever participated in head start, would mean yes
      5. Other preschool
      6. No preschool
      7. Tracking individual people through the years
         1. Perhaps have a subset of people who graduate high school early
         2. Subset of people who don’t graduate high school
      8. Crimerate: would HS effect on crime?
      9. Variable List
         1. head\_start: 0 or 1
         2. sibdiff : 0 or 1
         3. mom\_id: unique from the dataset
         4. Hispanic: 0 or 1
            1. Race\_Child
            2. Created new variable

Hispanic: 0 for not, 1 for yes

* + - 1. Black: 0 or 1
         1. Race\_Child
      2. Male: 0 or 1
         1. Sex\_Child
      3. Firstborn: 0 to 1
         1. BirthOrder
      4. lninc\_0to3: log income, 0 to 3,
      5. momed: years of mom edu
      6. dadhome\_0to3: research
      7. ppvt\_3: the vocabulary test scores
      8. lnbw: IDK
      9. comp\_score\_5to6: Preschool
      10. comp\_score\_9to10: Elementary School
      11. comp\_score\_13to14: Middle School
      12. comps\_score17to18: High School
      13. repeat: if you needed to repeat a grade
          1. created variable

repeatgrade

* + - 1. learndis: 0 or 1
         1. Created variable

Learndisability

* + - 1. hsgrad: 0 or 1
         1. Created variable

hsgrad

* + - 1. somecoll: 0 or 1
         1. Created variable

somecollege

* + - 1. idle:
      2. fphealth: self-reported health, 0-1
      3. Age of mother when child was born (Teen pregnancy): numeric
         1. Original Dataset:

Age\_Moth\_Birth

* + - 1. Age of child when attended HS (pre school age): numeric
         1. Original Dataset:

Age\_1stHS88

Age\_1stHS90

Age\_1stHS92

Age\_1stHS94

Age\_1stHS96

Age\_1stHS98

Age\_1stHS100

Age\_1stHS102

Age\_1stHS104

* + - 1. If child has mental disability (to control for): 0 or 1
         1. Created variable

“mental.disability”

* + - 1. If child has limiting conditions (in case it might result in repetition): 0 or 1
      2. Birthweight: numeric
         1. Original dataset:

BirthWeight

* 1. Data

1. Results
   1. Figure 1: recreate linear model from the paper, present table.
   2. Figure 2: present the best model to fit our research questions.
   3. Figure 3: we found that figure one (pros and cons)
   4. Figure 4: why our model would best answer our hypothesis.
2. Conclusion
3. Appendix