Number of Hours per Week and Age in Early Education Used to Predict Mathematics Success at Grade 5

(student name withheld)

**INTRODUCTION**

Statistical methods are widely used in pre-K education research to determine school readiness and grade outcome, especially in urban, rural and low-poverty stricken areas where children are at high risk. National policymakers and the public as a whole also realize the success of the United States depends on the successful functioning of the American education system. There is a growing awareness that school reform must include elementary and preschool and not just reform for secondary and postsecondary education (U.S. Department of Education, 2004).

“Efforts to expand and improve early education will benefit from insights gained through analyses of data from the large-scale, nationally representative, longitudinal Early Childhood Longitudinal Study-Kindergarten Class (ECLS-K) database. The ECLS-K database contains information about the types of preschool and elementary programs in which children participate, the services they receive, and repeated measures of the children’s cognitive skills and knowledge. The ECLS-K database also contains measures of children’s physical health and growth, social development, and emotional well-being, along with information on family background and the educational quality of their home environments.”- (U.S. Department of Education, 2004).

Using R and the raw data from the ICPSR: DS1: Early Head Start Research and Evaluation (EHSRE) Study (EHSRE, 2010), we will construct a CART model to predict the number of hours in EHS and age that has the most impact on mathematics achievement on the ECLS-K at Grade 5.

**RESEARCH QUESTION**

Can we use the number of hours per week of early education at ages 14, 24, and 36 months from the EHSRE study to predict math achievement on the ECLS-K math routing exam at Grade 5? An ECLS-K math score >= 10 is the national cutoff for math success.

The objective is to use predictor Early Head Start (EHS) variables hours per week at ages 14, 24, and 36 to predict outcome of Grade 5 mathematics success.

**METHOD**

Our research question above was derived from the original EHSRE longitudinal study. We constructed our data model using R and a subset of data from EHSRE to build our CART model.

**DATA SOURCES**

The EHSRE data was obtained from a link to the Inter-University Consortium for Political and Social Research (ICPSR) from the data.gov website. In the ICPSR - EHSRE study, baseline data were collected when participants first applied to an EHS program and additional data were collected during follow-up interviews and assessments based on the numbers of months since random assignment and the age of focus child. There were three phases of the study: (1) Birth to Three [“0-3”], (2) Pre-Kindergarten Follow-up [“PreK”] and (3) Elementary School Follow-up [“G5”]. There were about 3,000 applicants randomly selected for the impact study from 17 EHS programs. The amount of history the dataset covered ranged from 1996 to 2010. Academic Success on Mathematics assessment was defined as scoring an 8 or better on the Mathematics Routing Scale. (Peterson, Zhang, Roggman, Green, Cohen, Atwater, McKleavy & Korfmacher, 2006).

**Data Cleaning**

We had to clean up the data a bit by creating a subset that extracted negative values from the EHSRE dataset that were assigned to administrative codes in the range -1 through -8. These values were used to show which questions on the survey were either not answered, not on the survey version, invalid, etc. Our subset data is included in our Excel file.

**MEASURES**

There were over 1,000 variables in the EHSRE dataset; however, we only used a subset of these to develop our own constructs. We had to make some assumptions about which factors influenced academic success and so we omitted variables from the EHSRE dataset that were not appropriate to our analysis.  Our assumptions are as follows:

1. The number of Hours per week in EHS is a factor in future academic success
2. Participation in any Pre-K program is a factor in future academic success
3. Participation in EHS at age 3 is a factor in future academic success
4. Participation in EHS at age 4 is a factor in future academic success
5. Participation in EHS all three programs, Pre-K, age 3, age 4 is a factor in future academic success

The following are variables we used in our data model:

Predictors:

AGEHRSE4 - PSIs: HRS/WK IN EHS CARE AT 14 MO; Nominal.

AGEHRSE6 - PSIs: HRS/WK IN EHS CARE AT 24 MO; Nominal.

AGEHRSE8 - PSIs: HRS/WK IN EHS CARE AT 36MO; Nominal.

Target:

C5MATHRA: ECLS-K math routing score; Nominal. Dependent variable.

**LIMIATIONS**

One of the immediate limitations of the study was that the data was from a longitudinal study. Participants may drop out and reduce the sample size and ability to reach representativness. Also, a lot of the data in our raw data set could not be used because of lack of responsiveness on the survey or invalid or missing information. Basic statistics do not help us predict the path of events that will lead to successes. We used R and Cart modeling to create a decision tree to visualize observations in nodes and branches. However, there is also limitations with our technology, R. Therefore, we included an evaluation of risk in the accuracy of our model. It is important to review limitations of the model when making a decision based on model results. We find that the results of our model outweigh its limitations and so we feel comfortable with our decision.

**ANALYTIC APPROACH**

Our DA model is constructed to predict academic success in math for 5th Graders using predictor Early Head Start (EHS) variables (AGEHRSE) and outcome of Grade 5 math success (C5MATHRA). The dependent variable in our subset will be C5MATHRA- ECLS-K math routing score. A score of 10 or greater on C5MATRA- ECLS-K math is the cutoff for success on the math assessment. Our independent variables will be the number of hours per week in EHS or Pre-K (AGEHRSE). Our model will allow us to visualize and analyze the dataset to determine if participation in EHS can be associated with better academic outcomes on both math if a 5th grader was able to perform well on assessment tests in these areas given their early education path.

Ideally the analysis is to minimize error, bias and confounding inherent in all research studies, but sometimes things come up that you did not anticipate during the study. Therefore we control for it in our study using a comparison group, covariates, or historical records if raw data is unavailable. In this study, we can use the data from the comparison group of children who did not enroll in an EHS program to compare math scores at Grade 5 to determine if the effect on the number of hours per week in EHS has an impact. We can also look at possible covariates in our dataset such as income and child health (number of doctor / hospital visits).

**CODE**

> d4<- read.csv("~/coursework/snhu/prek.csv")

> require(tree)

> require(rpart)

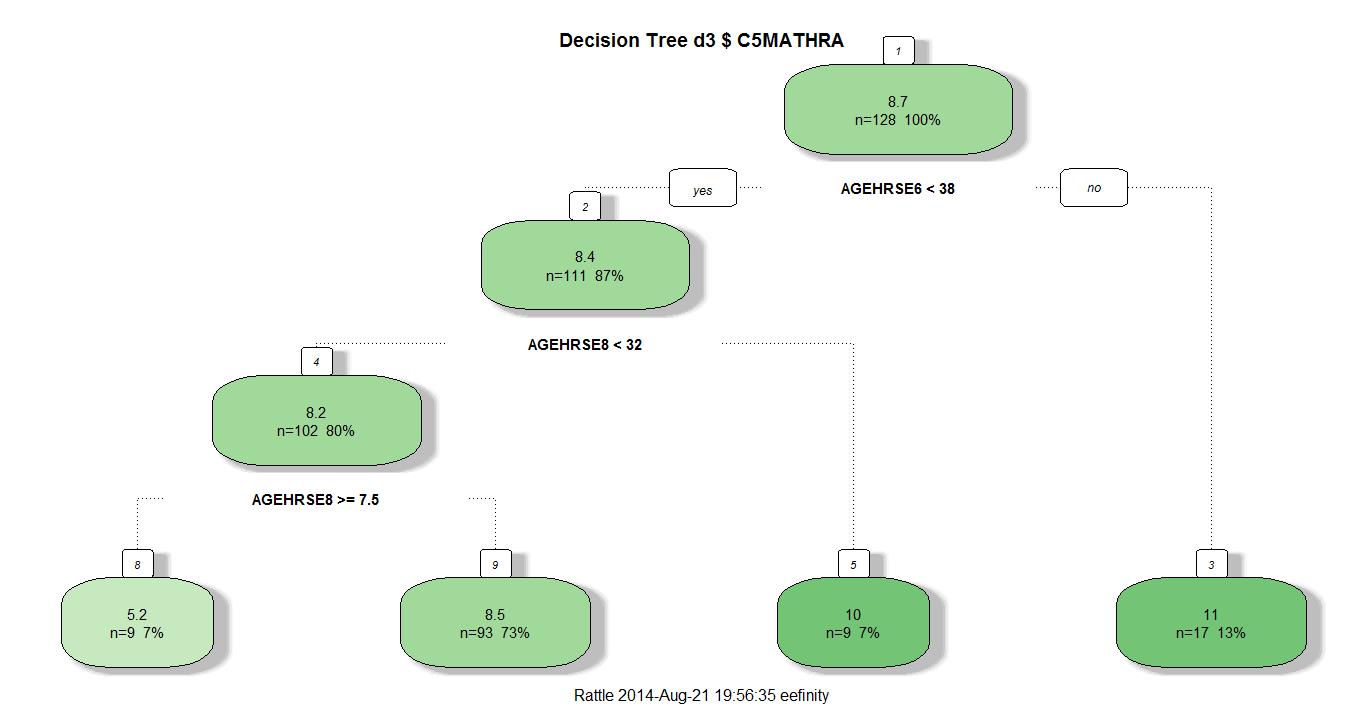
> require(rattle)

> rattle()

**RESULTS**

**CART MODEL**

To build our model for decision support, we used the function Rattle from Rpart in R. Here is our model plot below.



**Table 1. Cart Model**

Validity of Tree Classification:

 I ran the Error Matrix utility in Rattle to evaluate the validity of my prediction. These are the results.

Variables actually used in tree construction:

[1] AGEHRSE6 AGEHRSE8

Root node error: 2703/128 = 21.117

n=128 (43780 observations deleted due to missingness)

CP nsplit rel error xerror xstd

1 0.037228 0 1.00000 1.0121 0.088799

2 0.022569 1 0.96277 1.0739 0.095296

3 0.010000 3 0.91763 1.0474 0.093568

Time taken: 0.09 secs

So what does our tree in R tell us? Thirteen percent (13%) of children in the study who participated in EHS for at least 38 hrs/week at 24 mos (2yrs of age) had scores of at least 11 on the math assessment. Those who didn’t had a score of 8. This group not only had a higher math score than those receiving 32 hrs/week or less, but also scored higher than the cutoff for math success (cutoff = 10). Of the children who scored less than 11 on the math assessment and participated more than 32 hrs/wk in EHS program at age 36 months (3 yrs of age) scored a 10 on the math assessment. This group made the math success cut score of 10 and scored higher than children who received less than 32 hr/wk in early education by about 2 points. Focus children who receive >= 7.5 hrs/wk in EHS at age 36 months scored 8.5 points which was about 3 points higher than children who did not. Based on the data model, I would predict that having young children as early as 24 and 36 months (or 2 yrs and 3 yrs of age) can have a positive impact on their math assessment outcomes. We did not see any outcomes in the model as a result of EHS at 14 months. If you have any children it may be beneficial to get them enrolled in headstart between age 2 and 4.

**CONCLUSION**

In our data model we used R to construct a CART model. The object was to predict math success at Grade 5 given the number of hours per week in EHS at ages 14, 24, and 36 months. We have determined based on our model, that math scores rise at ages 24 and 36 months when the number of hours in EHS increases. Therefore we can use our model to predict future ECLS-K math outcomes at Grade 5 given the number of hours children are in EHS. The future implications of this model can lead to further predicting math achievement outcome in general from grade 5 and beyond.

**REFERENCES**

Peterson, C, Zhang, D; Roggman, L; Green, B; Chazan-Cohen, R; Atwater, J; McKelvey, L & Korfmacher, J (2006). Family participation and involvement in Early Head Start home visiting services: Relations with longitudinal outcomes. Springer US. Retrieved from website: <http://npcresearch.com/Files/Involvement_in_early_head_start_0106.pdf>

ICPSR (1996-2010). ICPSR 03804: Early Head Start Research and Evaluation (EHSRE) Study, 1996-2010: [United States]. Inter-university Consortium for Political and Social Research (ICPSR). Retrieved from website: <https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/3804?q=early+childhood+education&searchSource=find-analyze-home>

US Department of Education (2006). Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K) Base Year Public-Use. National Center for Education Statistics. Retrieved from website: <http://nces.ed.gov/pubs2001/data/2001029e_imputation.zip>