

Predicting Reading Scores

Data: NELS 88 (National Education Longitudinal Study 1988)

Shreya Goel

shreya.goel@Columbia.edu

Data Utilized

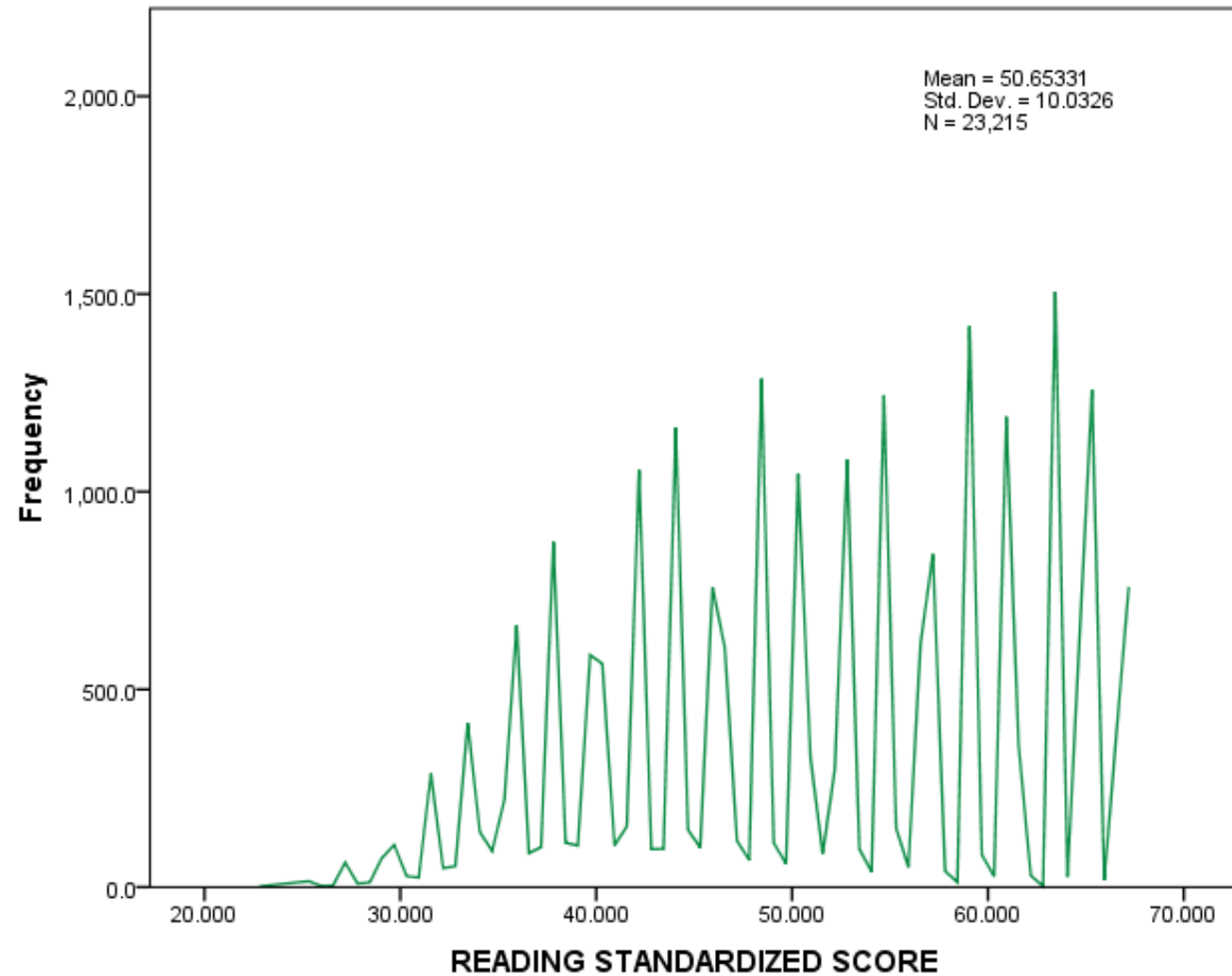
- ▶ Home Demographics
- ▶ Parents' Perception and Engagement
- ▶ Students' Perception and Level of Commitment and Engagement over time toward the subject and otherwise
- ▶ Reading Standardized Score

Current Research Focus

Built Regression Model to Predict Reading Scores based on Home Demographics, Parents and Students Actions and Perceptions

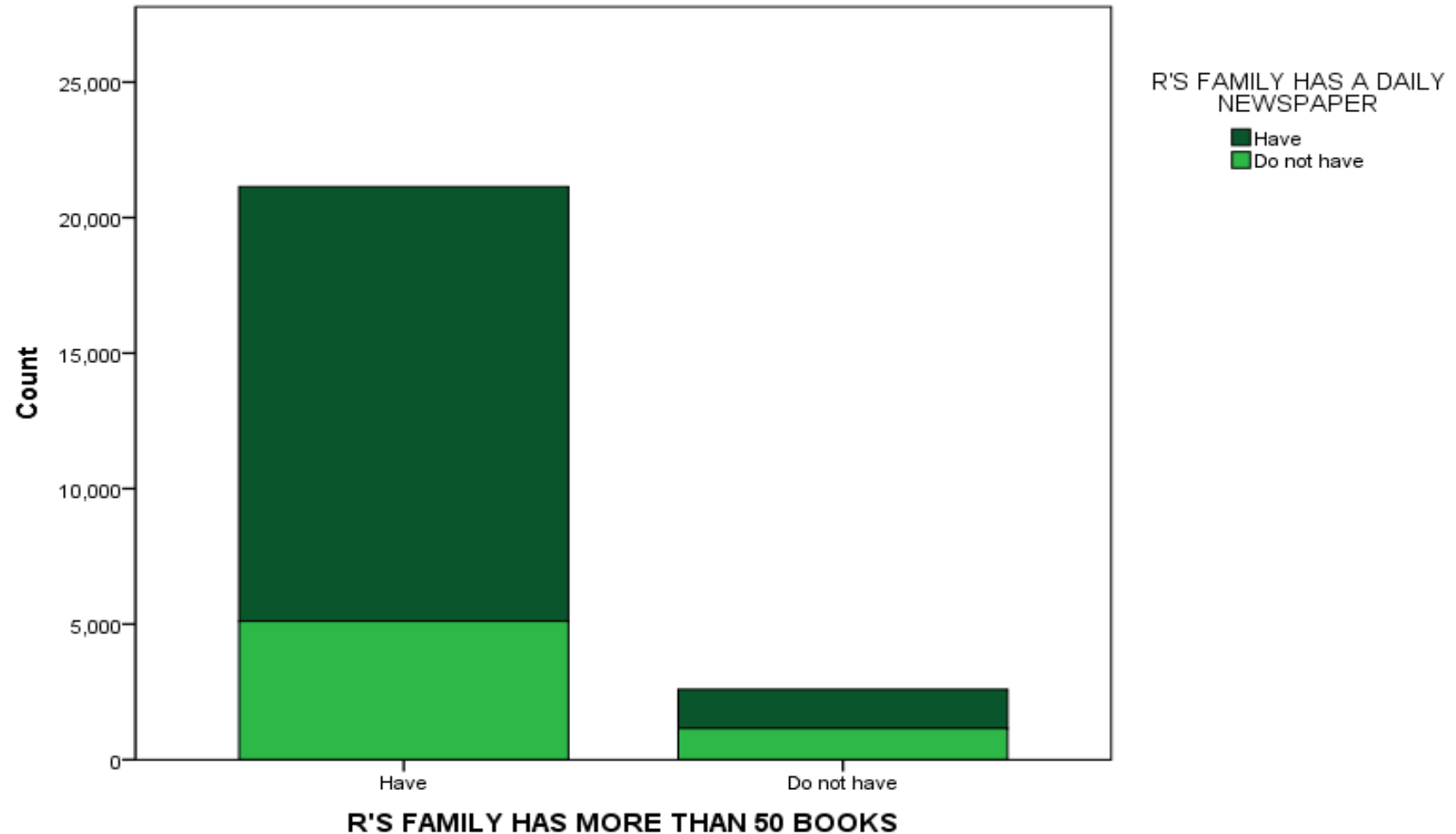
Initial Analysis

Reading Standardized Score



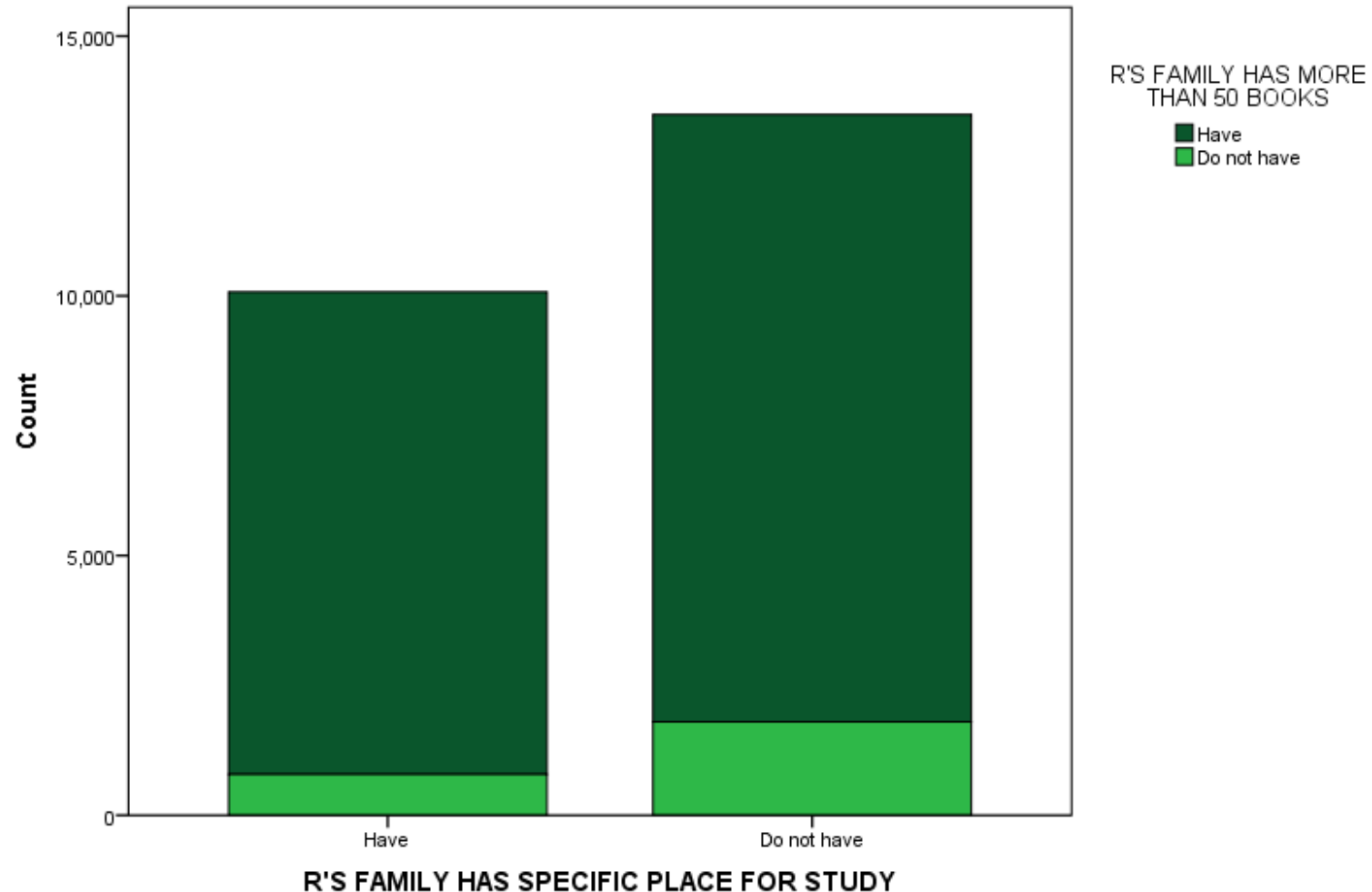
Number if Books

Most families have more 50 books and most of them also have subscribed for a daily newspaper

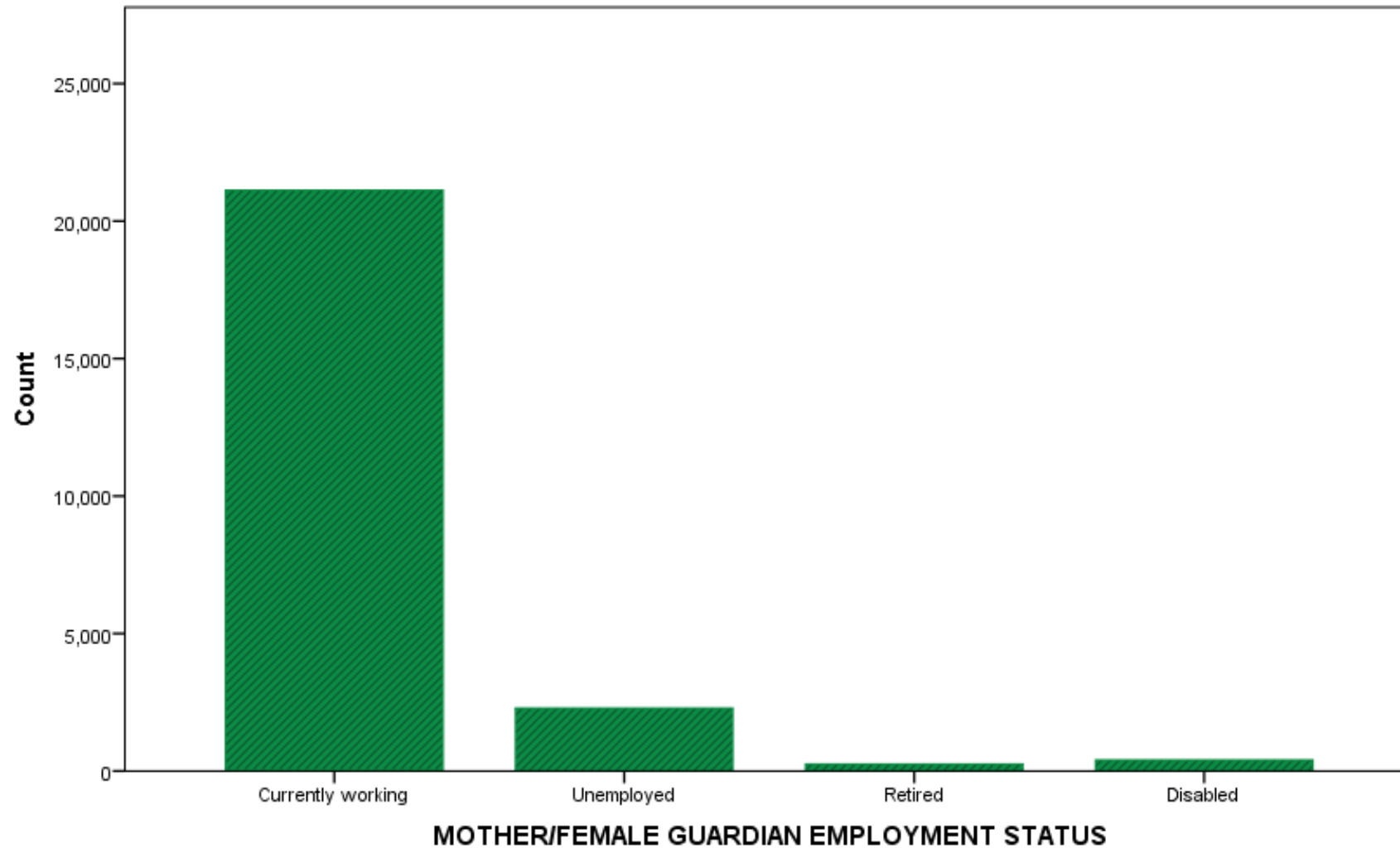


Household includes Specific Place for Study

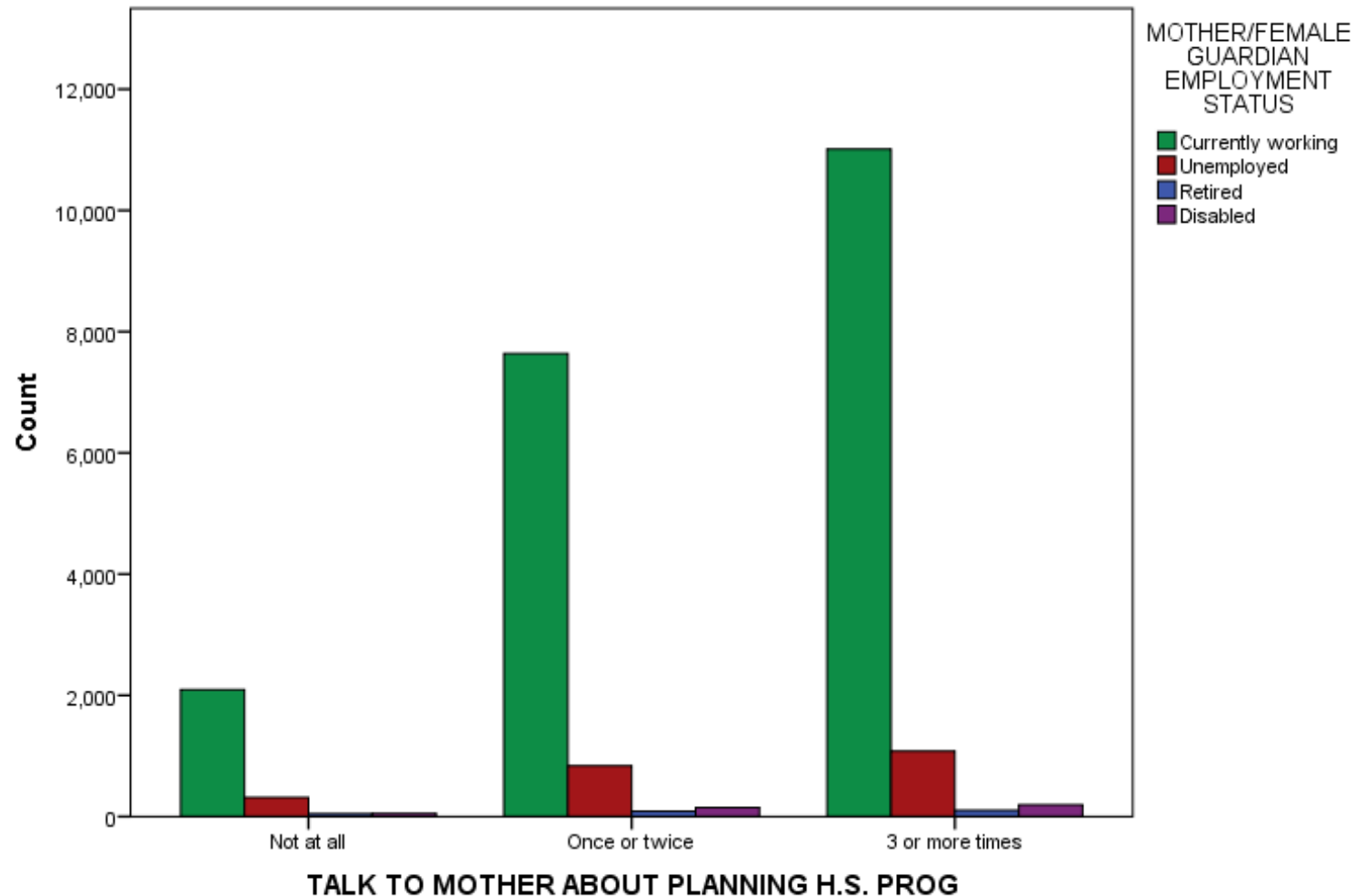
About 55% families do not have a specific place for study at home but most have more than 50 books at home



Most mothers or females guardians are employed

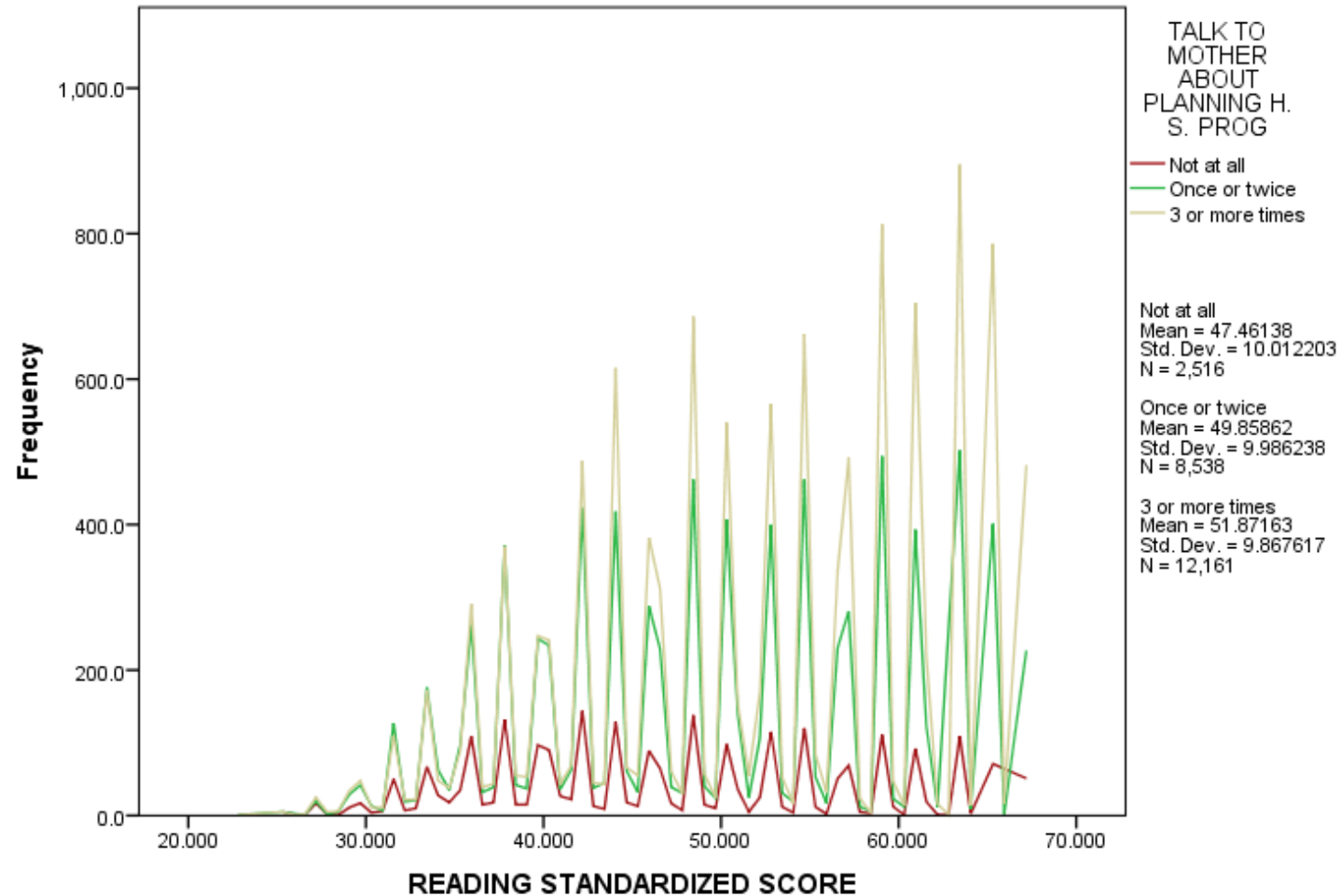


Students level of engagement clustered by employment status of mother or female guardian



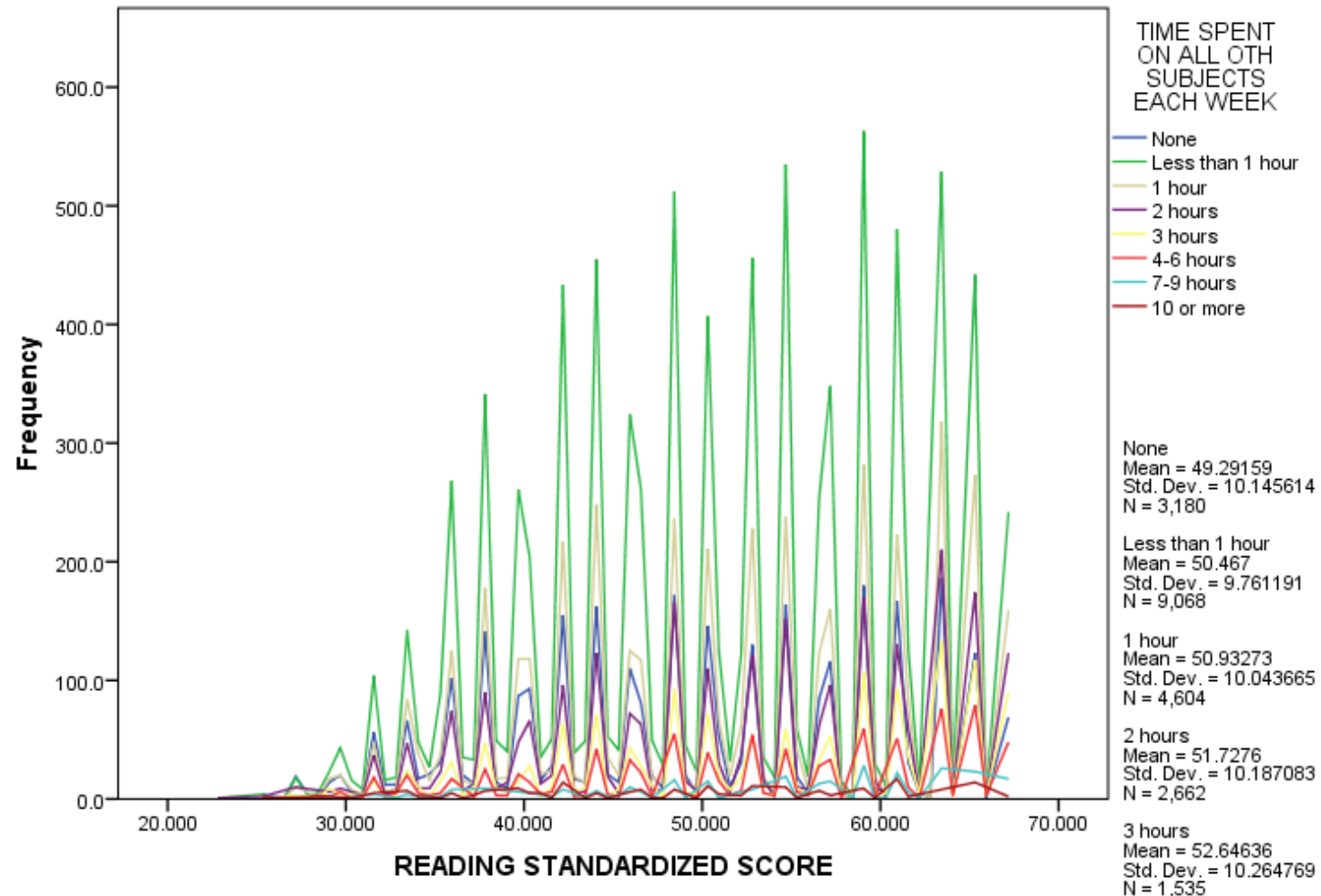
Reading Scores

clustered by students oral engagement with mother
regarding planning their higher studies



Reading Scores

clustered by students' amount of time
devoted to study subjects outside of school hours



Regression Model

To predict Reading Standardized Scores

Model Statistics:

8 Models were tested. Final Model which explains 20.3% of the variation in Reading Scores based on selected variables was selected.

Model	R square	Adjusted R square	F- Test
1	.047	.047	.000
2	.128	.128	.000
3	.125	.125	.000
4	.125	.125	.000
5	.102	.102	.000
6	.102	.102	.000
7	.203	.202	.000
8	.203	.202	.000

Model 8 – All significant variables related to reading

Variables	B	Sig.	Tolerance	VIF
(Constant)	38.441	.000		
R's family has specific place for study=Have	-1.446	.000	.941	1.062
R's family has a daily newspaper=Have	.745	.000	.968	1.034
R's family has more than 50 books=Have	3.802	.000	.936	1.068
In advanced, enriched, accelerated English=Yes	.752	.000	.975	1.026
Usually look forward to English class=1.0	-1.176	.000	.944	1.060
How much reading do you do on your own=1 hr or less per wk	2.421	.000	.543	1.840
How much reading do you do on your own =2 hours	3.706	.000	.584	1.712
How much reading do you do on your own =3 hours	5.300	.000	.680	1.471
How much reading do you do on your own =4-5 hours	7.178	.000	.751	1.331
How much reading do you do on your own =6 hrs or more per week	8.053	.000	.706	1.416
BYHOMEWK=5.50 to 10.49 hours	1.607	.000	.936	1.069
BYHOMEWK=10.50 to 12.99 hours	2.827	.000	.965	1.036
BYHOMEWK=13.00 to 20.99 hours	3.310	.000	.937	1.067
BYHOMEWK=21.00 and up hours	1.692	.000	.965	1.036
Mother/female guardian employment status =Currently working	2.595	.000	.874	1.144
Mother/female guardian employment status=Disabled	-1.240	.023	.881	1.135
Discuss school activities with parents=Once or twice	1.252	.000	.274	3.654
Discuss school activities with parents=3 or more times	3.344	.000	.250	3.999
Discuss things studied in class with parents=Once or twice	1.109	.000	.322	3.103
Discuss things studied in class with parents =3 or more times	3.111	.000	.288	3.475
Talk to mother about planning h.s. prog=Once or twice	.722	.002	.321	3.113
Talk to mother about planning h.s. prog=3 or more times	.913	.000	.293	3.408
How often parents check on r's homework=3	-1.790	.000	.667	1.498
How often parents check on r's homework=4	-2.504	.000	.640	1.561

a. Dependent Variable: READING STANDARDIZED SCORE

Summary

Final model which utilizes 11 features and 24 independent variables explains 20.3% of the variation in Reading Standardized Scores.

To reach the Final Model:

- ▶ Features were utilized for Descriptive Analysis
- ▶ New Variables/Features were created
- ▶ Total of 8 Regression Models were tested

Thank You

Shreya Goel

M.S. Learning Analytics | *Columbia University*
Marketing Analytics Intern | *Columbia University*
email: shreya.goel@columbia.edu

