

Chloe Chen, Songqiao Li, Yiqing Xin, Milca Girmai, Elizabeth Polston
 STA 198
 Professor Drake
 Healthy Kids Project Report

Healthy Kids Project Report

I. Abstract

With two data sets, T3 (time of food survey)/T4 (BMI measurement) and T9 (time of food survey)/T12 (BMI measurement), this report will respond to three questions on the Healthy Kids Project statement by establishing a food scale representing the healthy level of food intake and a proinflammatory scale showing the inflammation condition in kids' bodies. The report will reveal the relationship among BMI, food scale, and proinflammatory scale by utilizing descriptive analysis including mean and boxplot, correlational analysis, and regression analysis including the adjusted r square.

II. Introduction

In this project, we are going to answer the question of if BMI is correlated with the food intake of kids by introducing two scales: vegetable scale and proinflammatory scale. Moreover, we will use two data sets HealthyKidsT3T4 and HealthyKidsT9T12 to find the association between BMI and inflammation in kids' bodies.

In question 1, we aim to find out whether fruit and vegetable intake is related to BMI. We want to investigate this issue at both T3 (time of food survey)/T4 (BMI measurement) and T9 (time of food survey)/T12 (BMI measurement). To solve this question, we will provide some informative variable descriptives that involve tabulation, means, variances, quantiles, and boxplots for continuous variables at both time points respectively. A scale is created by adding all the vegetable intake responses into a single variable. A high value in the scale indicates a healthy diet, a low value indicates an unhealthy diet. This single variable is used to study the association between BMI and vegetation intake, in order to generate some representative insights concerning the variation among all intake levels (and time). We used the regression and correlation approaches for both time points stacked and for targeted variables at different time points.

In question 2, we focus on the correlation between the proinflammatory scale and the vegetable scale by constructing a proinflammatory scale with rankt4TNF, rankt4IL6, and rankt4CRP variables. To be more specific, we will add values of the above three variables together to form the proinflammatory scale and categorize it into three levels. The scale that is less than 100 is the low level, while the one between 100 and 200 is the median level. And the value being greater than 200 is a high level. A low level represents mild inflammation in kids' bodies whereas a high level shows severe inflammation. In the analysis part, we will show descriptives of two proinflammatory scales derived from T3/T4 and T9/T12 such as tabulation, means, standard deviation, and boxplots. Then, we will find the correlation between the proinflammatory scale and vegetable scale, or say food scale.

In this question 3, we were analyzing if there was an association between the proinflammatory scale score at T12 and the BMI at T4. To do this, we first did a simple regression comparing the BMI and proinflammatory scale at the same time point, for both T4 and T12. Then, I performed a regression on BMI at T4 and proinflammatory scale at T12 to determine whether BMI could be a predictor of future proinflammatory scores.

III. Statistical Methods and Data

In this report, the data consisted of 143 observations of multiple variables at 4 different time points (T3, T4, T9 and T12). The observations were children between 3 and 5 years old. T3 and T4 were taken as the initial values and T9 and T12 were the values at a return visit a year later. The variables of interest were several answers to a questionnaire on diet as well as proinflammatory markers in blood samples and BMI. The main methods utilized to analyze data are descriptive analysis, correlational analysis, and regression analysis.

IV. Analysis and Results

Question 1

To investigate the relationship between vegetable intake and the BMI, we first added up all the scores from the survey related to vegetable intake to form a vegetable scale, and then divided it into 3 categories.

Frequency

cat10q34	Frequency	Percent	Cumulative Frequency	Cumulative Percent
cat1	16	11.19	16	11.19
cat2	77	53.85	93	65.03
cat3	50	34.97	143	100.00

cat10q912	Frequency	Percent	Cumulative Frequency	Cumulative Percent
cat1	49	34.27	49	34.27
cat2	59	41.26	108	75.52
cat3	35	24.48	143	100.00

We can see from the table that in both time points, category 2 is the one with highest frequency.

T3T4 descriptive statistics

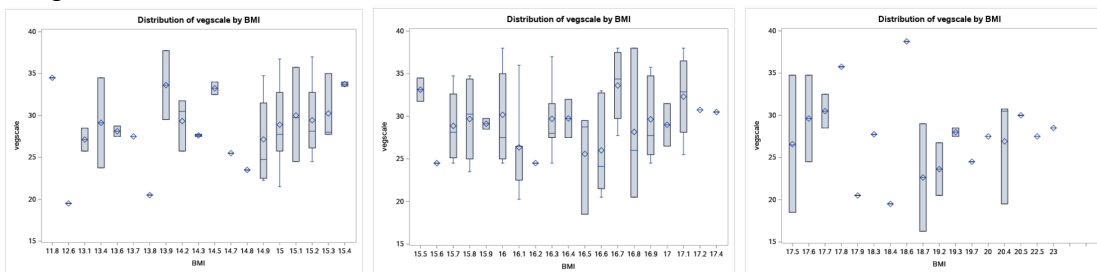
cat10q34	N Obs	Variable	Label	N	Minimum	Lower Quartile	Median	Upper Quartile	Maximum	Mean	Std Dev
cat1	16	T3HKQ9	T3HKQ9	6	1.000	2.000	3.000	3.000	3.000	2.500	0.837
		T3HKQ13	T3HKQ13	6	2.000	2.000	3.000	3.000	4.000	2.833	0.753
		T3HKQ25	T3HKQ25	6	1.000	3.000	3.500	4.000	5.000	3.333	1.366
		T3HKQ26	T3HKQ26	6	1.000	1.000	2.000	3.000	5.000	2.333	1.506
		T3HKQ27	T3HKQ27	6	1.250	2.500	2.500	2.500	2.500	2.292	0.510
		T3HKQ29	T3HKQ29	6	1.000	1.000	1.500	2.000	2.000	1.500	0.548
		T3HKQ37	T3HKQ37	6	1.000	1.000	2.000	2.000	2.000	1.667	0.516
		T3HKQ39	T3HKQ39	6	1.000	2.000	2.000	3.000	3.000	2.167	0.753
		vegscale		6	16.250	18.500	19.000	19.500	19.500	18.625	1.262
cat2	77	T3HKQ9	T3HKQ9	77	1.000	2.000	3.000	3.000	5.000	2.844	0.708
		T3HKQ13	T3HKQ13	77	2.000	4.000	5.000	5.000	5.000	4.286	0.825
		T3HKQ25	T3HKQ25	77	1.000	3.000	4.000	5.000	5.000	3.519	1.382
		T3HKQ26	T3HKQ26	77	1.000	3.000	3.000	4.000	5.000	3.325	0.966
		T3HKQ27	T3HKQ27	77	1.250	2.500	2.500	3.750	5.000	2.971	0.837
		T3HKQ29	T3HKQ29	77	1.000	2.000	2.000	3.000	5.000	2.299	0.796
		T3HKQ37	T3HKQ37	77	2.000	3.000	3.000	4.000	5.000	3.481	0.940
		T3HKQ39	T3HKQ39	77	1.000	2.000	3.000	4.000	5.000	3.091	1.194
		vegscale		77	20.250	24.500	26.500	27.750	30.000	25.815	2.649
cat3	50	T3HKQ9	T3HKQ9	50	2.000	3.000	3.000	3.000	5.000	3.320	0.844
		T3HKQ13	T3HKQ13	50	4.000	5.000	5.000	5.000	5.000	4.920	0.274
		T3HKQ25	T3HKQ25	50	1.000	4.000	4.000	5.000	5.000	4.220	0.954
		T3HKQ26	T3HKQ26	50	2.000	5.000	5.000	5.000	5.000	4.560	0.733
		T3HKQ27	T3HKQ27	50	2.500	2.500	3.750	5.000	5.000	3.825	0.959
		T3HKQ29	T3HKQ29	50	2.000	3.000	4.000	5.000	5.000	3.900	0.995
		T3HKQ37	T3HKQ37	50	3.000	5.000	5.000	5.000	5.000	4.780	0.507
		T3HKQ39	T3HKQ39	50	3.000	4.000	5.000	5.000	5.000	4.580	0.642
		vegscale		50	30.500	31.750	34.500	35.750	38.750	34.105	2.401

T9T12 descriptive statistics

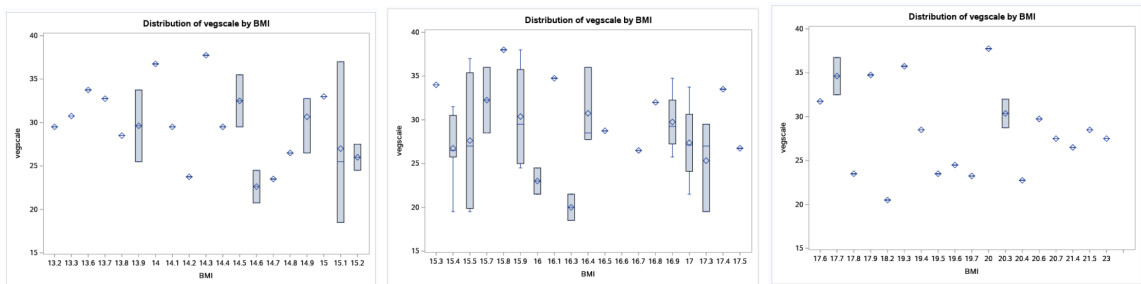
cat10q912	N Obs	Variable	Label	N	Minimum	Lower Quartile	Median	Upper Quartile	Maximum	Mean	Std Dev
cat1	49	T9HKQ9	T9HKQ9	5	2.000	3.000	3.000	3.000	3.000	2.800	0.447
		T9HKQ13	T9HKQ13	5	2.000	2.000	2.000	2.000	3.000	2.200	0.447
		T9HKQ25	T9HKQ25	5	3.000	3.000	4.000	4.000	5.000	3.800	0.837
		T9HKQ26	T9HKQ26	5	2.000	2.000	2.000	3.000	3.000	2.400	0.548
		T9HKQ27	T9HKQ27	5	2.500	2.500	2.500	2.500	2.500	2.500	0.000
		T9HKQ29	T9HKQ29	5	1.000	1.000	1.000	2.000	2.000	1.400	0.548
		T9HKQ37	T9HKQ37	5	2.000	2.000	2.000	2.000	2.000	2.000	0.000
		T9HKQ39	T9HKQ39	5	1.000	1.000	2.000	3.000	3.000	2.000	1.000
		vegscale		5	18.500	18.500	19.500	19.500	19.500	19.100	0.548
cat2	59	T9HKQ9	T9HKQ9	59	1.000	3.000	3.000	3.000	5.000	3.000	0.719
		T9HKQ13	T9HKQ13	59	2.000	4.000	5.000	5.000	5.000	4.305	0.856
		T9HKQ25	T9HKQ25	59	1.000	3.000	4.000	4.000	5.000	3.492	1.331
		T9HKQ26	T9HKQ26	59	1.000	3.000	3.000	4.000	5.000	3.424	0.855
		T9HKQ27	T9HKQ27	59	1.250	2.500	2.500	3.750	5.000	2.903	0.785
		T9HKQ29	T9HKQ29	59	1.000	2.000	2.000	3.000	5.000	2.305	0.836
		T9HKQ37	T9HKQ37	59	1.000	3.000	3.000	4.000	5.000	3.576	0.951
		T9HKQ39	T9HKQ39	59	1.000	2.000	3.000	5.000	5.000	3.305	1.316
		vegscale		59	20.250	24.500	26.500	28.500	29.750	26.309	2.693
cat3	35	T9HKQ9	T9HKQ9	35	2.000	3.000	3.000	3.000	5.000	3.143	0.692
		T9HKQ13	T9HKQ13	35	3.000	5.000	5.000	5.000	5.000	4.857	0.430
		T9HKQ25	T9HKQ25	35	2.000	4.000	5.000	5.000	5.000	4.400	0.946
		T9HKQ26	T9HKQ26	35	3.000	4.000	5.000	5.000	5.000	4.600	0.651
		T9HKQ27	T9HKQ27	35	2.500	3.750	3.750	5.000	5.000	3.929	0.865
		T9HKQ29	T9HKQ29	35	3.000	3.000	4.000	5.000	5.000	3.971	0.822
		T9HKQ37	T9HKQ37	35	4.000	4.000	5.000	5.000	5.000	4.743	0.443
		T9HKQ39	T9HKQ39	35	3.000	5.000	5.000	5.000	5.000	4.771	0.490
		vegscale		35	30.500	32.750	34.000	36.000	38.000	34.414	2.133

We can see from the table that in both time points, category 2 is the one with highest frequency.

Boxplot T3T4

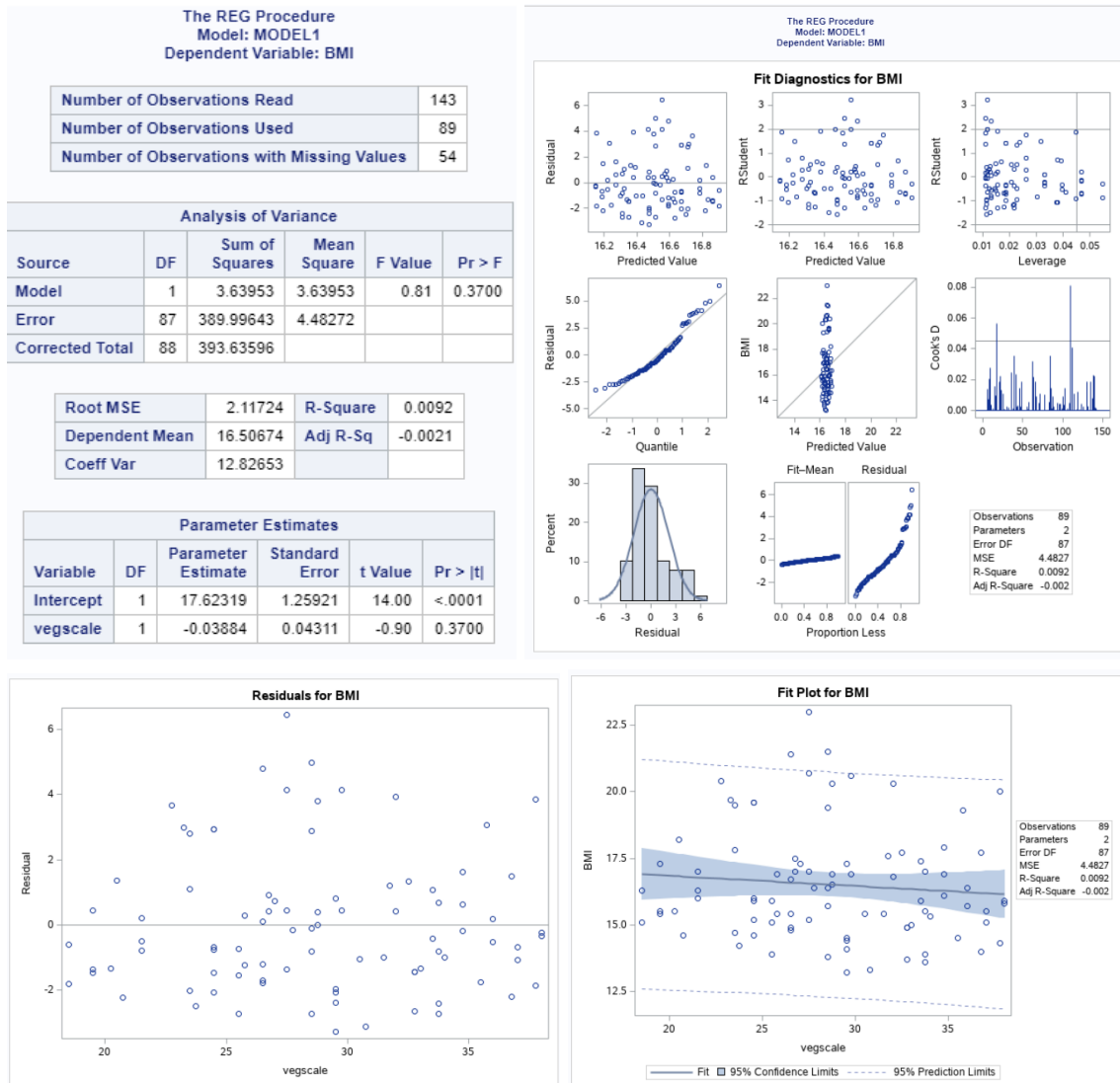


Boxplot T9T12



Looking at the boxplots, we can see that higher BMI results at a lower veg scale. And the vegscales for T9T12 are more spread out and slightly higher than T3T4.

Regression for stacked dataset (both)



In testing the association between the scale and BMI, we first looked at the regression analysis. The regression model with dependent variable of BMI and independent variable of vegscale is calculated to be $BMI = 17.62319 - 0.03884 (\text{vegscale})$. The negative coefficient of vegscale suggests that more vegetable intake results in lower BMI. However, the R-squared value of the model is 0.0092, which is a really low value, indicating that the independent variable is not explaining much in the variation of the dependent variable. Also, the small t-test and a high p-value of 0.37 suggests that we fail to reject the null hypothesis, and there are no significant correlations between the two variables. The residual plot has dispersive dots that indicate no normality exists.

Correlation for T3 scale and T12 BMI

The CORR Procedure

2 Variables: T12BMI vegscale34

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum	Label
T12BMI	93	16.50215	2.06897	16.10000	13.20000	23.00000	T12BMI
vegscale34	133	28.60714	5.17362	27.75000	16.25000	38.75000	

Pearson Correlation Coefficients

Prob > |r| under H0: Rho=0
Number of Observations

	T12BMI	vegscale34
T12BMI	1.00000	-0.06398
T12BMI		0.5423
	93	93
vegscale34	-0.06398	1.00000
	0.5423	
	93	133

Spearman Correlation Coefficients

Prob > |r| under H0: Rho=0
Number of Observations

	T12BMI	vegscale34
T12BMI	1.00000	-0.09088
T12BMI		0.3863
	93	93
vegscale34	-0.09088	1.00000
	0.3863	
	93	133

Kendall Tau b Correlation Coefficients

Prob > |tau| under H0: Tau=0
Number of Observations

	T12BMI	vegscale34
T12BMI	1.00000	-0.05783
T12BMI		0.4210
	93	93
vegscale34	-0.05783	1.00000
	0.4210	
	93	133

Next, we looked at the correlation between T12BMI and vegscale34. The Pearson correlation coefficient has a negative value of -0.06398, which suggests that there is a slight inverse correlation between the two variables. When one variable increases, the other one will slightly decrease.

Question 2

The main idea in this question is to discover the relationship between proinflammatory scale and food scale. In order to achieve this goal, we first construct the proinflammatory scale by adding three biomarkers together. And we find out there are more children who get a high inflammation at T3T4 time than those children T9T12 time. Also based on the following boxplot, we see there are more children with low-level inflammation at T9T12 than T3T4. This means the inflammation in kids' bodies may go down to the low level after several months.

Frequency for T3T4 and T9T12 respectively:

The FREQ Procedure

cat3bio34	Frequency	Percent	Cumulative Frequency	Cumulative Percent
hig	9	8.41	9	8.41
low	55	51.40	64	59.81
med	43	40.19	107	100.00
Frequency Missing = 36				

The FREQ Procedure

cat3bio912	Frequency	Percent	Cumulative Frequency	Cumulative Percent
low	76	60.32	76	60.32
med	50	39.68	126	100.00
Frequency Missing = 17				

Descriptive statistics for T3T4 and T9T12 respectively:

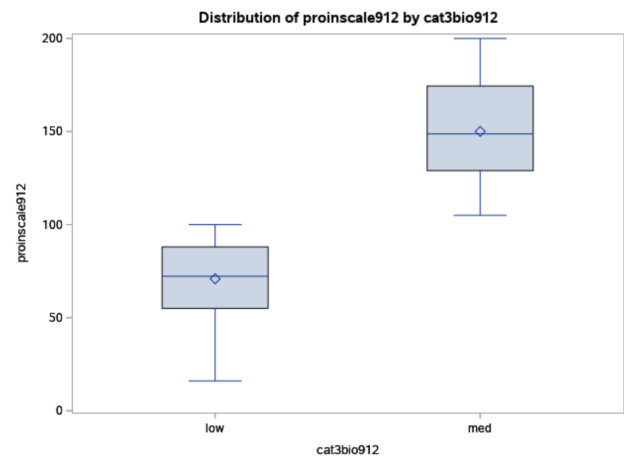
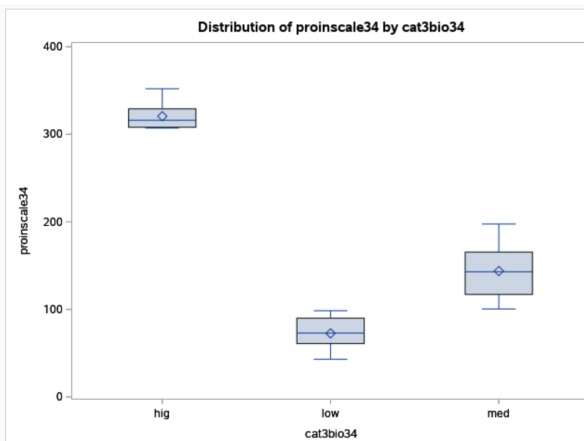
The MEANS Procedure

cat3bio34	N Obs	Variable	Label	N	Minimum	Lower Quartile	Median	Upper Quartile	Maximum	Mean	Std Dev
hig	9	proinscale34 T4BMI	T4BMI	9	307.000 13.400	308.000 14.200	316.000 15.900	329.000 17.600	352.000 22.500	320.556 16.322	15.509 2.826
low	55	proinscale34 T4BMI	T4BMI	13 45	43.000 13.600	61.000 15.000	73.000 16.200	90.000 17.000	98.500 20.500	72.615 16.273	18.662 1.784
med	43	proinscale34 T4BMI	T4BMI	43 43	100.500 11.800	117.000 15.200	143.000 16.000	165.500 16.900	197.500 20.400	143.767 16.000	28.777 1.722

The MEANS Procedure

cat3bio912	N Obs	Variable	Label	N	Minimum	Lower Quartile	Median	Upper Quartile	Maximum	Mean	Std Dev
low	76	proinscale912 T12BMI	T12BMI	26 31	16.000 13.200	55.000 14.400	72.250 15.900	88.000 16.600	100.000 21.400	70.923 15.874	20.590 1.888
med	50	proinscale912 T12BMI	T12BMI	50 47	105.000 13.800	129.000 15.200	148.750 16.000	174.500 17.600	200.000 23.000	150.090 16.702	27.982 2.098

Boxplot for T3T4 and T9T12 respectively



Correlation of proinflammatory scale and vegetable scale at T3T4 and T9T12 respectively

The CORR Procedure

2 Variables: proinscale34 vegscale34

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
proinscale34	101	188.82673	79.35948	19072	43.00000	352.00000
vegscale34	133	28.60714	5.17362	3805	16.25000	38.75000

Pearson Correlation Coefficients Prob > r under H0: Rho=0 Number of Observations		
	proinscale34	vegscale34
proinscale34	1.00000 101	-0.11536 0.2507 101
vegscale34	-0.11536 0.2507 101	1.00000 133

The CORR Procedure

2 Variables: proinscale912 vegscale912

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
proinscale912	93	144.13441	61.80220	13405	16.00000	277.00000
vegscale912	99	28.81061	5.06704	2852	18.50000	38.00000

Pearson Correlation Coefficients Prob > r under H0: Rho=0 Number of Observations		
	proinscale912	vegscale912
proinscale912	1.00000 93	-0.18628 0.0805 89
vegscale912	-0.18628 0.0805 89	1.00000 99

Regression of the proinflammatory scale and the vegetable scale at T3T4

The REG Procedure

Model: MODEL1

Dependent Variable: proinscale34

Number of Observations Read	143
Number of Observations Used	101
Number of Observations with Missing Values	42

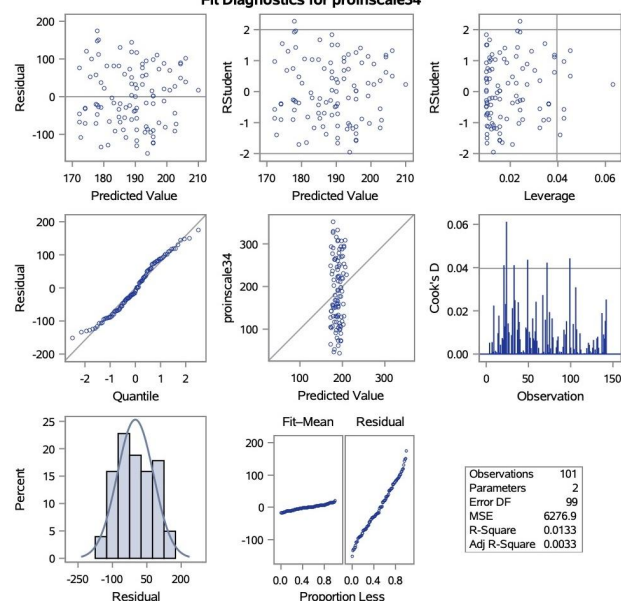
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	8380.87377	8380.87377	1.34	0.2507
Error	99	621412	6276.88731		
Corrected Total	100	629793			

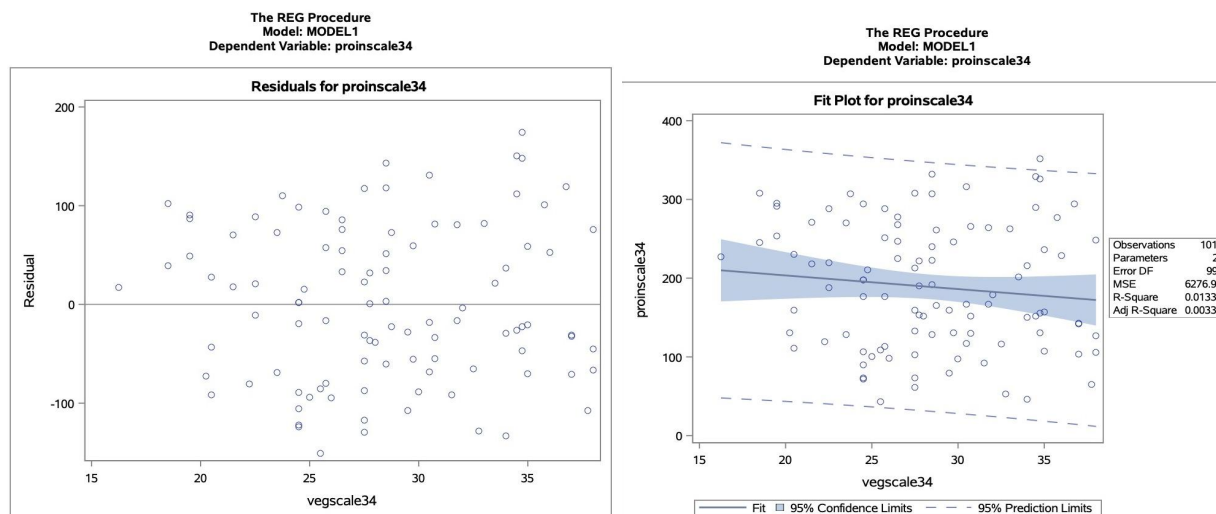
Root MSE	79.22681	R-Square	0.0133
Dependent Mean	188.82673	Adj R-Sq	0.0033
Coeff Var	41.95741		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	238.10383	43.36797	5.49	<.0001
vegscale34	1	-1.73248	1.49933	-1.16	0.2507

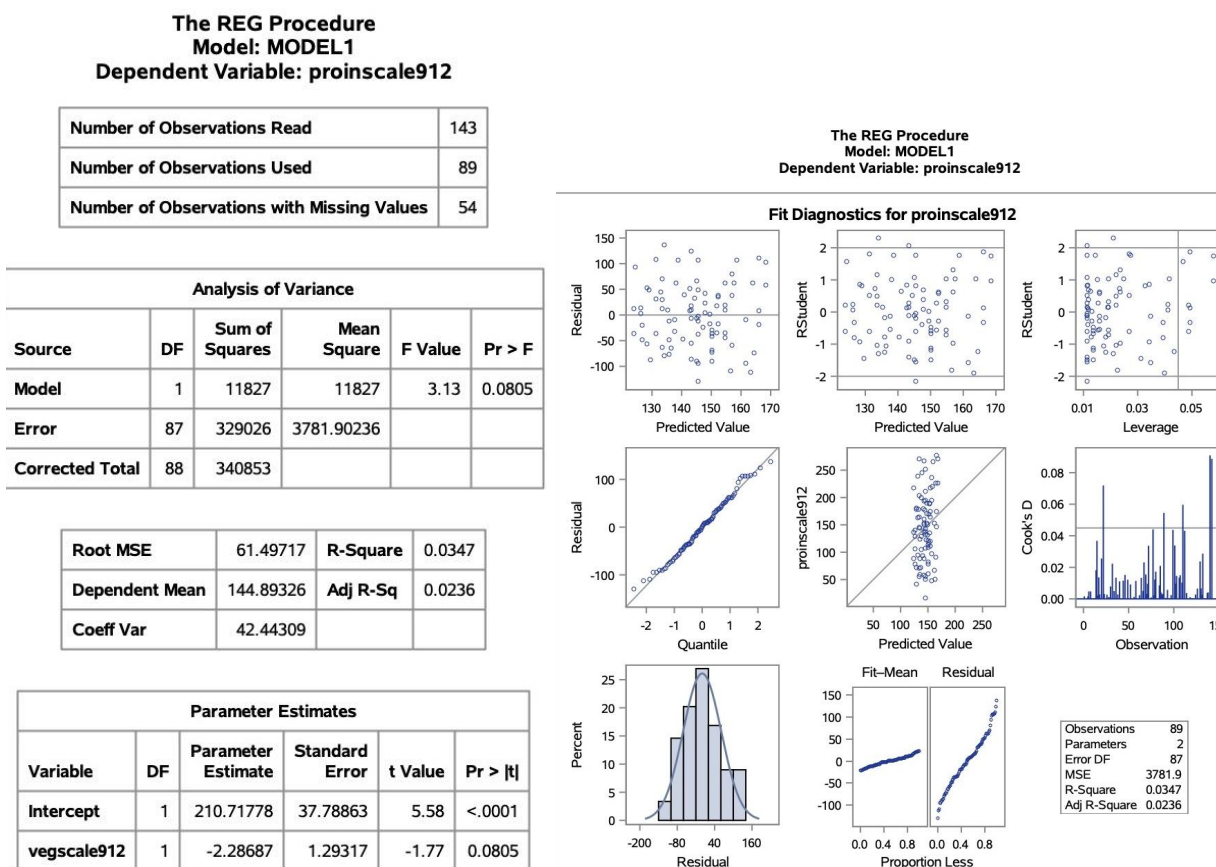
The REG Procedure
Model: MODEL1
Dependent Variable: proinscale34

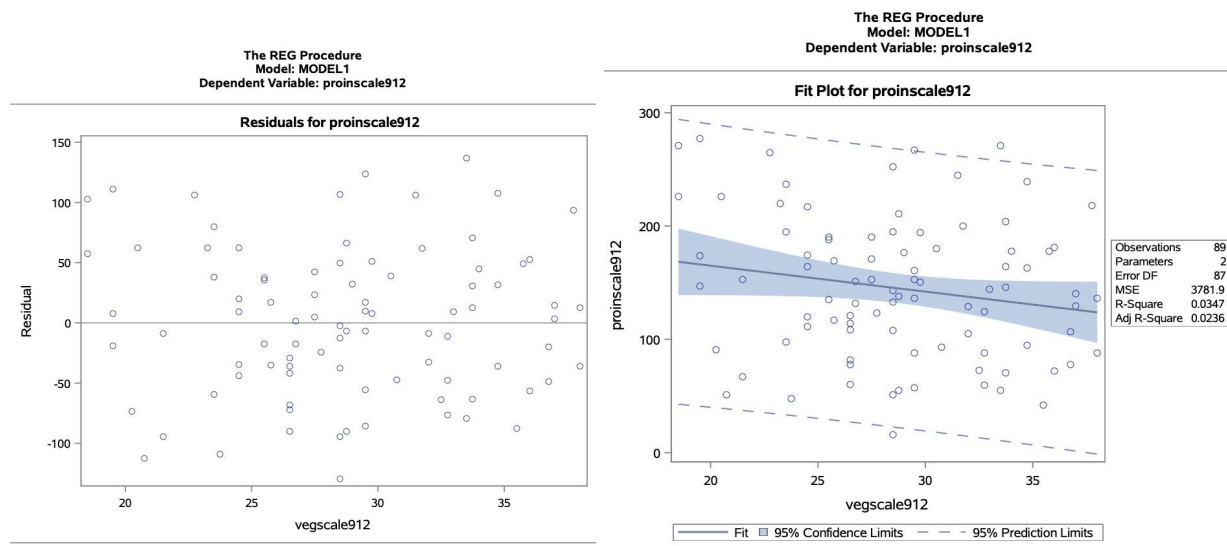
Fit Diagnostics for proinscale34





Regression of the proinflammatory scale and the vegetable scale at T9T12



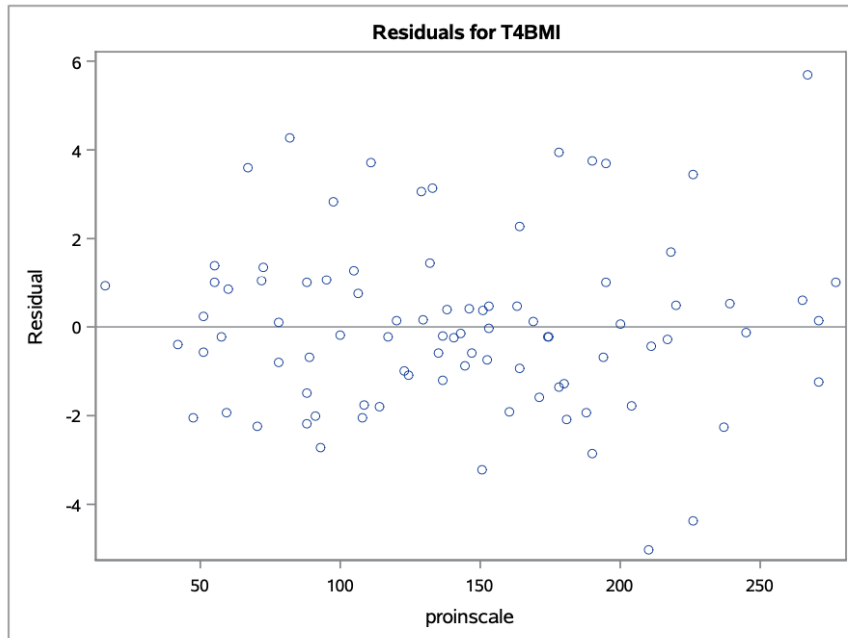


Through the use of regression and correlation, we can test to see if there is association between the food scale and inflammation. The data from four time intervals was collected and used to test the relation between the proinflammatory scale and food scale. Individually, at time 12, the residuals don't seem to have any significant pattern which indicates normality. In addition, the value from the t-test is smaller which leads to a larger p-value, so I fail to reject the null- there is no correlation between the variables. The value of R-squared supports this claim- a small value of R-squared indicates negligible correlation. Pearson's correlation leads to a negative value, so there may be a minimal negative relationship between the variables under study. The R-squared is not at an exact value of 0, so again the results support this conclusion.

Question 3

To better understand the correlation between BMI and proinflammatory scale, we first performed a simple regression accounting for time point. For a regression between BMI and proinflammatory scale at T4, there were 101 observations utilized. The adjusted R^2 value was -.0001, showing that almost none of the variation in proinflammatory markers at T4 could be explained by BMI at T4. For a regression between BMI and proinflammatory markers at T12, there were 87 observations utilized. The adjusted R^2 value was .0855, showing that very little of the variation in proinflammatory markers at T12 could be explained by BMI at T12.

Next, I moved on to a regression analysis of BMI at T4 being a predictor for the proinflammatory scale value at T12. Here 89 observations were utilized. The adjusted R^2 value was .0593, meaning that less of the proinflammatory scale value was explained by the T4 BMI than the T12 BMI. Further, the residuals of the regression model are large (some as large as 6 and -4). Together, these likely mean that BMI at T4 is not an accurate predictor of T12 proinflammatory scale.



V. Conclusion

Through the above analysis, we conclude answers to the three questions.

For question 1, there are more children in the median level of vegetable scale at T3 and T9 respectively, which means the total food intake for kids is most likely between 20 and 30. And after a year's visit, the number of children with low-level food scales is increasing so that there are more children in low-level status than the first visit. This means children may have a preference for eating unhealthy food. At the same time, even though the regression analysis shows eating unhealthy food may lead to an increase in BMI, there is not much correlation between BMI and food intake.

For question 2, when values of the proinflammatory scale at T3T4 show a higher inflammation level in kids' bodies, values of the scale at T9T12 imply the inflammation goes down after a year. The regression analysis shows some negative relationship between proinflammatory scales and food scales at T3T4 and T9T12 respectively. This means while a kid eats unhealthy food showing a low-level food scale, he may have a high-level inflammation in his body at the same time.

For question 3, the regression between BMI at T4 and proinflammatory scale at T12 shows little correlation between each other since the adjusted r^2 value is too small. It looks like healthy food intake will not influence the inflammation condition inside bodies over time. All in all, although all correlations above are weak, the analysis certainly implies that if a kid has an unhealthy diet, he may be overweight with a high BMI and get inflammation in his body.

Code

The following is the code for all three questions on the project statement. In order to run this code, the only thing you need to do is to change the library name “STA198sp” into your library name on your computer.

*Before we start, we need to create an overall data set for all three questions that contains all variables required, vegscale, and proinscale:

*Step1: create new datasets by selecting hkqs, vegscale, BMI, proinscale at T3/T4 and T9/T12 Separately:

```
data STA198sp.T3T4set;
set STA198sp.hksurveyt3t4;
*Create T3T4 vegscale:
vegscale34=t3hkq9+t3hkq13+t3hkq25+t3hkq26+t3hkq27+t3hkq29+t3hkq37+t3hkq39;
if vegscale34 le 20 then cat10q34="cat1";
if 20 lt vegscale34 le 30 then cat10q34="cat2";
if vegscale34 gt 30 then cat10q34="cat3";
*Create T3T4 proinscale:
proinscale34=rankt4TNF+rankt4IL6+rankT4CRP;
if proinscale34 le 100 then cat3bio34="low";
if 100 lt proinscale34 le 200 then cat3bio34="median";
if proinscale34 gt 300 then cat3bio34="high";
BMI = T4BMI;
time=1;
vegscale = vegscale34;
keep ID t3hkq9 t3hkq13 t3hkq25 t3hkq26 t3hkq27 t3hkq29 t3hkq37 t3hkq39 vegscale34
cat10q34
rankt4TNF rankt4IL6 rankT4CRP proinscale34 cat3bio34 T4BMI BMI time vegscale;
run;
```

```
data STA198sp.T9T12set;
set STA198sp.hksurveyt9t12;
*Create T9T12 vegscale:
vegscale912=t9hkq9+t9hkq13+t9hkq25+t9hkq26+t9hkq27+t9hkq29+t9hkq37+t9hkq39;
if vegscale912 le 20 then cat10q912="cat1";
if 20 lt vegscale912 le 30 then cat10q912="cat2";
if vegscale912 gt 30 then cat10q912="cat3";
*Create T9T12 proinscale:
proinscale912=rankt12TNF+rankt12IL6+rankT12CRP;
if proinscale912 le 100 then cat3bio912="low";
if 100 lt proinscale912 le 200 then cat3bio912="median";
if proinscale912 gt 300 then cat3bio912="high";
BMI = T12BMI;
time =2;
```

```

vegyscale = vegyscale912;
keep ID t9hkq9 t9hkq13 t9hkq25 t9hkq26 t9hkq27 t9hkq29 t9hkq37 t9hkq39 vegyscale912
cat10q912
rankt12TNF rankt12IL6 rankt12CRP proinscale912 cat3bio912 T12BMI BMI time vegyscale;
run;

```

***Step 2: Sort T3T4set and T9T12set by ID:**

```

proc sort data=STA198sp.T3T4set out=STA198sp.T3T4setsrt;
by ID;
run;

proc sort data=STA198sp.T9T12set out=STA198sp.T9T12setsrt;
by ID;
run;

```

***Merge sorted T3T4set and T9T12set by ID:**

```

data STA198sp.StackedT3T9;
merge STA198sp.T3T4setsrt STA198sp.T9T12setsrt;
by ID;
run;

```

*** Question 1:**

***frequency, both:**

```

proc freq data= STA198sp.T3T4set;
table cat10q34;

proc freq data= STA198sp.T9T12set;
table cat10q912;

```

***descriptives, mean, std...:**

```

proc means data= STA198sp.T3T4set n min q1 median q3 max mean std maxdec=3;
var t3hkq9 t3hkq13 t3hkq25 t3hkq26 t3hkq27 t3hkq29 t3hkq37 t3hkq39 vegyscale;
class cat10q34;
run;

proc means data= STA198sp.T9T12set n min q1 median q3 max mean std maxdec=3;
var t9hkq9 t9hkq13 t9hkq25 t9hkq26 t9hkq27 t9hkq29 t9hkq37 t9hkq39 vegyscale;
class cat10q912;
run;

```

***sorting:**

```
proc sort data= STA198sp.T3T4set out= STA198sp.T3T4setBMI;
by BMI;
run;
```

```
proc sort data=STA198sp.T9T12set out=STA198sp.T9T12setBMI;
by BMI;
run;
```

***box plots(replaceable variables):**

```
proc boxplot data=STA198sp.T3T4setBMI;
plot vegscale*BMI;
run;
```

```
proc boxplot data=STA198sp.T9T12setBMI;
plot vegscale*BMI;
run;
```

***regression and correlation (replaceable variables):**

```
proc reg data=STA198sp.StackedT3T9;
model BMI = vegscale;
run;
```

```
proc corr data=STA198sp.StackedT3T9 pearson spearman kendall;
var T12BMI vegscale34;
run;
```

Question 2:**frequency:**

```
proc freq data=STA198sp.StackedT3T9;
table cat3bio34;
run;
```

```
proc freq data=STA198sp.StackedT3T9;
table cat3bio912;
run;
```

***descriptives, mean, std...:**

```
proc means data=STA198sp.StackedT3T9 n min q1 median q3 max mean std maxdec=3;
var proinscale34 T4BMI;
class cat3bio34;
run;
```

```
proc means data=STA198sp.StackedT3T9 n min q1 median q3 max mean std maxdec=3;
var proinscale912 T12BMI;
class cat3bio912;
run;
```

***sorting by categories to draw box plots:**

```
proc sort data=STA198sp.StackedT3T9 out=sta198sp.proinscaleT3srt;
by cat3bio34;
run;
```

```
proc sort data=STA198sp.StackedT3T9 out=sta198sp.proinscaleT9srt;
by cat3bio912;
run;
```

***box plots:**

```
proc boxplot data=sta198sp.proinscaleT3srt;
plot proinscale34*cat3bio34;
run;
```

```
proc boxplot data=sta198sp.proinscaleT9srt;
plot proinscale912*cat3bio912;
run;
```

***correlation and regression between BMI and Proinflammatory scale for T3 and T9 separately:**

```
proc corr data = STA198sp.StackedT3T9;
var proinscale34 T4BMI;
run;
```

```
proc corr data = STA198sp.StackedT3T9;
var proinscale912 T12BMI;
run;
```

```
proc reg data=STA198sp.StackedT3T9;
model T4BMI=proinscale34;
run;
```

```
proc reg data=STA198sp.StackedT3T9;
model T12BMI=proinscale912;
run;
```

***correlation and regression between proinscale and vegscale for T3 and T9 separately:**

```
proc corr data = STA198sp.StackedT3T9;
var proinscale34 vegscale34;
run;
```

```
proc corr data = STA198sp.StackedT3T9;
var proinscale912 vegscale912;
run;
```

```
proc reg data=STA198sp.StackedT3T9;
model proinscale34=vegscale34;
run;
```

```
proc reg data=STA198sp.StackedT3T9;
model proinscale912=vegscale912;
run;
```

***Question 3:**

***Utilized data set with BMI at T4 and T12 and proinflammatory scale at each time point:**

***Time Point regression:**

```
proc reg data = STA198sp.StackedT3T9;
model T4BMI = proinscale34;
run;
```

```
proc reg data = STA198sp.StackedT3T9;
model T12BMI = proinscale912;
run;
```

***Regression between T4BMI and Proinflammatory scale:**

```
proc reg data = STA198sp.StackedT3T9;
model T4BMI = proinscale912 ;
run;
```