CHILDREN'S HEALTHY LIVING (CHL)

CENTER OF EXCELLENCE

SAS Programming Guide

Developed by the CHL Data Work Group
for use in the CHL Pacific Region

For further information, please contact:
CHL Data Center
Children's Healthy Living Center of Excellence
Ag Sciences 302, 1955 East West Road, Honolulu, HI 96822
Email: chldata@hawaii.edu
Website: www.chl-pacific.org

This project is supported by the Agriculture and Food Research Initiative Grant no. 2021-68012-35899 and Grant no. 2018-69001-27551 from the USDA National Institute of Food and Agricultural Science Enhancement Coordinated Agricultural Program and Grant no. 1P20GM139753-01A1 from the Centers of Biomedical Research Excellence (COBRE).

United States Department of Agriculture
National Institute of Food and Agriculture

REVISED 10.14.2024

Contents

Background

SAS is an integrated software suite for advanced analytics, business intelligence, data management, and predictive analytics. This guide provides basic SAS codes to conduct various data procedures and statistical analyses.

Importing Datasets into SAS

***Note**: Highlighted shows what you must fill in.

Importing Permanent SAS Dataset Files

```
libname FolderName "C:\Users\Folder Location";

data DatasetName;
      set FolderName.FileName;
run;
```

Note: *FolderName* & *DatasetName* can be any name you will use in your SAS program. However, *FileName* must be the name of the SAS dataset file you are importing from the folder (copy this file name from your folder location & paste into SAS code).

Example:

```
libname chl "C:\Desktop\User\CHL";

data chldata;
      set chl.chl_final;
run;
```

Importing CSV Files

```
proc import datafile= "C:\Users\Folder Location\File Name.csv"
        out=DatasetName
        dbms=csv
        replace;
        getnames=yes;
        guessingrows=max;
run;
```

Importing Excel Files

```
proc import datafile= "C:\Users\Folder Location\File Name.xlsx"
        out=DatasetName
        dbms=xlsx
        replace;
        sheet="Sheet1";
        getnames=yes;
        guessingrows=max;
run;
```

Note: *DatasetName* can be any name that you will use in your SAS program.

Dataset Contents

Check the contents of your imported dataset with the codes below. Contents will include number of observations, names of variables, type of your variables (character vs. numeric), and label of variables (describes what the variable is).

```
proc contents data=DatasetName; run;
```

Example:

```
proc contents data=chldata; run;
```

Output:

The SAS System

The CONTENTS Procedure

Data Set Name	WORK.CHLDATA	Observations	5499
Member Type	DATA	Variables	32
Engine	V9	Indexes	0
Created	08/27/2024 15:43:23	Observation Length	296
Last Modified	08/27/2024 15:43:23	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	WINDOWS_64		
Encoding	wlatin1 Western (Windows)		

Engine/Host Dependent Information	
Data Set Page Size	65536
Number of Data Set Pages	25
First Data Page	1
Max Obs per Page	221
Obs in First Data Page	196
Number of Data Set Repairs	0
ExtendObsCounter	YES
Filename	C:\Users\RICADE~1\AppData\Local\Temp\SAS Temporary Files_TD7452_NAP1254_\chldata.sas7bdat
Release Created	9.0401M7
Host Created	X64_10PRO
Owner Name	NAP1254\Rica Dela Cruz
File Size	2MB
File Size (bytes)	1703936

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat	Label
1	agegrp	Char	8			ANTHRO: 2-category age group at anthropometry measures (1=2-5 yrs, 2=6-8 yrs)
16	agemos	Num	8			ANTHRO: Computed participant age in months when anthropometry was measured
2	bmi_category	Char	20	$14.		ANTHRO: BMI in 4 categories using CDC 2000 reference
21	carbohydrate_adjForVar_WTMN	Num	8			FOODLOG Intake for weighted mean across days: Carbohydrate in g, adjusted for within person variance
5	community	Num	8			GENERAL: Community
32	community_cluster	Num	8			GENERAL: Community for cluster analysis
22	dietaryfiber_adjForVar_WTMN	Num	8			FOODLOG Intake for weighted mean across days: Dietary Fiber in g, adjusted for within person variance
18	education	Num	8			DEMO: Highest grade of school respondent completed (1=Never attended, 2=Grades 1-8, 3=Grades 9-11, 4=HS, 5=Some college, 6= 4 year college)
11	hei1_adjForVar_WTMN	Num	8			FOODLOG Intake for day: Whole Grains in oz, weighted for weekday/weekend days and adjusted for within person variance
12	hei2_adjForVar_WTMN	Num	8			FOODLOG Intake for day: Vegetables in cups, weighted for weekday/weekend days and adjusted for within person variance
13	hei3_adjForVar_WTMN	Num	8			FOODLOG Intake for day: Fruit in cups, weighted for weekday/weekend days and adjusted for within person variance
14	hei4_adjForVar_WTMN	Num	8			FOODLOG Intake for day: Milk in cups, weighted for weekday/weekend days and adjusted for within person variance
15	hei5_adjForVar_WTMN	Num	8			FOODLOG Intake for day: Meat in oz, weighted for weekday/weekend days and adjusted for within person variance

Data Steps

Data steps revise imported datasets to create new variables, remove specific data observations, remove variables from the imported dataset, change the format of variables (e.g. number to character format).

"Proc Print" statements allow you to view your dataset line by line. You can view the whole

Creating New Variables

```
data NewDatasetName;
      set DatasetName;

if Variable1 = "Original Category 1" then Variable1_New = "New Category Name 1";
else if Variable1 = "Original Category 2" then Variable1_New = "New Category Name 2";

if Variable2 = # then Variable2_New = New#;
else if Variable2 = # then Variable2_New = New#;

if # =< Variable3 =< # then NewVariable3 = "Description";
else if # =< Variable3 =< # then NewVariable3 = "Description";

run;

Proc Print data= NewDatasetName (obs=# observations);
Var id Variable1 NewVariable1 Variable2 NewVariable2;
run;
```

Note: `(obs= )` is an optional statement. It outputs a list of the number of observations specified. If not included, all observations in the dataset will be listed.

Example:

```
data chldata_new;
      set chldata;

if bmi_category = "Underweight" then bmi_category_new = "Underweight";
else if bmi_category = "Healthy weight" then bmi_category_new = "Healthy";
else if bmi_category = "Overweight" then bmi_category_new = "OWOB";
else if bmi_category = "Obese" then bmi_category_new = "OWOB";

if bmi_category = "Healthy weight" then owob = 0;
else if bmi_category = "Overweight" then bmi_category_new = 1;
else if bmi_category = "Obese" then bmi_category_new = 1;

if sex = 1 then sex_new = "Male";
else if sex = 2 then sex_new = "Female";

if 24 =< agemos < 72 then agegroup = "2-5 years old";
else if 72 =< agemos < 108 then agegroup = "6-8 years old ";

run;

Proc Print data=chldata_new (obs=10);
Var id bmi_category bmi_category_new owob sex sex_new agemos agegroup;
run;
```

"Proc Print" Output:

Obs	id	bmi_category	bmi_category_new	owob	sex	sex_new	agemos	agegroup
1	111001	Healthy weight	Healthy	0	2	Fema	53.2961	2-5 years old
2	111002	Healthy weight	Healthy	0	1	Male	95.4934	6-8 years old
3	111003	Healthy weight	Healthy	0	2	Fema	83.6908	6-8 years old
4	111004	Healthy weight	Healthy	0	1	Male	46.0658	2-5 years old
5	111005	Healthy weight	Healthy	0	1	Male	57.5263	2-5 years old
6	111006	Healthy weight	Healthy	0	2	Fema	71.2632	2-5 years old
7	111009	Healthy weight	Healthy	0	2	Fema	72.1645	6-8 years old
8	111010	Healthy weight	Healthy	0	1	Male	57.1645	2-5 years old
9	111011	Healthy weight	Healthy	0	1	Male	72.3947	6-8 years old
10	111012	Healthy weight	Healthy	0	1	Male	54.7566	2-5 years old

Other Dataset Procedures

Sorting Data

```
proc Sort data=DatasetName;
      by Variable;
run;
```

Merging Datasets

```
data MergedDatasetName;
      Merge DatasetName1 DatasetName2;
      by Variable;
run;
```

Note: Datasets must first be sorted by the variable you are merging by.

Removing duplicate observations

Add `nodupkey` to your code to remove duplicate values of specific variables.

```
proc sort data=DatasetName nodupkey out=NewDatasetName;
      by Variable;
run;
```

Frequency Procedures

Frequency Procedures or “Proc Freq” provide frequencies, i.e. counts and percent of data. This procedure is used with categorical variables, or variables that include categories or groups (e.g. sex, which includes categories of male and female).

Simple Frequencies

```
proc freq data=chldata;
table bmi_category;
run;
```

Note: You can list out as many variable names after `table`, which will output separate tables for each variable.

Output:

The SAS System

The FREQ Procedure

ANTHRO: BMI in 4 categories using CDC 2000 reference				
bmi_category	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Healthy weight	3739	68.91	3739	68.91
Obese	774	14.26	4513	83.17
Overweight	776	14.30	5289	97.48
Underweight	137	2.52	5426	100.00
Frequency Missing = 73				

Simple Cross Tabulation

Cross tabulations outputs a cross table of 2 variables of interest. The same **`proc freq`** statement is used, but includes a * between the 2 variables of interest.

```
proc freq data=chldata;
table worldbank*income2;
title "Income by World Bank for table 1";
run;
```

Note: Adding a `title` statement will include a title above your output table.

Output:

Income by World Bank for table 1

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of worldbank by income2		
worldbank(World Income Bank Level: Low, Middle, High)	income2(DEMO: Average annual houshold income in two categories (1= < $35K, 2= $35K))		
	1	2	Total
high	1829 42.67 66.22 55.76	933 21.77 33.78 92.74	2762 64.44
lower middle	430 10.03 96.63 13.11	15 0.35 3.37 1.49	445 10.38
upper middle	1021 23.82 94.62 31.13	58 1.35 5.38 5.77	1079 25.17
Total	3280 76.53	1006 23.47	4286 100.00
Frequency Missing = 1213			

Complex Tabulation

Complex tabulations allow for cross tabulations of more than 2 variables.

```
proc tabulate data=chldata missing;
class sex agegrp race_new_omb jurisnum ;
table ALL (sex agegrp race_new_omb)*(N colpctn), jurisnum(ALL);
title 'Demographics Table - Sex, age group, Race/ethnicity';
run;
```

Output:

Demographics Table - Sex, age group, Race/ethnicity

		GENERAL: Jurisdiction number											All
		1	2	3	4	5	6	7	8	9	10	11	
All	N	190	199	855	903	191	188	188	214	968	938	665	5499
DEMO: Sex of the CHL subject													
1	N	102	93	445	476	106	92	104	88	508	450	350	2814
	ColPctN	53.68	46.73	52.05	52.71	55.50	48.94	55.32	41.12	52.48	47.97	52.63	51.17
2	N	88	106	410	427	85	96	84	126	460	488	315	2685
	ColPctN	46.32	53.27	47.95	47.29	44.50	51.06	44.68	58.88	47.52	52.03	47.37	48.83
ANTHRO: 2-category age group at anthropometry measures (1=2-5 yrs, 2=6-8 yrs)													
2-5 yrs	N	144	140	457	599	105	168	113	80	643	692	487	3628
	ColPctN	75.79	70.35	53.45	66.33	54.97	89.36	60.11	37.38	66.43	73.77	73.23	65.98
6-8 yrs	N	46	59	398	304	86	20	75	134	325	246	178	1871
	ColPctN	24.21	29.65	46.55	33.67	45.03	10.64	39.89	62.62	33.57	26.23	26.77	34.02
DEMO: Child race following guidelines of US Office of Management and Budget (OMB) Definition													
	N	1	.	.	2	2	.	.	.	4	4	6	19
	ColPctN	0.53	.	.	0.22	1.05	.	.	.	0.41	0.43	0.90	0.35
AIAN	N	.	.	.	.	.	.	.	.	.	2	122	124
	ColPctN	.	.	.	.	.	.	.	.	.	0.21	18.35	2.25
Asian	N	6	2	80	302	.	.	.	.	.	90	5	485
	ColPctN	3.16	1.01	9.36	33.44	.	.	.	.	.	9.59	0.75	8.82
Black	N	.	.	.	.	.	.	1	.	1	1	13	16
	ColPctN	.	.	.	.	.	.	0.53	.	0.10	0.11	1.95	0.29
More than one race	N	35	6	117	146	22	19	14	23	35	503	172	1092
	ColPctN	18.42	3.02	13.68	16.17	11.52	10.11	7.45	10.75	3.62	53.62	25.86	19.86
NHPI	N	148	191	652	450	167	169	173	191	928	270	6	3345
	ColPctN	77.89	95.98	76.26	49.83	87.43	89.89	92.02	89.25	95.87	28.78	0.90	60.83
White	N	.	.	6	3	.	.	.	.	.	68	341	418
	ColPctN	.	.	0.70	0.33	.	.	.	.	.	7.25	51.28	7.60

Means Procedures

Simple Means

Use the code below to obtain the mean of continuous variables. You can add many variables to after `var` to get means of all variables stated.

```
proc means data=chldata;
var   carbohydrate_adjForVar_WTMN dietaryfiber_adjForVar_WTMN
            protein_adjForVar_WTMN  totalfat_adjforvar_WTMN;
title   'Total averages of macronutrients';
run;
```

Output:

Total averages of macronutrients

The MEANS Procedure

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
carbohydrate_adjForVar_WTMN	FOODLOG Intake for weighted mean across days: Carbohydrate in g, adjusted for within person variance	3529	235.2379859	72.7096414	44.3717940	634.7956812
dietaryfiber_adjForVar_WTMN	FOODLOG Intake for weighted mean across days: Dietary Fiber in g, adjusted for within person variance	3529	10.0162430	4.9121329	0.3263891	44.9461291
protein_adjForVar_WTMN	FOODLOG Intake for weighted mean across days: Protein in g, adjusted for within person variance	3529	67.5018750	18.5412207	11.2731005	153.2901749
totalfat_adjForVar_WTMN	FOODLOG Intake for weighted mean across days: Total Fat in g, adjusted for within person variance	3529	62.1630656	20.1460062	6.5308032	169.0512017

Means by specific groups & adjusting for jurisdiction strata & community cluster

```
proc SURVEYmeans data=chldata mean std stderr clm;
var   carbohydrate_adjForVar_WTMN dietaryfiber_adjForVar_WTMN
           protein_adjForVar_WTMN  totalfat_adjforvar_WTMN;
title   'Total averages by agegroup';
strata  jurisnum;
cluster community_cluster;
weight  wt_anthro_adj;
domain  agegrp;
run;
```

Note: Adding to the statement `std` includes the Standard Error of the Mean, `stderr` includes the Standard Error of Sum, and `clm` includes the 95% confident limits for the mean in the output statement. Other statements can also be included such as `max min median mode`.

Output:

Total averages by agegroup

The SURVEYMEANS Procedure

Statistics for agegrp Domains							
agegrp	Variable	Label	Mean	Std Error of Mean	95% CL for Mean		Std Error of Sum
2-5 yrs	carbohydrate_adjForVar_WTMN	FOODLOG Intake for weighted mean across days: Carbohydrate in g, adjusted for within person variance	224.943449	3.306968	217.850708	232.036190	44323
	dietaryfiber_adjForVar_WTMN	FOODLOG Intake for weighted mean across days: Dietary Fiber in g, adjusted for within person variance	9.336849	0.283940	8.727858	9.945840	2093.369378
	protein_adjForVar_WTMN	FOODLOG Intake for weighted mean across days: Protein in g, adjusted for within person variance	64.039844	0.660354	62.623527	65.456162	12530
	totalfat_adjForVar_WTMN	FOODLOG Intake for weighted mean across days: Total Fat in g, adjusted for within person variance	58.072946	0.739802	56.486229	59.659662	11984
6-8 yrs	carbohydrate_adjForVar_WTMN	FOODLOG Intake for weighted mean across days: Carbohydrate in g, adjusted for within person variance	246.809731	4.087420	238.043088	255.576375	34501
	dietaryfiber_adjForVar_WTMN	FOODLOG Intake for weighted mean across days: Dietary Fiber in g, adjusted for within person variance	10.584936	0.371781	9.787545	11.382326	1654.049881
	protein_adjForVar_WTMN	FOODLOG Intake for weighted mean across days: Protein in g, adjusted for within person variance	72.533541	0.944180	70.508476	74.558606	10502
	totalfat_adjForVar_WTMN	FOODLOG Intake for weighted mean across days: Total Fat in g, adjusted for within person variance	66.979681	0.973409	64.891927	69.067435	9719.492217

Statistical Procedures

Chi Square Test

Chi Square test is a statistical test to see whether there is a significant difference in the percent between categories/groups of categorical variables.

```
proc freq data=chldata;
table sex*bmi_category / chisq;
title "Chi Square Test of BMI categories between males & females";
run;
```

Output:

Chi Square Test of BMI categories between males & females

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of sex by bmi_category				
sex(DEMO: Sex of the CHL subject)	bmi_category(ANTHRO: BMI in 4 categories using CDC 2000 reference)				
	Healthy weight	Obese	Overweight	Underweight	Total
1	1851 34.11 66.85 49.51	457 8.42 16.50 59.04	389 7.17 14.05 50.13	72 1.33 2.60 52.55	2769 51.03
2	1888 34.80 71.06 50.49	317 5.84 11.93 40.96	387 7.13 14.57 49.87	65 1.20 2.45 47.45	2657 48.97
Total	3739 68.91	774 14.26	776 14.30	137 2.52	5426 100.00
Frequency Missing = 73					

Statistics for Table of sex by bmi_category

Statistic	DF	Value	Prob
Chi-Square	3	23.7502	<.0001
Likelihood Ratio Chi-Square	3	23.8800	<.0001
Mantel-Haenszel Chi-Square	1	3.1568	0.0756
Phi Coefficient		0.0662	
Contingency Coefficient		0.0660	
Cramer's V		0.0662	

Sample Size = 5426
Frequency Missing = 73

T Test

T test is a statistical test to see if there is a significant difference between the means of a continuous variable by a categorical variable with only 2 categories/groups (e.g., sex, which only has 2 groups – male or female).

```
proc ttest data=chldata;
var   carbohydrate_adjForVar_WTMN;
class sex;
title "t test of mean carbohydrate intake by sex";
run;
```

Output:

t test of mean carbohydrate intake by sex

The TTEST Procedure

Variable: carbohydrate_adjForVar_WTMN (FOODLOG Intake for weighted mean across days: Carbohydrate in g, adjusted for within person variance)

sex	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
1		1798	240.4	73.2909	1.7284	66.8758	634.8
2		1731	229.8	71.7257	1.7240	44.3718	537.3
Diff (1-2)	Pooled		10.5760	72.5274	2.4422		
Diff (1-2)	Satterthwaite		10.5760		2.4412		

sex	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1		240.4	237.0	243.8	73.2909	70.9713	75.7684
2		229.8	226.5	233.2	71.7257	69.4136	74.1983
Diff (1-2)	Pooled	10.5760	5.7877	15.3643	72.5274	70.8738	74.2605
Diff (1-2)	Satterthwaite	10.5760	5.7897	15.3624			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	3527	4.33	<.0001
Satterthwaite	Unequal	3526.1	4.33	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	1797	1730	1.04	0.3650

Note: Output will also present histogram plot and q-q plots of continuous variable disaggregated by categorical groups, which is not shown here.

ANOVA Test or Simple Linear Regression

ANOVA test is a statistical test to see if there is a significant difference between the means of a continuous variable by a categorical variable with more than 2 categories/groups.

ANOVA test is the simplest form of a general linear model (glm), so either **`proc anova`** or **`proc glm`** can be used.

```
proc anova data=chldata_new;
class bmi_category ;
model carbohydrate_adjForVar_WTMN = bmi_category;
title "ANOVA test of mean carbohydrate intake by BMI category";
run;

proc glm data=chldata_new;
class bmi_category ;
model carbohydrate_adjForVar_WTMN = bmi_category;
title "ANOVA test of mean carbohydrate intake by BMI category";
run;
```

Output:

ANOVA test of mean carbohydrate intake by BMI category

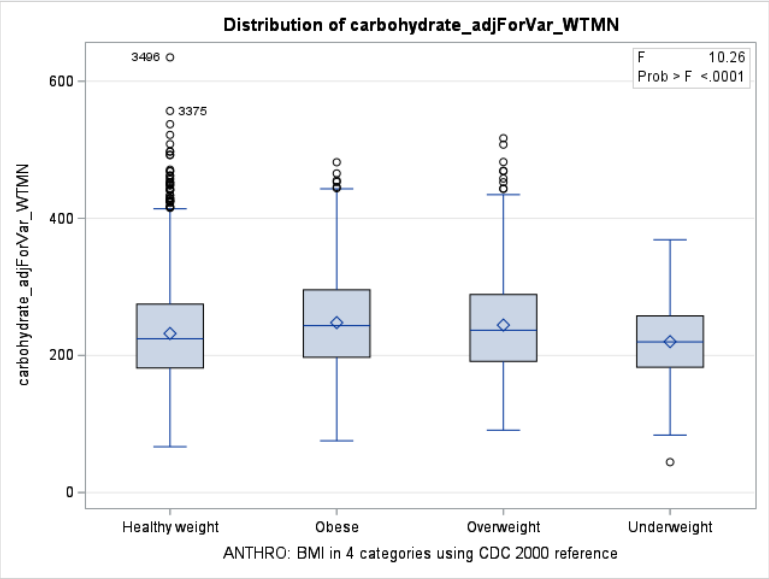
The ANOVA Procedure

Dependent Variable: carbohydrate_adjForVar_WTMN FOODLOG Intake for weighted mean across days: Carbohydrate in g, adjusted for within person variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	160745.05	53581.68	10.26	<.0001
Error	3476	18144964.47	5220.07		
Corrected Total	3479	18305709.52			

R-Square	Coeff Var	Root MSE	carbohydrate_adjForVar_WTMN Mean
0.008781	30.69114	72.25005	235.4101

Source	DF	Anova SS	Mean Square	F Value	Pr > F
bmi_category	3	160745.0531	53581.6844	10.26	<.0001



General Linear Regression Models

Multiple Linear Regression

Multiple Linear Regression models test the association between multiple independent variables and a dependent variable, which must be continuous and have a normal distribution.

```
proc glm data=chldata;
class race_new_omb sex;
model carbohydrate_adjForVar_WTMN = race_new_omb agemos sex / solution;
title "Multiple linear regression model testing association between
race/ethnicity & carbohydrate intake, adjusting for age & sex";
run;
```

Output:

Multiple linear regression model testing association between race/ethnicity & carbohydrate intake, adjusting for age & sex

The GLM Procedure

Dependent Variable: carbohydrate_adjForVar_WTMN FOODLOG Intake for weighted mean across days: Carbohydrate in g, adjusted for within person variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	1143296.53	163328.08	32.87	<.0001
Error	3508	17430821.84	4968.88		
Corrected Total	3515	18574118.37			

R-Square	Coeff Var	Root MSE	carbohydrate_adjForVar_WTMN Mean
0.061553	29.96912	70.49026	235.2882

Source	DF	Type I SS	Mean Square	F Value	Pr > F
race_new_omb	5	531587.9121	106317.5824	21.40	<.0001
agemos	1	501632.9224	501632.9224	100.95	<.0001
sex	1	110075.7001	110075.7001	22.15	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
race_new_omb	5	447426.0809	89485.2162	18.01	<.0001
agemos	1	504441.0164	504441.0164	101.52	<.0001
sex	1	110075.7001	110075.7001	22.15	<.0001

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	169.6255373	B	6.15362007	27.57	<.0001
race_new_omb AIAN	14.9721343	B	14.64311997	1.02	0.3066
race_new_omb Asian	3.1315536	B	6.51512181	0.48	0.6308
race_new_omb Black	-9.7382539	B	25.38428122	-0.38	0.7013
race_new_omb More than one race	4.9838193	B	5.55731945	0.90	0.3699
race_new_omb NHPI	27.5860158	B	5.06153347	5.45	<.0001
race_new_omb White	0.0000000	B	.	.	.
agemos	0.6072687		0.06027050	10.08	<.0001
sex 1	11.2022086	B	2.38005541	4.71	<.0001
sex 2	0.0000000	B	.	.	.

Multiple Linear Regression accounting for CHL complex sample design

Multiple Linear Regression model below accounts for the CHL complex sample design and adjusts for the jurisdiction strata and community cluster.

```
proc surveyreg data=chldata;
class race_new_omb sex;
model carbohydrate_adjForVar_WTMN = race_new_omb agemos sex / solution;
weight wt_anthro_adj;
strata jurisnum ;
cluster community_cluster;
title "Multiple linear regression model testing association between race/ethnicity &
carbohydrate intake, adjusting for age & sex, accounting for complex sample design";
run;
```

Output:

Multiple linear regression model testing association between race/ethnicity & carbohydrate intake, adjusting for age & sex, accounting for complex sample design

The SURVEYREG Procedure

Regression Analysis for Dependent Variable carbohydrate_adjForVar_WTMN

Data Summary	
Number of Observations	3516
Mean of carbohydrate_adjForVar_WTMN	235.28817
Sum of carbohydrate_adjForVar_WTMN	827273.2

Design Summary	
Number of Strata	11
Number of Strata Collapsed	6
Number of Clusters	25

Fit Statistics	
R-Square	0.06155
Root MSE	70.4903
Denominator DF	19

Class Level Information

CLASS Variable	Label	Levels	Values
race_new_omb	DEMO: Child race following guidelines of US Office of Management and Budget (OMB) Definition	6	AIAN Asian Black More than one race NHPI White
sex	DEMO: Sex of the CHL subject	2	1 2

Tests of Model Effects

Effect	Num DF	F Value	Pr > F
Model	7	27.63	<.0001
Intercept	1	473.78	<.0001
race_new_omb	5	8.00	0.0003
agemos	1	57.99	<.0001
sex	1	51.97	<.0001

Note: The denominator degrees of freedom for the F tests is 19.

Estimated Regression Coefficients

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	169.625537	6.8656171	24.71	<.0001
race_new_omb AIAN	14.972134	43.2790864	0.35	0.7332
race_new_omb Asian	3.131554	6.1568833	0.51	0.6169
race_new_omb Black	-9.738254	13.7062554	-0.71	0.4860
race_new_omb More than one race	4.983819	5.5479890	0.90	0.3803
race_new_omb NHPI	27.586016	7.4028330	3.73	0.0014
race_new_omb White	0.000000	0.0000000	.	.
agemos	0.607269	0.0797467	7.61	<.0001
sex 1	11.202209	1.5539234	7.21	<.0001
sex 2	0.000000	0.0000000	.	.

Logistic Regression Models

Multiple Logistic Regression

Multiple Logistic Regression models test the association between multiple independent variables and a dependent variable, which must be binary (i.e., only 2 groups).

Program below provides both beta coefficient estimates and odds ratios.

```
proc logistic data=chldata;
class race_new_omb (ref="White") sex (ref="1") / param = ref;
model owob (event="1") = race_new_omb agemos sex;
title "Multiple logistic regression model testing association between
race/ethnicity & overweight/obesity, adjusting for age & sex";
run;
```

Output:

Multiple logistic regression model testing association between race/ethnicity & overweight/obesity, adjusting for age & sex

The LOGISTIC Procedure

Model Information		
Data Set	WORK.CHLDATA_NEW	
Response Variable	owob	ANTHRO: Overweight or Obesity based >= 85th percentile for CDC growth curves (0=No, 1=Yes)
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	5499
Number of Observations Used	5407

Response Profile		
Ordered Value	owob	Total Frequency
1	0	3860
2	1	1547

Probability modeled is owob=1.

Class Level Information						
Class	Value	Design Variables				
race_new_omb	AIAN	1	0	0	0	0
	Asian	0	1	0	0	0
	Black	0	0	1	0	0
	More than one race	0	0	0	1	0
	NHPI	0	0	0	0	1
	White	0	0	0	0	0
sex	1	0				
	2	1				

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	6475.610	6450.003
SC	6482.206	6502.767
-2 Log L	6473.610	6434.003

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	39.6072	7	<.0001
Score	42.0019	7	<.0001
Wald	39.7028	7	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
race_new_omb	5	25.6951	0.0001
agemos	1	3.6436	0.0563
sex	1	11.8298	0.0006

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-1.0163	0.1457	48.6333	<.0001
race_new_omb	AIAN	1	0.7640	0.2134	12.8171	0.0003
race_new_omb	Asian	1	-0.0566	0.1501	0.1422	0.7061
race_new_omb	Black	1	1.4608	0.5286	7.6380	0.0057
race_new_omb	More than one race	1	-0.0190	0.1293	0.0216	0.8833
race_new_omb	NHPI	1	-0.00946	0.1169	0.0065	0.9355
agemos		1	0.00296	0.00155	3.6436	0.0563
sex	2	1	-0.2084	0.0608	11.8298	0.0006

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
race_new_omb AIAN vs White	2.147	1.413	3.262
race_new_omb Asian vs White	0.945	0.704	1.268
race_new_omb Black vs White	4.310	1.529	12.144
race_new_omb More than one race vs White	0.981	0.761	1.264
race_new_omb NHPI vs White	0.991	0.788	1.246
agemos	1.003	1.000	1.006
sex 2 vs 1	0.812	0.721	0.914

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	54.1	Somers' D	0.082
Percent Discordant	45.9	Gamma	0.082
Percent Tied	0.0	Tau-a	0.034
Pairs	5971420	c	0.541

Multiple Logistic Regression accounting for CHL complex sample design

Multiple Logistic Regression models below accounts for the CHL complex sample design and adjusts for the jurisdiction strata and community cluster.

Program below provides both beta coefficient estimates and odds ratios.

```
proc surveylogistic data = chldata_new;
class race_new_omb (ref="White") sex  / param = ref;
model owob (event="1") = race_new_omb agemos sex;
weight wt_anthro_adj;
strata jurisnum ;
cluster community_cluster;
title "Multiple logistic regression model testing association between
race/ethnicity & overweight/obesity, adjusting for age & sex, accouting for
complex sample design";
run;
```

Output:

Multiple logistic regression model testing association between race/ethnicity & overweight/obesity, adjusting for age & sex, accouting for complex sample design

The SURVEYLOGISTIC Procedure

Model Information		
Data Set	WORK.CHLDATA	
Response Variable	owob	ANTHRO: Overweight or Obesity based >= 85th percentile for CDC growth curves (0=No, 1=Yes)
Number of Response Levels	2	
Stratum Variable	jurisnum	GENERAL: Jurisdiction number
Number of Strata	11	
Cluster Variable	community_cluster	GENERAL: Community for cluster analysis
Number of Clusters	33	
Weight Variable	wt_anthro_adj	GENERAL: Sampling weight for each child based on the population of children at each community and adjusted so that the sum of weights = sample size.
Model	Binary Logit	
Optimization Technique	Fisher's Scoring	
Variance Adjustment	Degrees of Freedom (DF)	

Variance Estimation	
Method	Taylor Series
Variance Adjustment	Degrees of Freedom (DF)

Number of Observations Read	5499
Number of Observations Used	5407
Sum of Weights Read	5484.646
Sum of Weights Used	5394.155

Response Profile			
Ordered Value	owob	Total Frequency	Total Weight
1	0	3860	3868.2416
2	1	1547	1525.9130

Probability modeled is owob=1.

Class Level Information						
Class	Value	Design Variables				
race_new_omb	AIAN	1	0	0	0	0
	Asian	0	1	0	0	0
	Black	0	0	1	0	0
	More than one race	0	0	0	1	0
	NHPI	0	0	0	0	1
	White	0	0	0	0	0
sex	1	0				
	2	1				

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	6428.114	6390.635
SC	6434.707	6443.379
-2 Log L	6426.114	6374.635

Testing Global Null Hypothesis: BETA=0				
Test	F Value	Num DF	Den DF	Pr > F
Likelihood Ratio	5.45	3.1096	68.4117	0.0018
Score	4.67	7	16	0.0051
Wald	23.06	7	16	<.0001
NOTE: Second-order Rao-Scott design correction 1.2511 applied to the Likelihood Ratio test.				

Type 3 Analysis of Effects				
Effect	F Value	Num DF	Den DF	Pr > F
race_new_omb	34.98	5	18	<.0001
agemos	1.51	1	22	0.2319
sex	11.58	1	22	0.0026

Analysis of Maximum Likelihood Estimates					
Parameter		Estimate	Standard Error	t Value	Pr > |t|
Intercept		-1.2304	0.3388	-3.63	0.0015
race_new_omb	AIAN	0.9620	0.2160	4.45	0.0002
race_new_omb	Asian	0.2180	0.1951	1.12	0.2758
race_new_omb	Black	1.6920	0.3230	5.24	<.0001
race_new_omb	More than one race	0.2096	0.2273	0.92	0.3663
race_new_omb	NHPI	0.1619	0.1916	0.85	0.4072
agemos		0.00314	0.00255	1.23	0.2319
sex	2	-0.2155	0.0633	-3.40	0.0026
NOTE: The degrees of freedom for the t tests is 22.					

Odds Ratio Estimates			
Effect	Point Estimate	95% Confidence Limits	
race_new_omb AIAN vs White	2.617	1.672	4.096
race_new_omb Asian vs White	1.244	0.830	1.864
race_new_omb Black vs White	5.430	2.779	10.611
race_new_omb More than one race vs White	1.233	0.770	1.976
race_new_omb NHPI vs White	1.176	0.790	1.749
agemos	1.003	0.998	1.008
sex 2 vs 1	0.806	0.707	0.919
NOTE: The degrees of freedom in computing the confidence limits is 22.			

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	52.9	Somers' D	0.078
Percent Discordant	45.1	Gamma	0.080
Percent Tied	2.0	Tau-a	0.032
Pairs	5971420	c	0.539

Additional Support

For additional support on SAS programming, many online resources are available including:

SAS Help Center

UCLA Statistical Methods and Data Analytics

SAS Support Community

Searching on Google or asking ChapGPT are also helpful resources for SAS.

The CHL Data Center can also assist with your analysis. Submit a Service Request for consultation with an analyst.