Spring 5-2016

Lead Poisoning in United States Children

Zeren Zhou
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Mathematics

Lead Poisoning in United States Children

by

Zeren Zhou

A thesis presented to the
Graduate School of Arts & Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Master of Arts

May 2016
St. Louis, Missouri

Table of Contents

ACKNOWLEDGEMENT

I would first like to thank my thesis advisor Professor Edward Spitznagel of the Mathematic Department at Washington University in St.Louis. Prof. Spitznagel was always willing to answer my question whenever I ran into a trouble spot or had a question about my research or writing.

I would also like to acknowledge Professor Jimin Ding of the [of the Mathematic Department at Washington University in St.Louis as the second reader of this thesis, and I am gratefully indebted to her for her very valuable comments on this thesis.

I would also like to thank Centers for Disease Control and Prevention for providing such good data set, without it, this thesis could never be done.

Finally, I would like to thank to my parent for supporting my study.

Zeren Zhou

Washington University in St.Louis

April, 2016

ABSTRACT OF THE THESIS

Lead Poisoning in United States Children
by
Zeren Zhou

Master of Arts in Statistics

Washington University in St. Louis, 2016

Professor Edward Spitznagel, Chair

Professor Jimin Ding, Co-Chair

Objectives. The goals of this study were explore predictors of lead in childrens' blood and to develop a multivariate model using as many predictors as possible.

Methods. The logistic regression and linear regression were used, The analysis is conducted using SAS survey regression procedures that account for weighting, stratification, and clustering of the data.

Conclusions. The age, the family income and the race could be consider as predictors for blood lead, by using those three as predictors, we could develop a relatively good multivariate model for predicting blood lead of childrens.

CHAPTER I: Introduction

Although Lead has been widely used in human live, such as the production of gasoline, ceramic products, paints, batteries. Lead is a neurotoxic metal that affects areas of the brain that regulate behavior and nerve cell development. Children are more vulnerable to the effects of lead because the blood-brain barrier is not fully developed in them Lead-poisoning is one of the most common environmental health problems affecting young children. It can greatly affect children brain development, cause IQ decrease, learning problem. Hence it is important to develop a method to predict whether the children is blood lead at risk.

The goal of this thesis is to explore predictors for blood lead in childrens' blood and to develop a multivariate model as prognostic function that using as many predictors as possible.

Chapter II will discuss data and methodology I use to develop multivariate model. In particular, the methodology I use to get good predictor for blood lead.

Chapter III will discuss the process of getting good predictor by methodology in Chapter II.

Chapter IV will discuss how I get multivariate model for predicting whether children is blood lead at risk.

CHAPTER II: Data information and method

2.1 Data information

2.1.1 General information of data

The data is gather from The National Health and Nutrition Examination Survey (NHANES), NHANES is a program of studies designed to assess the health and nutritional status of adults and children in the United States. The survey is unique in that it combines interviews and physical examinations. NHANES is a major program of the National Center for Health Statistics (NCHS). NCHS is part of the Centers for Disease Control and Prevention (CDC) and has the responsibility for producing vital and health statistics for the Nation. NHANES has gather blood metal that is dangerous to human body (such as lead, cadmium, mercury) of adults and children in the United States 2007-2008, 2009-2010, 2011-2012, 2013-2014.

By combining those four data we obtain data set of the blood metal of adults and children in the U.S from 2007 to 2014, by deleting the data of blood metal of adults, and deleting the data of blood metal other than lead, we could the get the data set of blood lead of children at age 1-5 in the U.S from 2007 to 2014. The data set contains 4927 observations. Different observation correspond to different person.

The data set has many variables, only a few variables is meaningful variables for predict lead poisoning. Following variables are important:

LBXBPB: variable of blood lead, units as ug/dl.

RIAGENDR: variable of gender.

RIDAGEYR: variable of age in years at screening.

RIDRETH1: variable of race, there are 5 kind of races in NHANES, code as 1 to 5, 1 is for Mexican American, 2 is for other Hispanic, 3 is for Non-Hispanic White, 4 is for Non-Hispanic black and 5 is for other race (including multi-racial).

INDFMIN2: variable of annual family income, reported as a range value in dollars.

DMDHHSIZ: variable of total number of people in the household.

2.1.2 cluster, strata, and weights of data

The data of NHANES is collect as following, the collector would go to some areas and randomly talk to people to collect blood metal information and some personal information. Since blood metal is affected by the situation of areas, such as the condition fo water, local food supply, the data that NHANES obtain would have cluster, weights and strata.

The NHANES have already put some variables in the data set to deal with cluster, weights and strata of data that caused by survey design of data. The variables ***sdmvstra*** and ***sdmvpsu*** is design variables and could be used in calculating variance estimate. The variable ***wtmec2yr*** is the 2-year sample weights.

2.1.3 Recode income

Note that for variable of annual family income ***INDFMIN2***, since it is reported as a range value in dollars, and it has extremely values (77 and 99) caused by people refused to tell income in the survey. I add a new variable ***familyincome*** to reflect income of family, the transformation between ***INDFMIN2*** and ***familyincome***

$$\boldsymbol{familyincome} = (\boldsymbol{INDFMIN2} - \mathbf{1}) \ \boldsymbol{*2.5} + \mathbf{5}$$

if INDFMIN2 not equal to 77 or 99.

The $\boldsymbol{familyincome}$ variable is united as thousand dollar.

2.2 Methodology

2.2.1 How to select predictor

To find out predictor for lead poisoning, one way is to use correlation coefficients to see which variable has strong correlation with blood lead variable ***LBXBPB***, since the data set is cluster, strata, and weights, it is inaccurate to correlation coefficients directly. Hence we use logistic regression and linear regression for simple model to deal with cluster, strata, and weights. In the context of SAS procedures, we use SAS survey logistic regression and survey regression that account for clustering, weighting and stratification of data. That is, for variable ***a***, if the model

$$\mathrm{LBXBPB} = \beta_0 + \beta_1 * a + \varepsilon$$

has a significant non-zero β_1, then it is safety to conclude that variable ***a*** has a strong correlation with blood lead variable ***LBXBPB***.

Of course, since the goal of this thesis is to explore the predictors of lead poisoning in children, hence it seems more meaningful to use logistic regression to check correlation of certain variable and blood lead. That is, for variable ***a***, if the model

$$Logit(p) = \beta_0 + \beta_1 * a$$

has a significant non-zero β_1, then it is safety to conclude that variable ***a***, then it is safety to conclude that variable ***a*** has a strong correlation with blood lead variable ***LBXBPB***. Those two methods do not have much difference, we use both two find out predictors.

2.2.2 How to develop a multivariate model

After we get we find out what are relatively good predictors for blood lead of children (code as a1, a2, a3 and so on), we then could use multivariate linear regression or multivariate logistic regression to develop a multivariate model. Since the goal of this thesis is to develop predictors that could predict whether the child is lead concern or not, hence we would use multivariate logistic regression to develop model. That is, develop such model

$$Logit(p) = \beta_0 + \beta_1 * \boldsymbol{a1} + \beta_2 * \boldsymbol{a2} + \cdots$$

where a1, a2 are the predictors that we confirm through simple logistic regression and simple linear regression.

CHAPTER III: Find predictor of blood lead

3.1 Find predictor through logistic regression

3.1.1 Determine blood lead level as blood lead at risk

We use simple logistic regression to find out what is a relatively good indicator for lead poisoning. Then the first thing to do is to determine what kind of blood lead level could be considered as blood lead at risk. Many studies suggest blood lead level greater or equal to 10 ug/dL is extremely dangerous to children aged 1 to 5, and there are more and more scientific evidence of adverse effects of blood lead levels below 10 μg/dL in children, for instance, it can result in decreased IQ, academic failure and behavior problems blood lead level greater than 5 ug/dL is considered be dangerous to children, hence we set blood lead level equal or greater than 4.5 ug/dL is blood lead at risk.

In this way, I define p is the probability that blood lead of children aged 1 to 5 is greater or equal to 4.5 ug/dL, that is

$$p = \mathrm{pr}(\mathrm{LBXBPB} \geq 4.5)$$

3.1.2 Age variable

First consider age variable, that is consider model

$$Logit(p) = \beta_0 + \beta_1 * RIDAGEYR$$

where logit() is logit function, $\log it(x) = ln(\frac{x}{1-x})$.

We can get the result as following

Testing Global Null Hypothesis: BETA=0					
Test		Chi-Square	DF		Pr > ChiSq
Wald		4.2284	1		0.0398
Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	3.6774	0.2599	200.1818	<.0001
RIDAGEYR	1	0.1811	0.0881	4.2284	0.0398

As we can see, the P-value is 0.0398, hence I conclude the model test is significant. Hence the age variable RIDAGEYR could be consider as a good indicator for lead poisoning.

3.1.3 Family income variable

We now consider whether family income is a indicator for lead poisoning, that is consider model

$$Logit(p) = \beta_0 + \beta_1 * familyincome$$

where logit() is logit function, $\log it(x) = ln(\frac{x}{1-x})$.

We can get the result as following:

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Wald	7.0292	1	0.0080

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	2.7858	0.4143	45.2067	<.0001
familyincome	1	0.071	0.000027	7.0292	0.0080

as the result shows, the P-value is 0.008, hence it is safely to conclude the model test is significant. Hence the family income variable could be consider as an indicator for lead poisoning.

3.1.4 Race variable

Right now we consider whether the race could be consider as indicator for lead poisoning. We use dummy variable to code different race.

if RIDRETH1=1 then set Race1=1; Race2=0; Race3=0; Race4=0;

if RIDRETH1=2 then set; Race1=0; Race2=1; Race3=0; Race4=0;

if RIDRETH1=3 then set; Race1=0; Race2=0; Race3=1; Race4=0;

if RIDRETH1=5 then set; Race1=0; Race2=0; Race3=0; Race4=1;

if RIDRETH1=4 then set; Race1=0; Race2=0; Race3=0; Race4=0;

This is we set Non-Hispanic black as standard.

Consider the model

$$Logit(p) = \beta_0 + \beta_1 * Race1 + \beta_2 * Race2 + \beta_3 * Race3 + \beta_4 * Race4$$

We can get the result as following:

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Wald	48.5161	6	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.7408	0.4932	12.4557	0.0004
RIDAGEYR	1	0.1669	0.0885	3.5617	0.0591
familyincome	1	0.0691	0.0253	7.4580	0.0063
Race1	1	1.3803	0.4335	10.1367	0.0015
Race2	1	1.5534	0.4422	12.3390	0.0004
Race3	1	0.5083	0.4286	1.4070	0.2356
Race4	1	0.7702	0.4622	2.7776	0.0956

hence it is safely to conclude the model test is significant. Hence the race could be consider as an indicator for lead poisoning.

3.1.5 Gender variable and family size variable

Consider model

$$Logit(p) = \beta_0 + \beta_1 * RIAGENDE$$

where logit() is logit function, $\log it(x) = ln(\frac{x}{1-x})$.

That is consider whether gender is an indicator for lead poisoning.

we can get result

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Wald	0.1596	1	0.6895

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	4.0665	0.3853	111.3622	<.0001
RIAGENDR	1	0.0829	0.2076	0.1596	0.6895

The P-value is 0.6895, hence the model is not significant and gender is not an indicator for lead poisoning.

For model

$$Logit(p) = \beta_0 + \beta_1 * DMDFMSIZ$$

where logit() is logit function, $\log it(x) = ln(\frac{x}{1-x})$.

That is consider whether family size could be an indicator for lead poisoning.

we can get result

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Wald	1.7405	1	0.1871

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	4.7782	0.5614	72.4528	<.0001
DMDFMSIZ	1	-0.1294	0.0981	1.7405	0.1871

The P-value is 0.1871, hence the model is not significant and family size is not an indicator for lead poisoning.

from logistic regression, that is the age variable RIDAGEYR, the family income variable Familyincome and the race variable Race1-Race4 could be consider as predictors for blood lead

3.2 Find predictor through linear regression

3.2.1 Age variable

We use simple linear regression to find out what is a relatively good indicator for lead poisoning. First consider age variable, that is consider model

$$\text{LBXBPB} = \beta_0 + \beta_1 \ * RIDAGEYR + \varepsilon$$

The result is

Tests of Model Effects			
Effect	Num DF	F Value	Pr > F
RIDAGEYR	1	28.10	<.0001

Estimated Regression Coefficients				
Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1.7502163	0.09369963	18.68	<.0001
RIDAGEYR	-.1177694	0.02221591	-5.30	<.0001

As we can see, the P-value is less than 0.0001, hence I conclude the model test is significant. Hence the age variable RIDAGEYR could be consider as a good predicator for lead poisoning.

3.2.2 Family income variable

We now consider whether family income is an indicator for lead poisoning, that is consider model

$$\text{LBXBPB} = \beta_0 + \beta_1 \ * \ + \varepsilon$$

We can get the result as following:

Tests of Model Effects			
Effect	Num DF	F Value	Pr > F
familyincome	1	30.07	<.0001

Estimated Regression Coefficients				
Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1.9226476	0.10949853	17.56	<.0001
familyincome	-0.0235	0.00000428	-5.48	<.0001

as the result shows, the P-value is less than 0.0001, hence it is safely to conclude the model test is significant. Hence the family income variable could be considered as an indicator for lead poisoning.

3.2.3 Race variable

we consider whether the race could be consider as indicator for lead poisoning. We use dummy variable to same as in logistic regression to code different race.

consider the model

$$\mathrm{LBXBPB} = \beta_0 + \beta_1 * Race1 + \beta_2 * Race2 + \beta_3 * Race3 + \beta_4 * Race4 + \varepsilon$$

We can get the result as following:

Tests of Model Effects			
Effect	Num DF	F Value	Pr > F
Model	4	12.76	<.0001
Intercept	1	472.38	<.0001
Race1	1	42.82	<.0001
Race2	1	31.53	<.0001
Race3	1	15.96	0.0002
Race4	1	24.19	<.0001

Estimated Regression Coefficients				
Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1.8454244	0.08490849	21.73	<.0001
Race1	-0.6011248	0.09186820	-6.54	<.0001
Race2	-0.5769378	0.10274790	-5.62	<.0001
Race3	-0.5008340	0.12538240	-3.99	0.0002
Race4	-0.5544454	0.11273429	-4.92	<.0001

hence it is safely to conclude the model test is significant. Hence the race could be consider as an indicator for lead poisoning.

3.2.4 Gender variable and family size variable

Consider model

$$\mathrm{LBXBPB} = \beta_0 + \beta_1 \ * RIAGENDR + \varepsilon$$

That is consider whether gender is an indicator for lead poisoning.

we can get result

Tests of Model Effects			
Effect	Num DF	F Value	Pr > F
RIAGENDR	1	2.04	0.1584

Estimated Regression Coefficients				
Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1.5156218	0.10415431	14.55	<.0001
RIAGENDR	-.0851080	0.05963961	-1.43	0.1584

The P-value is 0.1584, hence the model is not significant and gender is not an indicator for lead poisoning.

For model

$$\mathrm{LBXBPB} = \beta_0 + \beta_1 * DMDFMSIZ + \varepsilon$$

That is consider whether family size could be an indicator for lead poisoning.

we can get result

Tests of Model Effects			
Effect	Num DF	F Value	Pr > F
DMDFMSIZ	1	1.68	0.1995

Estimated Regression Coefficients				
Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1.25998141	0.12654422	9.96	<.0001
DMDFMSIZ	0.02844896	0.02194684	1.30	0.1995

The P-value is 0.1995, hence the model is not significant and family size is not an indicator for lead poisoning.

The result we get from simple linear regression is the same as the result we get from logistic regression, that is the age variable RIDAGEYR, the family income variable

Familyincome and the race variable Race1-Race4 could be consider as predictors for blood lead

CHAPTER IV: Results of multivariate model

From above, we can see the age variable RIDAGEYR, the family income variable Familyincome and the race variable Race1-Race4 could be consider as indicator for children lead poison. Now consider the full model to see whether those variable are truly good indicators.

Consider model

$$Logit(p) = \beta_0 + \beta_1 * Race1 + \beta_2 * Race2 + \beta_3 * Race3 + \beta_4 * Race4 + \beta_5 * RIDAGEYR + \beta_6 * familyincome$$

We could get the result as following:

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Wald	48.5161	6	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.7408	0.4932	12.4557	0.0004
RIDAGEYR	1	0.1669	0.0885	3.5617	0.0591
familyincome	1	0.0691	0.0253	7.4580	0.0063
Race1	1	1.3803	0.4335	10.1367	0.0015
Race2	1	1.5534	0.4422	12.3390	0.0004
Race3	1	0.5083	0.4286	1.4070	0.2356
Race4	1	0.7702	0.4622	2.7776	0.0956

As we can see, all of those variables are good indicator for lead poison. And the predicted model for lead poisoning child is

$$Logit(p) = 2.5110 + 1.3803 * Race1 + 1.5534 * Race2 + 0.5083 * Race3$$
$$+ 0.7702 * Race4 + 0.1669 * RIDAGEYR + 0.0691 * familyincome$$

As the following table shows

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	69.2	Somers' D	0.430
Percent Discordant	26.2	Gamma	0.451
Percent Tied	4.6	Tau-a	0.017
Pairs	406640	c	0.715

The coefficient of concordance is 0.715, which is relatively good, hence the model

$$Logit(p) = 2.5110 + 0.6101(Race1) + 0.7831(Race2) - 0.2619(Race3) - 0.7702(Race4)$$
$$+ 0.1669(RIDAGEYR) + 0.000069(familyincome)$$

is a good predicted model for blood lead at risk.

REFERENCE

Min-ming Li , Jia Cao et al. **The trend of lead poisoning rate in Chinese population aged 0–18 years old: a meta-analysis**. *BMC Public Health*. 2015; **15**:756

Daniel R. Berg, Erin T. Eckstein et al. **Childhood lead poisoning prevention through prenatal housing inspection**. *Am J Obstet Gynecol*. 2012 Mar;206(3):199.e1-4

Kaufmann, R.B., Clouse, T.L., Olson, D.R., and Matte, T.D. **Elevated blood lead levels and blood lead screening among US children aged one to five years: 1988-1994.** *Pediatrics*. 2000; 106: E79

Pat McLaine, Rebecca Lee et al. Elevated Blood Lead Levels and Reading Readiness at the Start of Kindergarten. Pediatrics. 2013 Jun;131(6):1081-9.

Pocock, S.J., Smith, M., Baghurst, P. **Environmental lead and children's intelligence: a systematic review of the epidemiological evidence.** *BMJ*. 1994;309:1189–1197.

Jaime Raymond, Mary Jean Brown, CDC. Childhood Blood Lead Levels — United States, 2007–2012. *MMWR*. October 23, 2015 / 62(54);76-80.

Kaplowitz, S., Perlstadt, H., Post, L. **Comparing lead poisoning risk assessment methods: census block group characteristics vs zip codes as predictors.** *Public Health Rep*. 2010;125:234–245.

Ronald P.Cody, Jeffery K.Smith. **Applied Statistics and SAS Programming Language, Fifth edition**.

E.L.Lehmann, Joseph P. Romano. **Testing Statistical Hypotheses. Third edition**.

APPENDIX

SAS code

```
Data allData;
set Lead_E Lead_F Lead_G Lead_H;
IF LBXBPB ge 4.5 then do; leadPoison=1;end;
else  do; leadPoison=0;end;
run;

Data child;
set allData;
IF ridageyr ge 1 and ridageyr le 5;
run;

proc surveyreg data=child;/*reg for age*/
 model LBXBPB=ridageyr;
 weight  wtmec2yr;
 cluster sdmvpsu;
 strata  sdmvstra;
 output out=regchild;
 run;

 proc gchart data=child;
vbar LBXBPB/group=RIAGENDR;
run;

 proc surveyreg data=child;/*reg for gender*/
 model LBXBPB=RIAGENDR;
 weight  wtmec2yr;
```

```
cluster sdmvpsu;
strata  sdmvstra;
run;

Data child3;/*dummy coding for race*/
set child;
if RIDRETH1=1 then do; Race1=1;Race2=0;Race3=0;Race4=0;end;
if RIDRETH1=2 then do; Race1=0;Race2=1;Race3=0;Race4=0;end;
if RIDRETH1=3 then do; Race1=0;Race2=0;Race3=1;Race4=0;end;
if RIDRETH1=5 then do; Race1=0;Race2=0;Race3=0;Race4=1;end;
if RIDRETH1=4 then do; Race1=0;Race2=0;Race3=0;Race4=0;end;
run;

proc surveyreg data=child3;/*reg for race*/
model LBXBPB=Race1 Race2 Race3 Race4;
weight  wtmec2yr;
cluster sdmvpsu;
strata  sdmvstra;
run;

 proc surveyreg data=child;/*reg for familysize*/
model LBXBPB=dmdfmsiz;
weight  wtmec2yr;
cluster sdmvpsu;
strata  sdmvstra;
run;

  Data child4;/*reset income*/
set child3;
if indfmin2=77 or indfmin2=99 then do; familyincome=.;end;
```

```
 else do; familyincome=(indfmin2-1)*2.5+5;end;
 logincome=LOG(familyincome);
 run;

proc surveyreg data=child4;/*reg for income*/
 model LBXBPB=familyincome;
 weight  wtmec2yr;
 cluster sdmvpsu;
 strata  sdmvstra;
 run;

  proc surveyreg data=child4;/*reg for log-income*/
 model LBXBPB=logincome;
 weight  wtmec2yr;
 cluster sdmvpsu;
 strata  sdmvstra;
 run;

proc surveyreg data=child4;
 model LBXBPB=ridageyr familyincome;
 weight  wtmec2yr;
 cluster sdmvpsu;
 strata  sdmvstra;
 run;

proc surveylogistic data=child;/*logistic regression for age*/
 model leadPoison=ridageyr;
   weight  wtmec2yr;
 cluster sdmvpsu;
 strata  sdmvstra;
```

```
output out=regchild;
run;

proc surveylogistic data=child;/*logistic regression for gender*/
model leadPoison=RIAGENDR;
weight  wtmec2yr;
cluster sdmvpsu;
strata  sdmvstra;
output out=regchild;
run;

proc surveylogistic data=child3;/*logistic regression for race*/
model leadPoison=Race1 Race2 Race3 Race4;
weight  wtmec2yr;
cluster sdmvpsu;
strata  sdmvstra;
run;

  proc surveylogistic data=child;/*logistic regression for familysize*/
model leadPoison=dmdfmsiz;
weight  wtmec2yr;
cluster sdmvpsu;
strata  sdmvstra;
run;

 proc surveylogistic data=child4;/*logistics reg for income*/
model leadPoison=familyincome;
weight  wtmec2yr;
cluster sdmvpsu;
strata  sdmvstra;
```

```
run;

  proc surveylogistic data=child4;/*logistics reg for income*/
model leadPoison=logincome;
weight  wtmec2yr;
cluster sdmvpsu;
strata  sdmvstra;
run;

proc surveylogistic data=child4;/*logistics reg for all factor*/
model leadPoison= ridageyr familyincome race1 race2 race3 race4;
weight  wtmec2yr;
cluster sdmvpsu;
strata  sdmvstra;
run;

proc surveylogistic data=child4;/*logistics reg for all factor*/
model leadPoison= ridageyr logincome race1 race2 race3 race4;
weight  wtmec2yr;
cluster sdmvpsu;
strata  sdmvstra;
run;
```