暨南大學
JINAN UNIVERSITY

Course Paper for Undergraduate Students

Course Type： Subject Optional

Course Name： Introduction to Statistical Methods

Course Code： 60080079

Instructor： 梁倩茹, Qianru Liang

Course Paper for Introduction to Statistical Methods

Name H3Art

Student ID xxxxxxxxxx

College International School

Major CST

Submit Date： x / x / xxxx

Predicting Reading Scores: The Influence of Home Environment, Parental Involvement, and Student Attitudes Using NELS: 88 Data

21 CST H3Art

1. Introduction

Previous studies have demonstrated the significant impact of home environment, parental involvement, and student attitudes on reading achievement. This research extends these findings by utilizing the **National Educational Longitudinal Survey of 1988 (NELS:88)**, which collected data from 24,599 eighth-grade students across the U.S according to the survey's overview page on the National Center for Education Statistics website (https://nces.ed.gov/surveys/nels88/). This study focuses on predicting standardized reading scores using demographic information, parental and student actions, and perceptions regarding reading. By analyzing the 1988 survey results, I aim to **identify key predictors** of reading performance beyond traditional measures such as test scores and school performance.

2. Methods

The study utilized SPSS for **multiple regression analysis** to predict standardized reading scores. Three models were developed based on different sets of predictors: **home demographics**, **student actions and perceptions**, and **parental involvement**. Variables were recoded into dummy variables where necessary. Model significance was tested using the **omnibus test**, and multicollinearity was assessed with **Variance Inflation Factors (VIF)**. Non-significant variables and those causing multicollinearity were iteratively removed to refine the models. The **final model combined significant predictors from all three initial models**, achieving an adjusted R^2 of 20.2%.

2.1 Data Descriptive Statistics

The dependent variable in this study is the standardized reading score (BYTXRSTD) from the NELS:88 dataset. This variable ranges from 23.098 to 67.499, with missing values coded as 999.998. From the 23,932 observations, 882 with missing values were omitted. I selected independent variables based on previous research and logical review.

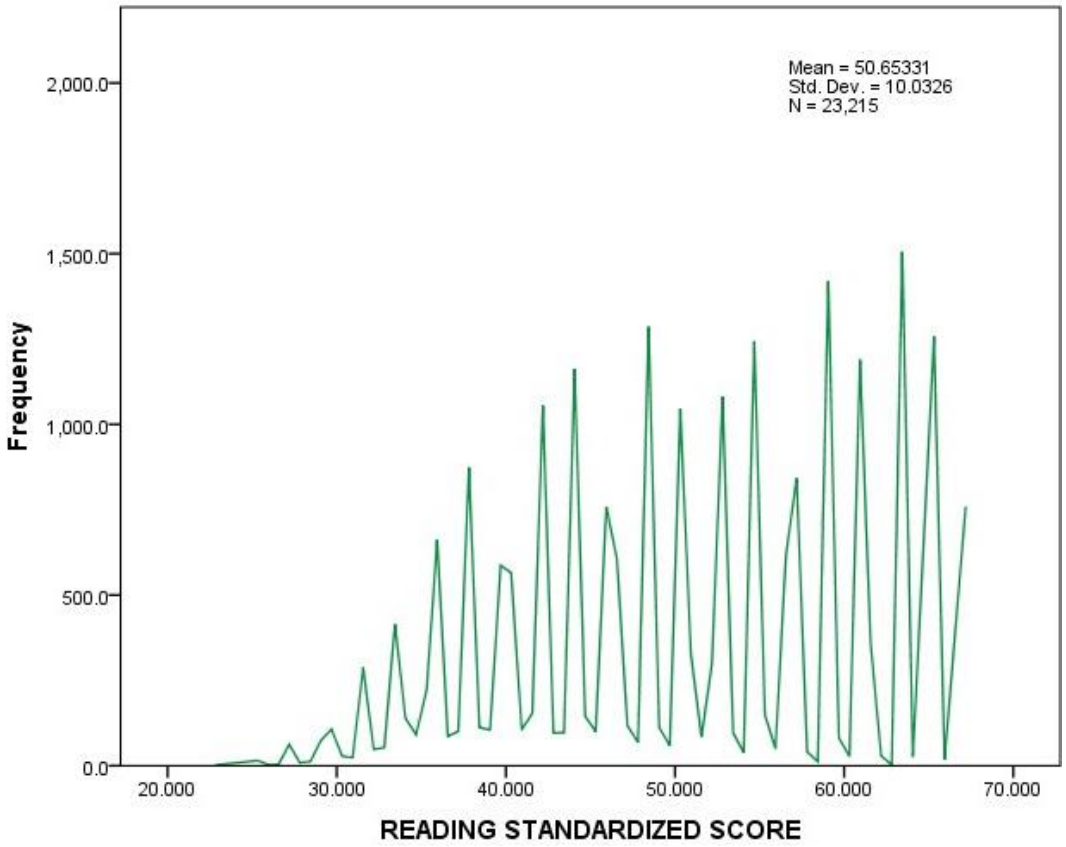

Figure 1: Score Line of Reading Standardized

Three models were created: one focusing on home demographics, another on student actions and perceptions, and a third on parental involvement. These models excluded other standardized scores to avoid multicollinearity. Each model's significance was assessed, and non-significant variables were removed to refine the final predictive model.

2.2 Question for Home Demographics Related to Reading

This model examined the impact of home demographics on reading scores using three variables:

1. Whether the family has a specific place for study (BYS35A).
2. Whether the family receives a daily newspaper (BYS35B).
3. Whether the family has more than 50 books (BYS35M).

These variables were recoded into dummy variables with "No" as the reference category. The model was significant ($p < .001$), explaining roughly 5% of the variation in reading scores. Interestingly, **having a specific place to study negatively impacted reading scores**, while having more than 50 books and receiving a daily newspaper had positive effects.

Table 1: Model of Home Demographics Related to Reading

Variables	B	Sig.	Tolerance	VIF
(Constant)	44.294	.000		

Has Specific Place for Study=Have	-.807	.000	.988	1.012
Family Has a Daily Newspaper=Have	1.380	.000	.976	1.025
Family Has More Than 50 Books=Have	6.466	.000	.974	1.027

a. Dependent Variable: READING STANDARDIZED SCORE

2.3 Question for Student Actions and Perceptions Related to Reading

This model explored the relationship between student actions and perceptions and their reading scores using several variables:

1. Hours spent on homework per week (BYHOMEWK), categorized into eight levels (1=none, 2=.5 to 1.99, 3=2 to 2.99, 4=3-5.49, 5=5.50 to 10.49, 6=10.50 to 12.99, 7=13.00 to 20.99, 8=21 hours or more).
2. Enrollment in advanced English classes (BYS66A).
3. Attitude towards English class (BYS70A).
4. Amount of reading done independently (BYS80), categorized into six levels (1=none, 2=1 hour or less, 3=2 hours or less, 4=3 hours or less, 5=3-5 hours, 6=6 or more hours and omitting missing values).

Initially, all variables showed significance ($p < .001$), explaining 12.8% of the reading score variation. Due to **high collinearity**, I removed one **homework level (3-5.49 hours)**, reducing the adjusted R^2 to 12.5%. Further adjustments excluded two **non-significant homework levels (.5-2.99 hours)**, resulting in a final model explaining 12.5% of the variation, with all remaining variables significant.

Table 2: Model of Student Actions and Perceptions Related to Reading

Variables	B	Sig.	Tolerance	VIF
(Constant)	46.516	.000		
In advanced, enriched, accelerated English=Yes	.469	.001	.981	1.020
Usually look forward to English class=1.0	-1.367	.000	.962	1.039
How much reading do you do on your own=1 hour or less per week	3.192	.000	.565	1.770
How much reading do you do on your own=2 hours	4.737	.000	.611	1.635
How much reading do you do on your own=3 hours	6.581	.000	.709	1.411
How much reading do you do on your own=4-5 hours	8.801	.000	.780	1.281
How much reading do you do on your own=6 hours or more per week	9.819	.000	.738	1.355
Number of hours spent on homework per week=5.50-10.49 hours	2.033	.000	.945	1.058
Number of hours spent on homework per week=10.50-12.99 hours	3.686	.000	.973	1.028
Number of hours spent on homework per week=13.00-20.99 hours	4.250	.000	.952	1.050
Number of hours spent on homework per week=21.00 and up hours	2.547	.000	.973	1.027

a. Dependent Variable: READING STANDARDIZED SCORE

2.4 Question for Parent Involvement and Perceptions Regarding Reading

This model assessed the impact of parental involvement and perceptions on student reading scores using the following variables:

1. Mother's employment status (BYS4A), categorized as working, unemployed, or disabled.
2. Frequency of discussing school activities with parents (BYS36B), categorized into not at all, once or twice, and three or more times.
3. Frequency of discussing things studied in class with parents (BYS36C), using the same categories.
4. Frequency of parents checking homework (BYS38A), categorized into never, rarely, sometimes, and often.
5. Frequency of talking to the mother about planning high school courses (BYS50B), using the same categories as discussions.

The initial model was significant ($p < .001$), explaining 10.2% of the reading score variation. All variables were significant **except for the retired category of mother's employment**. Removing this category, the final model maintained a 10.2% adjusted R^2, with all remaining variables significant.

Table 3: Model of Parents Involvement Related to Reading

Variables	B	Sig.	Tolerance	VIF
(Constant)	42.766			
Mother/female guardian employment status=Currently working	3.166	.000	.874	1.144
Mother/female guardian employment status=Disabled	-1.661	.000	.880	1.136
Discuss school activities with parents=Once or twice	1.563	.002	.290	3.447
Discuss school activities with parents=3 or more times	4.254	.000	.266	3.764
Discuss things studied in class with parents=Once or twice	1.723	.000	.337	2.968
Discuss things studied in class with parents=3 or more times	4.568	.000	.304	3.285
Talk to mother about planning high school prog=Once or twice	1.137	.000	.334	2.998
Talk to mother about planning high school prog=3 or more times	1.797	.000	.307	3.258
How often parents check on r's homework=2	-.512	.000	.440	2.270
How often parents check on r's homework=3	-2.367	.044	.349	2.863
How often parents check on r's homework =4	-3.005	.000	.314	3.187

a. Dependent Variable: READING STANDARDIZED SCORE

3. Results

The final model combined significant variables from the previous models to predict reading scores comprehensively. The initial model included 25 variables from home demographics, student actions and perceptions, and parental involvement. This model was significant ($p < .001$) and explained 20.2% of the variation in reading scores.

After identifying **a non-significant variable (parental homework checking frequency at the "rarely" level)**, I removed it and reran the model. The refined final model, with 24 variables, maintained an adjusted R^2 of 20.2%, achieving significance for all included variables, thus providing a robust prediction of reading scores based on a combination of demographic, student, and parental factors.

Table 4: Model of 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 hour or less per week	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 hours 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 high school prog=Once or twice	.722	.002	.321	3.113
Talk to mother about planning high school 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

I attempted to use demographic, involvement or action, and perception-related variables in order to predict Reading Standardized Scores rather than other traditional methods such as other test scores or school performance. Based on my model, I was able to predict **20.2% of the variation** in the NELS:88 Reading Standardized Scores using a combination of home demographic information relating to reading, parent involvement and perceptions regarding reading, and student actions and perceptions regarding reading.

Table 5: Model Statistics

Model	R^2	Adjusted R^2	F-test
1	.047	.047	.000
2	.125	.125	.000
3	.102	.102	.000
Final Model	**.203**	**.202**	**.000**

Given the variables I used, this is a significant prediction model. For future research, it would be logical to try other variables and/or combinations from this dataset as well as potentially explore other research surveys that might provide additional demographic or other related information that was not available in this particular dataset.

4. Conclusion

This analysis of NELS:88 data aimed to predict students' reading scores based on parental and student actions, perceptions, and home demographics. The study contributes to the literature by identifying factors beyond traditional test scores and school performance that affect reading achievement. Results indicate that home demographics, student perceptions and actions, and parental involvement significantly predict reading scores. Student perceptions and actions, along with parental involvement, have a greater impact than home demographics. The combined model explains 20.2% of the variation in standardized reading scores.