

## Factors that affect early childhood education

Win Suvansinpan, Geoffrey Li, Roshan Ramkeesoon

Wednesday, March 13, 2019

## Outline

- ▶ Dataset
- ▶ Experiment 1. TV watching
- ▶ Experiment 2. Bedtime
- ▶ Experiment 3. Age

## Dataset

- ▶ Early Childhood Longitudinal Study
- ▶ Sampling and data collection methodology
- ▶ Availability
- ▶ Common response variable

## Early Childhood Longitudinal Study

- ▶ conducted by the Institute for Education Sciences, part of U.S. federal government<sup>1</sup>
- ▶ collected data on factors that affect educational outcomes by interviewing students, their teachers, and their parents/guardians
- ▶ tracking 3,341 variables ranging from student bedtime to the teacher's level of interest in teaching, the study is one of the largest datasets of its kind
- ▶ We only use the 2002 dataset, which pertains to 15,000 students in spring 3rd grade from the Kindergarten class of 1998-1999

---

<sup>1</sup>U.S. Dept of Education, National Center for Education Statistics. EARLY CHILDHOOD LONGITUDINAL STUDY [UNITED STATES]: KINDERGARTEN CLASS OF 1998-1999, THIRD GRADE [Computer file]. ICPSR version. Washington, DC: U.S. Dept of Education, Institute of Education Sciences [producer], 2004. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2004.

## Sampling and data collection methodology

- ▶ 15,000 students nationwide
- ▶ multi-stage stratified cluster sampling
  - ▶ schools were randomly selected across U.S.A.
  - ▶ average of 23 kindergartners randomly selected from each selected school
- ▶ sample is diverse (socioeconomic status, race, private/public schools, etc)
- ▶ parents and teachers were interviewed over phone, students were interviewed in-person
- ▶ all interviews conducted by trained staff

## Availability

- ▶ ECLS dataset is available for download on the ECLS website
- ▶ High level (>90%) of completeness for most fields

## Common response variables

- ▶ math, reading, and science exams administered during student surveys
- ▶ “math t-score”: raw scores in each subject were also provided as normalized to  $N(\mu = 50, sd = 10)$
- ▶ normalization is useful so we can make comparisons in response across the three different subjects
- ▶ will examine average changes in child’s normalized score associated with predictors

## Experiment 1. TV watching

- ▶ Background
- ▶ Hypothesis
- ▶ Variables
- ▶ Method
- ▶ Results

## Background

- ▶ While television is a decades old technology, the interaction between time spent consuming media and child education outcomes is still of interest
- ▶ prior studies have shown that there is a statistically significant negative effect of early childhood television watching on educational outcomes at high school level
- ▶ prior studies indicate that time spent after classes on student activities and homework have significant positive influence on grades
- ▶ interested in matching prior study's association between total television watching with grades in our data
- ▶ interested in association of afterschool television watching on grades

## Hypotheses

- ▶  $H_{A0}$ : there is no association between hours of television watching **total per week** on normalized student grades in any subject
- ▶  $H_{B0}$ : there is no association between hours of television watching **after school** on normalized student grades in any subject

## Variables

- ▶ Control for parent income a numeric proxy of socioeconomic status which we think may have an effect on parent and child behavior
- ▶ Predictors are hours of tv watched in each period:
  - ▶ weekdays before 8am
  - ▶ weekdays between 8am and 3pm
  - ▶ weekdays between 3pm and dinner
  - ▶ weekdays between dinner and bedtime
  - ▶ saturdays
  - ▶ sundays

## Method

- ▶ To test  $H_{A0}$ : Use linear regression to test for significance of hours of tv watched **total per week** on each subject score while controlling for parent income
- ▶ To test  $H_{B0}$ : Use linear regression to test for significance of hours of tv watched **after school** on each subject score while controlling for parent income

## Television watched weekly

```
[1] "MATH"
```

	Estimate	Std. Error	t value	Pr(> t )
tv_watched	-5.357392e-02	1.932118e-02	-2.772808	5.568096e-03
income	7.416452e-05	2.055512e-06	36.080806	1.285431e-267

```
[1] "READING"
```

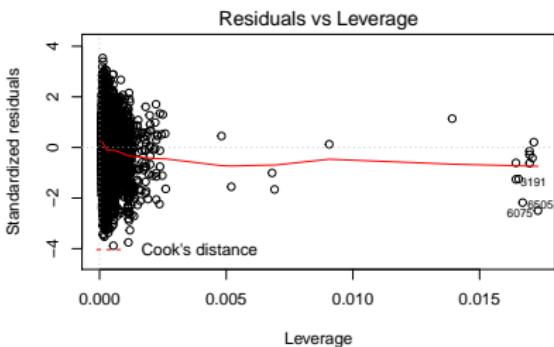
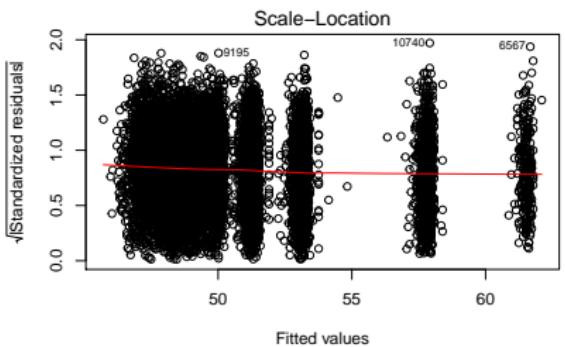
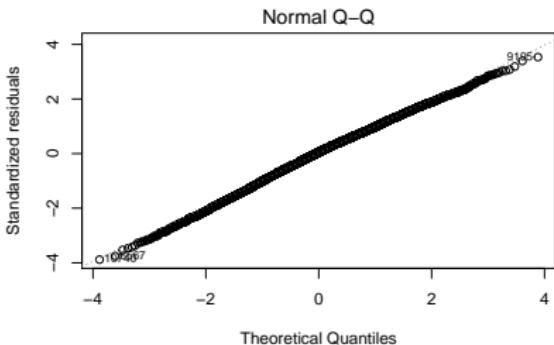
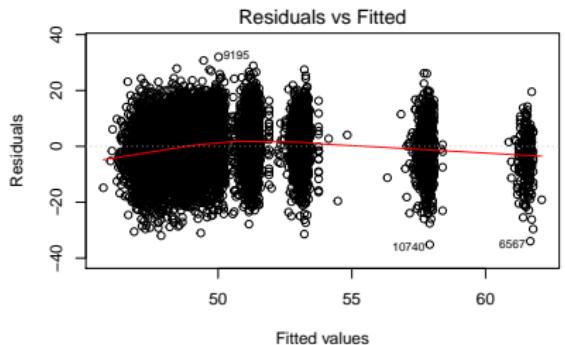
	Estimate	Std. Error	t value	Pr(> t )
tv_watched	-6.596743e-02	2.165083e-02	-3.046877	2.318537e-03
income	8.196384e-05	2.303356e-06	35.584538	8.600593e-261

```
[1] "SCIENCE"
```

	Estimate	Std. Error	t value	Pr(> t )
tv_watched	-5.813635e-02	1.977590e-02	-2.939757	3.292442e-03
income	7.951489e-05	2.103888e-06	37.794259	9.888020e-292

- ▶ All tv watching coefficients statistically significant, have negative association with test scores
- ▶ Income coefficient is statistically significant, positive association with grades
- ▶ Each additional hour of tv watched per week is associated with an average 0.05 to 0.07 drop in mean test scores for a given income
- ▶ We have reproduced the result that we reject the null hypothesis that there is no association between hours of television watched and grades

## Assess model assumptions



## Assess model assumptions

- ▶ slight deviance from normality in the residuals vs fitted plot
- ▶ qqplot shows normality of residuals
- ▶ scale-location plot shows constant variance of residuals across fitted values
- ▶ striation patterns in the residuals versus fitted plot might be explained by large jumps in data for parent's income
- ▶ similar results for comparisons with the other two tests (see report)
- ▶ seems like assumptions hold sufficiently for parameters of interest to be meaningful

## Television watched after school

```
[1] "MATH"
```

	Estimate	Std. Error	t value	Pr(> t )
tv_watched	-2.341768e-01	9.588090e-02	-2.442372	1.460883e-02
income	7.426279e-05	2.055343e-06	36.131572	2.551378e-268

```
[1] "READING"
```

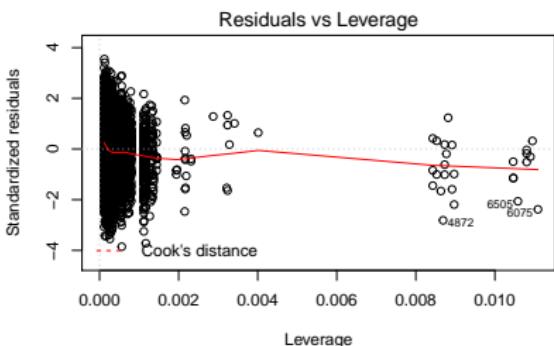
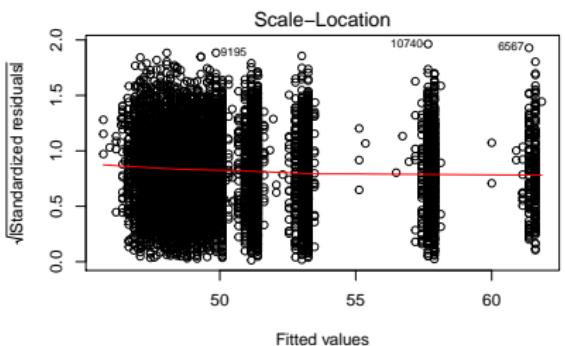
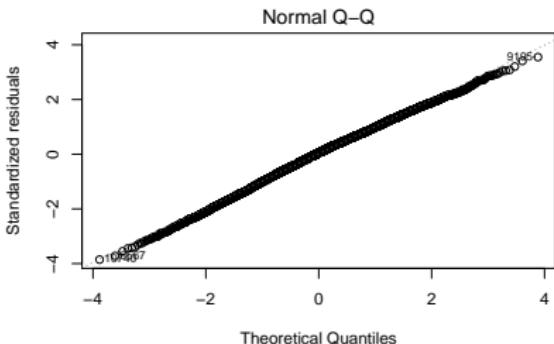
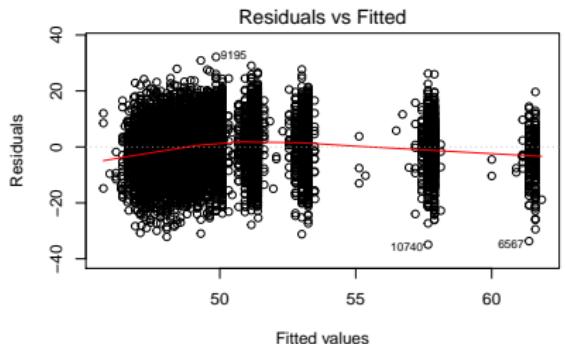
	Estimate	Std. Error	t value	Pr(> t )
tv_watched	-0.376264734	1.074157e-01	-3.502884	4.623227e-04
income	0.000081832	2.302608e-06	35.538829	3.626932e-260

```
[1] "SCIENCE"
```

	Estimate	Std. Error	t value	Pr(> t )
tv_watched	-2.754657e-01	9.813259e-02	-2.807077	5.009303e-03
income	7.956013e-05	2.103612e-06	37.820734	4.130156e-292

- ▶ All the coefficients are statistically significant and negative
- ▶ Income is significantly significant and positive association with grades
- ▶ Each additional hour of tv watched after school is associated with an average 0.23 to 0.37 drop in mean test scores for a given income
- ▶ We reject the null hypothesis that after school television watching does not have an association with grades

## Assess model assumptions - math scores



## Assess model assumptions

- ▶ slight deviance from normality in the residuals vs fitted plot
- ▶ qqplot shows normality of residuals
- ▶ scale-location plot shows constant variance of residuals across fitted values
- ▶ striation patterns in the residuals versus fitted plot might be explained by large jumps in data for parent's income
- ▶ similar results to reading and science (see report)
- ▶ seems like assumptions hold sufficiently for parameters of interest to be meaningful

## Compare coefficients from both experiments

- ▶ Coefficients found from the two linear regression models mentioned on previous slides:

Subject	Weekly.Hours	AfterSchool.Hours
1 Math	-0.0536	-0.23
2 Reading	-0.0660	-0.38
3 Science	-0.0581	-0.28

- ▶ The coefficients are higher than for total television watching indicating the isolated effect of each "after school" tv watching hour associated with worse outcomes than each hour weekly tv watching