Does Household Income Affect Weight Gain in New York State

Eric Ching

Joashaha Drakes

Destini Hall

MPH 302

Dr. Taehwan Park

Introduction

Usually, income is understood to be affected by variables such as education level, location, and even in some cases, ethnicity. As such, the relationship between income and weight gain (obesity rates) involves the exposure (income) influencing the outcome weight gain (obesity rates).

Income affects the health of children as well as their development in a plethora of ways. The association between income status and weight in children is observed across geographical locations. In New York State, data collected through the state's health department for elementary and middle/high school includes weight data across the state's 57 counties along with income data. The state with its 57 counties currently reports that a third of its children have higher obesity rates when compared to their counterparts elsewhere. Income as independent variable shares an inverse correlation with other outcomes but with weight, it is observed that lower income is linked to higher risks of obesity. Obesity is a key public health issue. In children, it is an indicator for chronic health problems in this target population such as diabetes and heart disease. It is a main risk factor for all-cause deaths and poor quality of life.

Statistical Data Analysis provides a mathematical and graphical framework for understanding the relationship between income and obesity and social determinants of health provide a description for this the observed correlation. Social determinants of health show that physical conditions restrict access to health and healthy food and at the same time influence health behaviors such as exercise and eating habits as well as psychosocial factors that arise from relative distress including stress and insecurity.

Study Objectives:

1. To determine if income affects obesity rates in grade school children within New York State.
2. To investigate correlation between income and obesity rates.

Research Question: How Income Affects Obesity Rates in Grade School Children within New York State?

Hypothesis: Income has an inverse correlation to weight within grade school children. The null hypothesis is based on income having no correlation whatsoever with weight for children in grade school.

Data sources from proposal report:

- New York State Health Department
- IndexMundi

Methods

To test if household median income affected weight and obesity in students, our project used datasets from two sources. Our first dataset was from New York State Health Department (https://health.data.ny.gov/Health/Student-Weight-Status-Category-Reporting-Results-B/es3k-2aus/data), which displays the health data of students from elementary and middle/high school from various date ranges between 2010 to 2019. Specifically, this dataset shows the number of students that have a healthy weight or are considered overweight or obese within areas of each of New York State's 57 counties. The second dataset comes from Index Mundi (https://www.indexmundi.com/facts/united-states/quick-facts/new-york/median-household-income#table) a data portal which turns raw data into useful information such as visual charts, tables, and maps. This dataset provided by IndexMundi shows the population data from each of New York State Counties. Due to limitations, we could only find the census data pertaining to the household median income of New York State's 57 counties for 2018. Ideally, we would want our datasets to take place over multiple years up to 2019, however since we could only find data on 2018's household median income our study was limited to only look at students within 2018.

In order to conduct our study several adjustments were made to our student weight dataset. Due to SAS OnDemand having issues processing datasets over 3megabytes, our student weight dataset was too large and running tests became slow or even unresponsive. To allow our student weight dataset to effectively run, unnecessary data was culled. Such data that was excluded were data collected before or after 2018, data that only looked exclusively at elementary or middle/high school levels, data that exclusively looked at the number of students being overweight or obese and data that exclusively looked at male or female students. One issue with the student weight dataset was that it did not look exclusively at just New York State

Counties but within several areas within each county and did not calculate the total amount of students within one county. This would not be an issue however it was not possible to obtain the median household income of each specific area listed. In order to merge our datasets, we would need both our student weight data and our census data to have an observation of 57 (the number of New York State Counties). The data for total number of students within each county that are obese or overweight and total number of healthy weight students in each county was calculated and entered manually in order to only have an observation of 57. Ideally our study would want to control for sex, but due to time constraint, manually indexing the individual total students that are either Male, Female or Both for each County was not practical.

In our study we identified our population as "County," our dependent variable as the "Percentage of Students Overweight or Obese," and our independent variable as "Median Household Income." While we had the number of students within each county that are overweight or obese and number of healthy weight students, we did not have a percentage in our dataset. "Percentage of Students Overweight or Obese" was added in a data step of our SAS program after calculating the total number of students and dividing it with the number of students that are overweight or obese.

```
DATA schooldatatotal;
SET work.schooldata;
WHERE 'Location Code'n = 'TOTAL';
'Total Students'n = 'Number Overweight or Obese'n +'Number Healthy Weight'n;
'Percent Overweight or Obese'n = ROUND(('Number Overweight or Obese'n / 'Total Students'n) * 100, 0.1);
'Percent Healthy Weight'n = ROUND(('Number Healthy Weight'n / 'Total Students'n) * 100, 0.1);
KEEP 'county'n 'Location Code'n'Number Overweight or Obese'n 'Number Healthy Weight'n
'Total Students'n 'Percent Overweight or Obese'n 'Percent Healthy Weight'n;
RUN;
```

Figure 1. Student Weight Dataset Sorted to only include total students from each County

In order to prove if median household income affects weight gain and obesity rates, we split median household income into quartiles and classified each county into 4 separate groups. By splitting median household income into quartiles, we can effectively show whether a county has a high or low median household income. The median household income quartiles were obtained by performing a univariate procedure on the variable. After Obtaining the quartile numbers a new variable called “group” is made in a new data step by using conditional statements. Group 1 indicates the richest counties while group 4 indicates the poorest counties.

Quantiles (Definition 5)	
Level	Quantile
100% Max	111240
99%	111240
95%	96675
90%	80839
75% Q3	60736
50% Median	55673
25% Q1	52268
10%	49305
5%	47033
1%	45332
0% Min	45332

Figure 2. Quartiles obtained after using a Univariate Procedure

```
DATA censusdatarefined;
SET WORK.censusdata;
'County'n = UPCASE('County'n);
IF 'Median household income (in 2018'n < 52268  THEN 'GROUP'n = 4;
ELSE IF 'Median household income (in 2018'n >= 52268 and
 'Median household income (in 2018'n < 55673 THEN 'GROUP'n = 3;
ELSE IF 'Median household income (in 2018'n <= 60736 and
 'Median household income (in 2018'n >= 55673 THEN 'GROUP'n = 2;
 ELSE IF 'Median household income (in 2018'n >= 60736 THEN 'GROUP'n = 1;
/*Group is seperates county in quartiles (1 being the richest county and 4 being the poorest)
 by reversing the numbers the graph shows that the poorer a county is the bigger likelyhood of
 a county having a larger percentage of a population being obese or overweight. Data can be reversed*/
KEEP 'County'n  'Percentage of persons in poverty'n 'Median household income (in 2018'n 'Poverty Status'n
 'Per capita income in past 12 mon'n 'GROUP'n;
RUN;
```

Figure 3. Group variable created using quartile ranges to indicate poorest to richest groups

Since our study only had two distinctive variables, “Percentage of Students Overweight or Obese,” our dependent variable, and “Median Household Income,” our independent variable, and with no control variables, we decided that the best statistical test to use would be a simple linear regression. We chose linear regression over multiple linear regression simply due to limitation in our data. When conducting our linear regression test, we replaced our independent variable “Median Household Income” with our new variable “Group.” Since this variable separated “Median Household Income” into 4 groups, it effectively allowed us to understand the results of our linear regression test.

```
PROC REG DATA= alldata;
MODEL 'Percent Overweight or Obese'n = 'group'n;
TITLE 'Linear Regresion';
```

Figure 4. Linear Regression Test

Results

The purpose of our research is to analyze the effects that income has on obesity rates in grade school children by county in New York State. Our fit plot analyzed the percentage of children that are overweight or obese by group. Our groups went 1-4, respectfully; 2018 median

household incomes that were greater than or equal to $60,736, greater than or equal to $55,673, less than $55,673, and incomes less than $52,268. The groups are separated into quartiles with the higher quartile containing the richest counties in New York State and lessening by the increase in numbers to show that the larger the quartile the more likely that the population is overweight. As you can see, as the income of the county quartiles decreases, the percent of overweight or obese children in said quartile also increases. As with any natural environment there are some outliers where some of the wealthy counties have a high percentage of overweight children and vice versa, but for the most part, there is a steady linear relationship. Our prediction limits also mimic linear relationships between the mentioned variables.

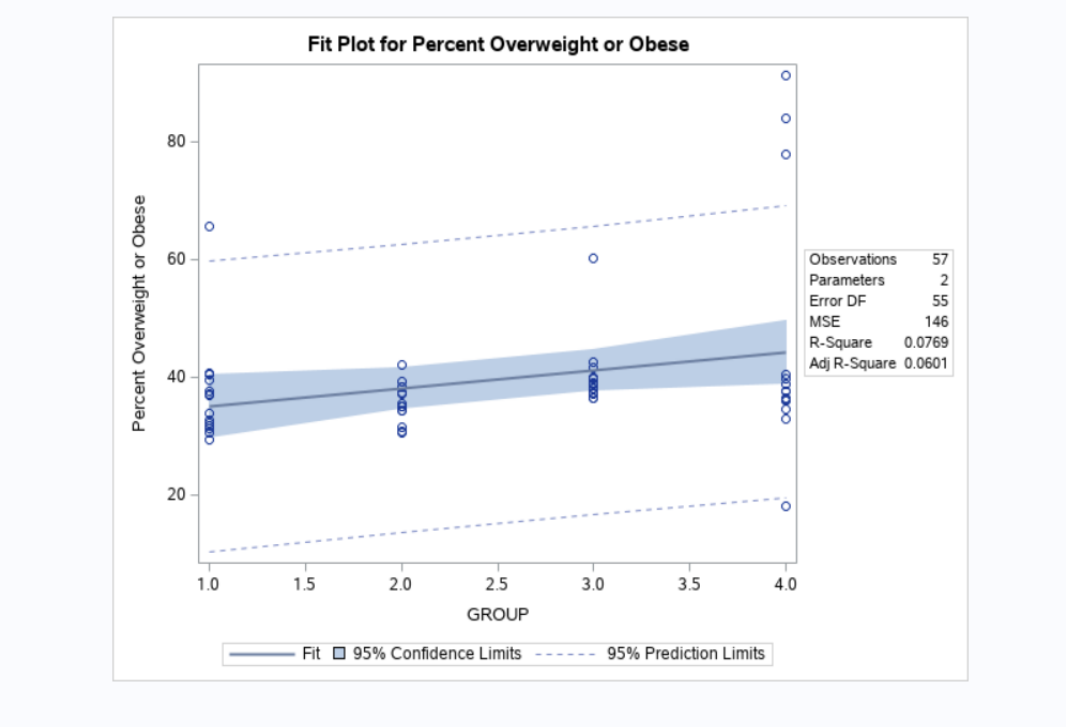

Figure 5. Fit Plot Graph

First, we will mention our R-Square value which is a goodness-of-fit measure for our below procedure. This number determines what percentage of variation in our dependent variable

(percentage overweight or obese) is explained by our independent variable, which in this specific procedure is the group variable (grouped county incomes by quartile). Our R-squared of roughly .077 means that our group variable is 77% responsible for the variation in the percentage of overweight or obese children. Our parameter estimate is the analysis of our coefficients used in this specific linear regression procedure. Our intercept of around 32 shows that when controlling for all other variables, 31% of children would still be overweight or obese. As our group variable increases by 1 unit (or 1 quartile), while controlling for all other variables, the percentage of children that are overweight or obese would increase by approximately 3.1% per quartile. Lastly, we will discuss the p-value which allows us to validate our hypothesis against the research question at hand. Once again, we are supporting the idea that income variation ('group' variable) influences percentages of overweight or obese children in New York State. A p-value of .037, which is less than <.05, shows that we have statistically significant evidence against the null hypotheses. In more specificality, there is a 3.7% chance that the null hypothesis is even true.

Linear Regresion

The REG Procedure
Model: MODEL1
Dependent Variable: Percent Overweight or Obese

Number of Observations Read	57
Number of Observations Used	57

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	668.61476	668.61476	4.58	0.0368
Error	55	8029.92033	145.99855		
Corrected Total	56	8698.53509			

Root MSE	12.08299	R-Square	0.0769
Dependent Mean	39.63860	Adj R-Sq	0.0601
Coeff Var	30.48288		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	31.95275	3.93197	8.13	<.0001
GROUP	1	3.08516	1.44167	2.14	0.0368

Figure 6. Linear Regression Procedure

Application

This data that has been analyzed can be applied to public health measures as obesity in grade school children is classified as a nation-wide problem (CDC, 2021). There is plenty of information on how the government can help underprivileged areas gain access to healthier foods, such as supporting the local pantries with fresh produce and grains that align with the U.S. Department of Agriculture's MyPlate initiative (USDA, 2022). It is also important to note the variance in the types of meals being served in grade schools that are in poorer neighborhoods and how funding plays a role in 1-2 meals (breakfast and lunch) that students may eat a day while in their learning environments. With that much control on student health such as nutrient intake, it is an important thing to note how counties divide their funds per district and per school.

Conclusion

The findings of this research study suggest that there is a relationship between income and weight gain (obesity rates) as there is statistical significance. Income and weight gain (obesity rates) share an inversely proportional relationship whereby a low income is related to weight gain. To combat this, access must be widened to ensure that healthy food and care facilities, as well as providers, are within reach of individuals regardless of income. Programs such as Supplemental Nutrition Assistance Program (SNAP) and Women, Infants, and Children (WIC) are known nutrition programs that have been established and made available to New York Stae residents. Yet, many households in New York State struggle to put meals 'on the table' and healthy ones at that. This is due to the need for policy changes at the legislative level of the state

that would see the expansion of the eligibility criteria (that currently excludes and/or inhibits immigrants from accessing) and provision of culturally competent meal planning assistance to serve individuals from all backgrounds. Further, it would ensure that aid (like the current grant from the Robert Wood Johnson Foundation's Healthy Kids, Healthy Communities Fund to battle overweight and obesity in children) is utilized for all and not just a select few.

The analysis of income and obesity revealed a significant linkage between the two. However, the extent to which this relationship exists is not strong. This could be due to a protective effect as a result of confounders that have a separate relationship with both the independent variable (income) and the dependent variable (weight gain). In this case, the role of confounders in the relationship between income and weight gain needs to be explored to determine the true extent of the causal relationship between the two variables. Moreover, reversed causality may exist between the two variables, and in this case, it could be hypothesized that obesity is not the outcome but rather the exposure that leads to lowered income. Therefore, an examination of probable reverse causality may be required to more thoroughly tackle income-correlated health inequalities and inequities in obesity.

Reference

"Are You Making Every Bite Count?" *MyPlate*, U.S. Department of Agriculture, http://www.myplate.gov/.

Goisis A, Sacker A, Kelly Y. Why are poorer children at higher risk of obesity and overweight? A UK cohort study. *Eur J Public Health* 2016;26:7–13. 10.1093/eurpub/ckv219

Hoyt LT, Kushi LH, Leung CW, et al. . Neighborhood influences on girls' obesity risk across the transition to adolescence. *Pediatrics* 2014;134:942–9. 10.1542/peds.2014-1286

Kakinami L, Séguin L, Lambert M, et al. . Poverty's latent effect on adiposity during childhood: evidence from a Québec birth cohort. *J Epidemiol Community Health* 2014;68:239–45. 10.1136/jech-2012-201881

"New York Median Household Income (in 2018 Dollars), 2014-2018 by County." *New York Median Household Income (in 2018 Dollars), 2014-2018 by County*, IndexMundi, https://www.indexmundi.com/facts/united-states/quick-facts/new-york/median-household-income#table.

"Obesity." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 21 Sept. 2021, https://www.cdc.gov/healthyschools/obesity/index.htm.

"Student Weight Status Category Reporting Results: Beginning 2010: State of New York." *Student Weight Status Category Reporting Results: Beginning 2010*, New York State

Department of Health, https://health.data.ny.gov/Health/Student-Weight-Status-Category-Reporting-Results-B/es3k-2aus/data.

```
/*Eric Ching
Joashaha Drakes
  Destini Hall
*/


PROC IMPORT
DATAFILE =
        '/home/u60876078/data_analysis/SAS/CLASS/Project/Student_Weight_Status_Categor
        y_Reporting_Results__Beginning_2010 total.xlsx'
OUT= schooldata DBMS = xlsx
REPLACE;
GETNAMES = yes;
RUN;
/*PROC PRINT DATA = schooldata;*/

DATA schooldatatotal;
SET work.schooldata;
WHERE 'Location Code'n = 'TOTAL';
'Total Students'n = 'Number Overweight or Obese'n +'Number Healthy Weight'n;
'Percent Overweight or Obese'n = ROUND(('Number Overweight or Obese'n / 'Total
        Students'n) * 100, 0.1);
'Percent Healthy Weight'n = ROUND(('Number Healthy Weight'n / 'Total Students'n) * 100,
        0.1);
KEEP 'county'n 'Location Code'n'Number Overweight or Obese'n 'Number Healthy Weight'n
'Total Students'n 'Percent Overweight or Obese'n 'Percent Healthy Weight'n;
RUN;

PROC PRINT DATA = schooldatatotal; RUN;
/*PROC CONTENTS DATA= schooldatatotal; RUN;*/

PROC IMPORT
DATAFILE = '/home/u60876078/data_analysis/SAS/CLASS/Project/2018 income and
        poverty percentage.xlsx'
OUT= censusdata DBMS = xlsx
REPLACE;
GETNAMES = yes;
RUN;

/*PROC PRINT DATA= censusdata; RUN;/*
/*PROC CONTENTS DATA= censusdata; RUN;*/
/*
Observations for both schooldatatotal and censusdata must be 57 to reference all 57 countys
        once
*/
```

```
DATA censusdatarefined;
SET WORK.censusdata;
'County'n = UPCASE('County'n);
IF 'Median household income (in 2018'n < 52268  THEN 'GROUP'n = 4;
ELSE IF 'Median household income (in 2018'n >= 52268 and
'Median household income (in 2018'n < 55673 THEN 'GROUP'n = 3;
ELSE IF 'Median household income (in 2018'n <= 60736 and
'Median household income (in 2018'n >= 55673 THEN 'GROUP'n = 2;
ELSE IF 'Median household income (in 2018'n >= 60736 THEN 'GROUP'n = 1;
/*Group is seperates county in quartiles (1 being the richest county and 4 being the poorest)
by reversing the numbers the graph shows that the poorer a county is the bigger likelyhood of
a county having a larger percentage of a population being obese or overweight. Data can be
      reversed*/
KEEP 'County'n        'Percentage of persons in poverty'n  'Median household income (in
      2018'n 'Poverty Status'n
'Per capita income in past 12 mon'n 'GROUP'n;
RUN;

/*PROC PRINT DATA= censusdatarefined; RUN;*/

PROC SORT DATA = schooldatatotal; BY county; RUN;
PROC SORT DATA = censusdatarefined; BY county; RUN;
DATA alldata;
MERGE censusdatarefined schooldatatotal;
by 'county'n;
KEEP 'county'n 'Percent Overweight or Obese'n 'Percent Healthy Weight'n       'Median
      household income (in 2018'n 'Poverty Status'n 'GROUP'n;
RUN;

PROC CONTENTS DATA=alldata; RUN;
PROC PRINT DATA = alldata; RUN;
TITLE 'All Data';
RUN;
/*
proc freq data=alldata;
where 'group'n = 1;
run;
*/

ODS GRAPHICS ON;
PROC UNIVARIATE DATA= alldata;
VAR 'Median household income (in 2018'n;
```

```
HISTOGRAM 'Median household income (in 2018'n /NORMAL; PROBPLOT 'Median
      household income (in 2018'n;
RUN;

PROC REG DATA= alldata;
MODEL 'Percent Overweight or Obese'n = 'group'n;
TITLE 'Linear Regresion';
/*
R square is less than 1 meaning that there is little effect
*/
/*
As income decreases per quartile there is a 3% predticted increase in the rate of obesity and
      overweightness
*/
```