CHAPTER 7 ASDA ANALYSIS EXAMPLES REPLICATION-SPSS/PASW V18 COMPLEX SAMPLES

GENERAL NOTES ABOUT ANALYSIS EXAMPLES REPLICATION
These examples are intended to provide guidance on how to use the commands/procedures for analysis of complex sample survey data and assume all data management and other preliminary work is done. The relevant syntax for the procedure of interest is shown first along with the associated output for that procedure(s). In some examples, there may be more than one block of syntax and in this case all syntax is first presented followed by the output produced.

In some software packages certain procedures or options are not available but we have made every attempt to demonstrate how to match the output produced by Stata 10+ in the textbook. Check the ASDA website for updates to the various software tools we cover.

NOTES ABOUT LINEAR REGRESSION ANALYSIS IN SPSS/PASW V18 COMPLEX SAMPLES MODULE

SPSS/PASW CSGLM can perform nearly all of the analysis of linear regression examples presented in Chapter 7 of ASDA. Some of the fine points of these procedures are the use of a SUBPOP statement for subpopulation analyses, various output statistics specified on the STATISTICS subcommand, and use of an analysis Plan file for all Complex Samples commands. The plan file should be prepared prior to working with any Complex Samples commands and offers the ability to declare weights and design variables to the program.

Throughout this chapter we use both original coding and reverse coding for the categorical or factor variables that serve as predictor variables. This is required due to the desire to match the omitted categories to Stata's default of the lowest category as the referent. The codes are simply reversed such that 5=1, 4=2, 3=3, 2=4, 1=5 for a 5 category variable. This enables use of the factor statement in the regression models. The goal of obtaining the correct omitted category could also be accomplished by using 0/1 indicator variables but hypothesis testing of categorical variables becomes more complex. We demonstrate both techniques but favor a reverse coding scheme for the ease of programming.

This command also includes the ability to save predicted values and residuals from CSGLM and use for further diagnostic analyses. This is demonstrated during the diagnostic work with the final model selected to predict diastolic blood pressure in the NHANES adult sample. This version of the Complex Samples module does not include the ability to produce Qnorm or Symplot of Residuals and are omitted however the Simple Random Sample version of these linear regression diagnostics would be an acceptable alternative. Please see the SPSS/PASW documentation for details.

Another technique included in this chapter's output is custom hypothesis tests for the interactions of race*age and race*age squared along with gender*age and gender*age squared. These interactions are specified for tests of the null hypothesis of the interactions being equal to zero. These types of custom tests are possible using syntax based analysis in SPSS v18 through use of the /CUSTOM subcommand.

* ANALYSIS EXAMPLE 7.5: DIASTOLIC BLOOD PRESSURE PREDICTED BY RACE/ETHNICITY NHANES ADULT DATA

* Complex Samples General Linear Model.
CSGLM bpxdi1_1 BY revrace
  /PLAN FILE='F:\applied_analysis_book\csplan_nhanes.csaplan'
  /DOMAIN VARIABLE=age18p(1)
  /MODEL revrace
  /INTERCEPT INCLUDE=YES SHOW=YES
  /STATISTICS PARAMETER SE CINTERVAL TTEST DEFF
  /PRINT SUMMARY VARIABLEINFO SAMPLEINFO
  /TEST TYPE=F PADJUST=LSD
  /MISSING CLASSMISSING=EXCLUDE
  /CRITERIA CILEVEL=95.

Complex Samples: General Linear Model

Sample Design Information

		N
Unweighted Cases	Valid	4581
	Invalid	982
	Total	5563
Population Size		1.900E8
Subpopulation Size		1.900E8
Stage 1	Strata	15
	Units	30
Sampling Design Degrees of Freedom		15

a. Subpopulation: age18p = 1

Factor Information

		Weighted Count	Weighted Percent
revrace	1	9915591.033	5.2%
	2	2.047E7	10.8%
	3	1.380E8	72.6%
	4	6407271.118	3.4%
	5	1.519E7	8.0%
	Subpopulation Size	1.900E8	100.0%

Subpopulation: age18p = 1

Model Summary[a]

R Square	.005

Subpopulation: age18p

= 1

a. Model: bpxdi1_1 =

(Intercept) + revrace

Tests of Model Effects[a]

Source	df1	df2	Wald F	Sig.
(Corrected Model)	4.000	12.000	6.229	.006
(Intercept)	1.000	15.000	18774.016	.000
revrace	4.000	12.000	6.229	.006

Subpopulation: age18p = 1

a. Model: bpxdi1_1 = (Intercept) + revrace

Parameter Estimates[b]

Parameter	Estimate	Std. Error	95% Confidence Interval		Hypothesis Test			Design Effect
			Lower	Upper	t	df	Sig.	
(Intercept)	68.300	.412	67.420	69.179	165.587	15.000	.000	.470
[revrace=1.00]	1.785	1.030	-.410	3.980	1.733	15.000	.104	1.081
[revrace=2.00]	3.728	.753	2.122	5.333	4.949	15.000	.000	.802
[revrace=3.00]	2.428	.554	1.246	3.609	4.380	15.000	.001	.754
[revrace=4.00]	1.592	1.109	-.771	3.956	1.436	15.000	.172	.957
[revrace=5.00]	.000[a]	.	.	.	.	.	.	.

Subpopulation: age18p = 1

a. Set to zero because this parameter is redundant.

b. Model: bpxdi1_1 = (Intercept) + revrace

NOTE: CODES FOR REVRACE 1=OTHER 2=BLACK 3=WHITE

4=OTHER HISPANIC 5=MEXICAN

```
* ANALYSIS EXAMPLE 7.5: DIASTOLIC BLOOD PRESSURE PREDICTED BY AGE
* Complex Samples General Linear Model.
CSGLM  bpxdi1_1 WITH agec
  /PLAN FILE='F:\applied_analysis_book\csplan_nhanes.csaplan'
  /DOMAIN VARIABLE=age18p(1)
  /MODEL agec
  /INTERCEPT INCLUDE=YES SHOW=YES
  /STATISTICS PARAMETER SE TTEST
  /PRINT SUMMARY VARIABLEINFO SAMPLEINFO
  /TEST TYPE=F PADJUST=LSD
  /MISSING CLASSMISSING=EXCLUDE
  /CRITERIA CILEVEL=95.
```

Complex Samples: General Linear Model

Sample Design Information

		N
Unweighted Cases	Valid	4581
	Invalid	982
	Total	5563
Population Size		1.900E8
Subpopulation Size		1.900E8
Stage 1	Strata	15
	Units	30
Sampling Design Degrees of Freedom		15

a. Subpopulation: age18p = 1

Variable Information

		Mean
Dependent Variable	bpxdi1_1	70.61
Covariates	agec	-.0717

Subpopulation: age18p = 1

Model Summary[a]

R Square	.006

Subpopulation: age18p

= 1

a. Model: bpxdi1_1 =

(Intercept) + agec

Tests of Model Effects[a]

Source	df1	df2	Wald F	Sig.
(Corrected Model)	1.000	15.000	7.696	.014
(Intercept)	1.000	15.000	40780.471	.000
agec	1.000	15.000	7.696	.014

Subpopulation: age18p = 1

a. Model: bpxdi1_1 = (Intercept) + agec

Parameter Estimates[a]

Parameter	Estimate	Std. Error	Hypothesis Test		
			t	df	Sig.
(Intercept)	70.616	.350	201.942	15.000	.000
agec	.057	.021	2.774	15.000	.014

Subpopulation: age18p = 1

a. Model: bpxdi1_1 = (Intercept) + agec

```
* ANALYSIS EXAMPLE7 7.5: DIASTOLIC BLOOD PRESSURE PREDICTED BY FEMALE
* Complex Samples General Linear Model.
CSGLM  bpxdi1_1 WITH female
  /PLAN FILE='F:\applied_analysis_book\csplan_nhanes.csaplan'
  /DOMAIN VARIABLE=age18p(1)
  /MODEL female
  /INTERCEPT INCLUDE=YES SHOW=YES
  /STATISTICS PARAMETER SE TTEST
  /PRINT SUMMARY VARIABLEINFO SAMPLEINFO
  /TEST TYPE=F PADJUST=LSD
  /MISSING CLASSMISSING=EXCLUDE
  /CRITERIA CILEVEL=95.
```

Complex Samples: General Linear Model

Sample Design Information

		N
Unweighted Cases	Valid	4581
	Invalid	982
	Total	5563
Population Size		1.900E8
Subpopulation Size		1.900E8
Stage 1	Strata	15
	Units	30
Sampling Design Degrees of Freedom		15

a. Subpopulation: age18p = 1

Variable Information

		Mean
Dependent Variable	bpxdi1_1	70.61
Covariates	female	.51

Subpopulation: age18p = 1

Model Summary[a]

R Square	.013

Subpopulation: age18p

= 1

a. Model: bpxdi1_1 =

(Intercept) + female

Tests of Model Effects[a]

Source	df1	df2	Wald F	Sig.
(Corrected Model)	1.000	15.000	56.430	.000
(Intercept)	1.000	15.000	29368.085	.000
female	1.000	15.000	56.430	.000

Subpopulation: age18p = 1

a. Model: bpxdi1_1 = (Intercept) + female

Parameter Estimates[a]

Parameter	Estimate	Std. Error	Hypothesis Test		
			t	df	Sig.
(Intercept)	72.069	.421	171.371	15.000	.000
female	-2.844	.379	-7.512	15.000	.000

Subpopulation: age18p = 1

a. Model: bpxdi1_1 = (Intercept) + female

```
* ANALYSIS EXAMPLE 7.5: DIASTOLIC BLOOD PRESSURE PREDICTED BY MARITAL STATUS
* Complex Samples General Linear Model.
CSGLM  bpxdi1_1 BY revMARCAT
  /PLAN FILE='F:\applied_analysis_book\csplan_nhanes.csaplan'
  /DOMAIN VARIABLE=age18p(1)
  /MODEL revmarcat
  /INTERCEPT INCLUDE=YES SHOW=YES
  /STATISTICS PARAMETER SE TTEST
  /PRINT SUMMARY VARIABLEINFO SAMPLEINFO
  /TEST TYPE=F PADJUST=LSD
  /MISSING CLASSMISSING=EXCLUDE
  /CRITERIA CILEVEL=95.


NOTE: CODES FOR REVMARCAT 1=NEVER MARRIED 2=PREVIOUSLY MARRIED 3=MARRIED
```

Complex Samples: General Linear Model

Sample Design Information

		N
Unweighted Cases	Valid	4578
	Invalid	985
	Total	5563
Population Size		1.898E8
Subpopulation Size		1.898E8
Stage 1	Strata	15
	Units	30
Sampling Design Degrees of Freedom		15

a. Subpopulation: age18p = 1

Variable Information

	Mean
Dependent Variable bpxdi1_1	70.61

Subpopulation: age18p = 1

Factor Information

		Weighted Count	Weighted Percent
revmarcat	1	3.325E7	17.5%
	2	3.351E7	17.7%
	3	1.231E8	64.8%
	Subpopulation Size	1.898E8	100.0%

Subpopulation: age18p = 1

Model Summary[a]

R Square	.018

Subpopulation: age18p

= 1

a. Model: bpxdi1_1 =

(Intercept) +

revmarcat

Tests of Model Effects[a]

Source	df1	df2	Wald F	Sig.
(Corrected Model)	2.000	14.000	37.480	.000
(Intercept)	1.000	15.000	44999.359	.000
revmarcat	2.000	14.000	37.480	.000

Subpopulation: age18p = 1

a. Model: bpxdi1_1 = (Intercept) + revmarcat

Parameter Estimates[b]

Parameter	Estimate	Std. Error	Hypothesis Test		
			t	df	Sig.
(Intercept)	71.392	.468	152.696	15.000	.000
[revmarcat=1.00]	-4.386	.573	-7.654	15.000	.000
[revmarcat=2.00]	-.073	.681	-.108	15.000	.916
[revmarcat=3.00]	.000[a]	.	.	.	.

Subpopulation: age18p = 1

a. Set to zero because this parameter is redundant.

b. Model: bpxdi1_1 = (Intercept) + revmarcat

```
* ANALYSIS EXAMPLE 7.5: DIASTOLIC BLOOD PRESSURE WITH ALL PREDICTORS IN MODEL NO WEIGHTS OR COMPLEX DESIGN VARIABLES.
* Generalized Linear Models.
GENLIN bpxdi1_1 BY revrace revmarcat WITH agec female
  /MODEL revrace revmarcat agec female INTERCEPT=YES
 DISTRIBUTION=NORMAL LINK=IDENTITY
  /CRITERIA SCALE=1 COVB=MODEL PCONVERGE=1E-006(ABSOLUTE)
 SINGULAR=1E-012 ANALYSISTYPE=3(WALD) CILEVEL=95 CITYPE=WALD LIKELIHOOD=FULL
  /MISSING CLASSMISSING=EXCLUDE
  /PRINT CPS DESCRIPTIVES MODELINFO FIT SUMMARY SOLUTION.
```

Generalized Linear Models

Model Information

Dependent Variable	bpxdi1_1
Probability Distribution	Normal
Link Function	Identity

Case Processing Summary

	N	Percent
Included	4578	82.3%
Excluded	985	17.7%
Total	5563	100.0%

Categorical Variable Information

			N	Percent
Factor	revrace	1	184	4.0%
		2	1050	22.9%
		3	2228	48.7%
		4	141	3.1%
		5	975	21.3%
		Total	4578	100.0%
	revmarcat	1	1042	22.8%
		2	864	18.9%
		3	2672	58.4%
		Total	4578	100.0%

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Dependent Variable	bpxdi1_1	4578	4	124	68.99	12.868
Covariate	agec	4578	-27.60	39.40	-.6104	19.91572
	female	4578	0	1	.52	.500

Goodness of Fit[b]

	Value	df	Value/df
Deviance	712510.734	4569	155.945
Scaled Deviance	712510.734	4569	
Pearson Chi-Square	712510.734	4569	155.945
Scaled Pearson Chi-Square	712510.734	4569	
Log Likelihood[a]	-360462.268		
Akaike's Information Criterion (AIC)	720942.536		
Finite Sample Corrected AIC (AICC)	720942.575		
Bayesian Information Criterion (BIC)	721000.397		
Consistent AIC (CAIC)	721009.397		

Dependent Variable: bpxdi1_1

Model: (Intercept), revrace, revmarcat, agec, female

a. The full log likelihood function is displayed and used in computing information criteria.

b. Information criteria are in small-is-better form.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
45386.916	8	.000

Dependent Variable: bpxdi1_1

Model: (Intercept), revrace, revmarcat, agec, female

a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

| Source | Type III | | |
	Wald Chi-Square	df	Sig.
(Intercept)	5238709.396	1	.000
revrace	10445.093	4	.000
revmarcat	10992.740	2	.000
agec	1804.848	1	.000
female	12861.121	1	.000

Dependent Variable: bpxdi1_1

Model: (Intercept), revrace, revmarcat, agec,

female

Parameter Estimates

| Parameter | B | Std. Error | 95% Wald Confidence Interval | | Hypothesis Test | | |
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	69.672	.0372	69.599	69.745	3510778.205	1	.000
[revrace=1.00]	2.312	.0804	2.154	2.470	826.028	1	.000
[revrace=2.00]	4.508	.0451	4.420	4.597	9982.011	1	.000
[revrace=3.00]	1.672	.0394	1.595	1.749	1804.707	1	.000
[revrace=4.00]	1.898	.0901	1.722	2.075	443.681	1	.000
[revrace=5.00]	0[a]	.	.	.	.	.	.
[revmarcat=1.00]	-4.216	.0408	-4.296	-4.136	10656.388	1	.000
[revmarcat=2.00]	.327	.0418	.245	.409	61.112	1	.000
[revmarcat=3.00]	0[a]	.	.	.	.	.	.
Agec	.039	.0009	.037	.041	1804.848	1	.000
Female	-3.402	.0300	-3.461	-3.343	12861.121	1	.000
(Scale)	1[b]						

Dependent Variable: bpxdi1_1

Model: (Intercept), revrace, revmarcat, agec, female

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

* ANALYSIS EXAMPLE 7.5: DIASTOLIC BLOOD PRESSURE WITH ALL PREDICTORS IN MODEL WITH WEIGHT BUT NO COMPLEX DESIGN VARIABLES.
* NOTE THAT WITH GENLIN YOU CAN USE THE ORIGINAL CODING FOR RACE/ETHNICITY AND MARITAL STATUS AND SPECIFY THE ORDER = DESCENDING OPTION TO
MATCH THE OMITTED CATEGORY OF LOWEST (IN STATA)

* Generalized Linear Models.
GENLIN bpxdi1_1 BY RIDRETH1 marcat (ORDER=DESCENDING) WITH agec female
 /MODEL RIDRETH1 marcat agec female INTERCEPT=YES SCALEWEIGHT=WTMEC2YR
 DISTRIBUTION=NORMAL LINK=IDENTITY
 /CRITERIA SCALE=1 COVB=MODEL PCONVERGE=1E-006(ABSOLUTE) SINGULAR=1E-012
 ANALYSISTYPE=3(WALD) CILEVEL=95 CITYPE=WALD LIKELIHOOD=FULL
  /MISSING CLASSMISSING=EXCLUDE
  /PRINT CPS DESCRIPTIVES MODELINFO FIT SUMMARY SOLUTION.

Generalized Linear Models

Model Information

Dependent Variable	bpxdi1_1
Probability Distribution	Normal
Link Function	Identity
Scale Weight Variable	Full Sample 2 Year MEC Exam Weight

Case Processing Summary

	N	Percent
Included	4578	82.3%
Excluded	985	17.7%
Total	5563	100.0%

Categorical Variable Information

			N	Percent
Factor	1=mex 2=oth hisp 3=white 4=black 5=other	5	184	4.0%
		4	1050	22.9%
		3	2228	48.7%
		2	141	3.1%
		1	975	21.3%
		Total	4578	100.0%
	1=married 2=prev married 3=never married	3	1042	22.8%
		2	864	18.9%
		1	2672	58.4%

Categorical Variable Information

			N	Percent
Factor	1=mex 2=oth hisp 3=white 4=black 5=other	5	184	4.0%
		4	1050	22.9%
		3	2228	48.7%
		2	141	3.1%
		1	975	21.3%
		Total	4578	100.0%
	1=married 2=prev married 3=never married	3	1042	22.8%
		2	864	18.9%
		1	2672	58.4%
		Total	4578	100.0%

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Dependent Variable	bpxdi1_1	4578	4	124	68.99	12.868
Covariate	agec	4578	-27.60	39.40	-.6104	19.91572
	female	4578	0	1	.52	.500
Scale Weight	Full Sample 2 Year MEC Exam Weight	4578	1363	156152	41469.6640	31722.36577

Goodness of Fit[b]

	Value	df	Value/df
Deviance	2.769E10	4569	6059872.965
Scaled Deviance	2.769E10	4569	
Pearson Chi-Square	2.769E10	4569	6059872.965
Scaled Pearson Chi-Square	2.769E10	4569	
Log Likelihood[a]	-1.384E10		
Akaike's Information Criterion (AIC)	2.769E10		
Finite Sample Corrected AIC (AICC)	2.769E10		
Bayesian Information Criterion (BIC)	2.769E10		
Consistent AIC (CAIC)	2.769E10		

Dependent Variable: bpxdi1_1

Model: (Intercept), RIDRETH1, marcat, agec, female

a. The full log likelihood function is displayed and used in computing information criteria.

b. Information criteria are in small-is-better form.

Tests of Model Effects

Source	Type III		
	Wald Chi-Square	df	Sig.
(Intercept)	2.106E11	1	.000
RIDRETH1	1.718E8	4	.000
marcat	4.229E8	2	.000
agec	1.221E7	1	.000
female	4.187E8	1	.000

Dependent Variable: bpxdi1_1

Model: (Intercept), RIDRETH1, marcat, agec, female

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	70.678	.0003	70.678	70.679	6.809E10	1	.000
[RIDRETH1=5]	1.958	.0004	1.958	1.959	2.294E7	1	.000
[RIDRETH1=4]	4.409	.0003	4.408	4.409	1.667E8	1	.000
[RIDRETH1=3]	2.192	.0003	2.191	2.192	6.417E7	1	.000
[RIDRETH1=2]	1.787	.0005	1.786	1.787	1.437E7	1	.000
[RIDRETH1=1]	0[a]	.	.	.	.	.	.
[marcat=3]	-4.356	.0002	-4.357	-4.356	4.188E8	1	.000
[marcat=2]	.017	.0002	.017	.018	7118.694	1	.000
[marcat=1]	0[a]	.	.	.	.	.	.
agec	.017	4.8743E-6	.017	.017	1.221E7	1	.000
female	-2.997	.0001	-2.998	-2.997	4.187E8	1	.000
(Scale)	1[b]						

Dependent Variable: bpxdi1_1

Model: (Intercept), RIDRETH1, marcat, agec, female

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

```
* ANALYSIS EXAMPLE 7.5: ALL PREDICTORS IN MODEL AND WEIGHT APPLIED AND USE OF COMPLEX DESIGN VARIABLES
* Complex Samples General Linear Model.
CSGLM  bpxdi1_1 by revrace revmarcat with female agec
  /PLAN FILE='F:\applied_analysis_book\csplan_nhanes.csaplan'
  /DOMAIN VARIABLE=age18p(1)
  /MODEL  revrace revmarcat  female agec
  /INTERCEPT INCLUDE=YES SHOW=YES
  /STATISTICS PARAMETER SE CINTERVAL TTEST DEFF
  /PRINT SUMMARY VARIABLEINFO SAMPLEINFO
  /TEST TYPE=F PADJUST=LSD
  /MISSING CLASSMISSING=EXCLUDE
  /CRITERIA CILEVEL=95.
```

Sample Design Information

		N
Unweighted Cases	Valid	4578
	Invalid	985
	Total	5563
Population Size		1.898E8
Subpopulation Size		1.898E8
Stage 1	Strata	15
	Units	30
Sampling Design Degrees of Freedom		15

a. Subpopulation: age18p = 1

Variable Information

		Mean
Dependent Variable	bpxdi1_1	70.61
Covariates	female	.51
	agec	-.0840

Subpopulation: age18p = 1

Factor Information

		Weighted Count	Weighted Percent
Revrace	1	9915591.033	5.2%
	2	2.047E7	10.8%
	3	1.379E8	72.6%
	4	6407271.118	3.4%
	5	1.519E7	8.0%
Revmarcat	1	3.325E7	17.5%
	2	3.351E7	17.7%
	3	1.231E8	64.8%
Subpopulation Size		1.898E8	100.0%

Sample Design Information

		N
Unweighted Cases	Valid	4578
	Invalid	985
	Total	5563
Population Size		1.898E8
Subpopulation Size		1.898E8
Stage 1	Strata	15
	Units	30
Sampling Design Degrees of Freedom		15

Subpopulation: age18p = 1

Model Summary[a]

R Square	.039

Subpopulation: age18p
= 1

a. Model: bpxdi1_1 =
(Intercept) + revrace
+ revmarcat + female
+ agec

Tests of Model Effects[a]

Source	df1	df2	Wald F	Sig.
(Corrected Model)	8.000	8.000	12.663	.001
(Intercept)	1.000	15.000	16880.579	.000
revrace	4.000	12.000	7.557	.003
revmarcat	2.000	14.000	30.554	.000
female	1.000	15.000	81.941	.000
agec	1.000	15.000	.607	.448

Subpopulation: age18p = 1

a. Model: bpxdi1_1 = (Intercept) + revrace + revmarcat + female
+ agec

Parameter Estimates[b]

Parameter	Estimate	Std. Error	95% Confidence Interval		Hypothesis Test			Design Effect
			Lower	Upper	t	df	Sig.	
(Intercept)	70.678	.501	69.611	71.745	141.141	15.000	.000	.627
[revrace=1.00]	1.958	.988	-.148	4.064	1.982	15.000	.066	1.045
[revrace=2.00]	4.409	.761	2.786	6.031	5.792	15.000	.000	.839
[revrace=3.00]	2.192	.605	.903	3.481	3.624	15.000	.002	.901
[revrace=4.00]	1.787	1.142	-.648	4.221	1.564	15.000	.139	1.039
[revrace=5.00]	.000[a]	.	.	.	.	.	.	.
[revmarcat=1.00]	-4.356	.565	-5.560	-3.152	-7.710	15.000	.000	1.118
[revmarcat=2.00]	.017	.718	-1.513	1.547	.024	15.000	.981	1.772
[revmarcat=3.00]	.000[a]	.	.	.	.	.	.	.
female	-2.997	.331	-3.703	-2.292	-9.052	15.000	.000	.858
agec	.017	.022	-.030	.064	.779	15.000	.448	2.619

Subpopulation: age18p = 1

a. Set to zero because this parameter is redundant.

b. Model: bpxdi1_1 = (Intercept) + revrace + revmarcat + female + agec

```
* ANALYSIS EXAMPLE 7.5: ALL PREDICTORS IN MODEL AND WEIGHT APPLIED AND USE OF COMPLEX DESIGN VARIABLES, ADD AGE SQUARED TO MODEL
* Complex Samples General Linear Model.
CSGLM  bpxdi1_1 by revrace  revmarcat WITH female agec agecsq
  /PLAN FILE='F:\applied_analysis_book\csplan_nhanes.csaplan'
  /DOMAIN VARIABLE=age18p(1)
  /MODEL revrace revmarcat female agec agecsq
  /INTERCEPT INCLUDE=YES SHOW=YES
  /STATISTICS PARAMETER SE CINTERVAL TTEST DEFF
  /PRINT SUMMARY VARIABLEINFO SAMPLEINFO
  /TEST TYPE=F PADJUST=LSD
  /MISSING CLASSMISSING=EXCLUDE
  /CRITERIA CILEVEL=95.
```

Complex Samples: General Linear Model

Sample Design Information

		N
Unweighted Cases	Valid	4578
	Invalid	985
	Total	5563
Population Size		1.898E8
Subpopulation Size		1.898E8
Stage 1	Strata	15
	Units	30
Sampling Design Degrees of Freedom		15

a. Subpopulation: age18p = 1

Variable Information

		Mean
Dependent Variable	bpxdi1_1	70.61
Covariates	female	.51
	agec	-.0840
	agecsq	299.4190

Subpopulation: age18p = 1

Factor Information

		Weighted Count	Weighted Percent
revrace	1	9915591.033	5.2%
	2	2.047E7	10.8%
	3	1.379E8	72.6%
	4	6407271.118	3.4%
	5	1.519E7	8.0%
revmarcat	1	3.325E7	17.5%
	2	3.351E7	17.7%
	3	1.231E8	64.8%
Subpopulation Size		1.898E8	100.0%

Subpopulation: age18p = 1

Model Summary[a]

R Square	.134

Subpopulation: age18p

= 1

a. Model: bpxdi1_1 =

(Intercept) + revrace

+ revmarcat + female

+ agec + agecsq

Tests of Model Effects[a]

Source	df1	df2	Wald F	Sig.
(Corrected Model)	9.000	7.000	87.115	.000
(Intercept)	1.000	15.000	22223.287	.000
revrace	4.000	12.000	5.118	.012
revmarcat	2.000	14.000	1.506	.256
female	1.000	15.000	64.985	.000
agec	1.000	15.000	71.462	.000
agecsq	1.000	15.000	266.864	.000

Subpopulation: age18p = 1

a. Model: bpxdi1_1 = (Intercept) + revrace + revmarcat + female

+ agec + agecsq

Parameter Estimates[b]

Parameter	Estimate	Std. Error	95% Confidence Interval		Hypothesis Test			Design Effect
			Lower	Upper	t	df	Sig.	
(Intercept)	73.859	.455	72.889	74.829	162.369	15.000	.000	.548
[revrace=1.00]	1.189	.934	-.803	3.180	1.272	15.000	.223	1.019
[revrace=2.00]	3.465	.779	1.804	5.126	4.447	15.000	.000	.983
[revrace=3.00]	1.781	.631	.436	3.125	2.823	15.000	.013	1.109
[revrace=4.00]	1.189	1.087	-1.127	3.505	1.094	15.000	.291	1.032
[revrace=5.00]	.000[a]	.	.	.	.	.	.	.
[revmarcat=1.00]	-.343	.582	-1.583	.897	-.590	15.000	.564	1.213
[revmarcat=2.00]	1.040	.622	-.285	2.366	1.673	15.000	.115	1.465
[revmarcat=3.00]	.000[a]	.	.	.	.	.	.	.
female	-2.721	.338	-3.441	-2.002	-8.061	15.000	.000	.985
agec	.125	.015	.094	.157	8.454	15.000	.000	1.345
agecsq	-.012	.001	-.014	-.011	-16.336	15.000	.000	1.496

Subpopulation: age18p = 1

a. Set to zero because this parameter is redundant.

b. Model: bpxdi1_1 = (Intercept) + revrace + revmarcat + female + agec + agecsq

* ANALYSIS EXAMPLE 7.5: ALL PREDICTORS PLUS INTERACTION OF RACE TIMES AGE AND AGE SQUARED

```
CSGLM  bpxdi1_1 BY revrace revmarcat WITH agesq agecent female
  /PLAN FILE='F:\applied_analysis_book\SPSS Analysis Examples Replication\Analysis Examples Replication Winter 2010
SPSSv18\csplan_nhanes.csaplan'
  /DOMAIN VARIABLE=age18(1)
  /MODEL revrace revmarcat agesq agecent female revrace*agecent revrace*agesq
  /INTERCEPT INCLUDE=YES SHOW=YES
 /STATISTICS PARAMETER SE CINTERVAL TTEST DEFF
  /PRINT SUMMARY VARIABLEINFO SAMPLEINFO
  /TEST TYPE=F PADJUST=LSD
  /MISSING CLASSMISSING=EXCLUDE
 /CRITERIA CILEVEL=95
/CUSTOM LABEL="RACE TIMES AGE AND AGE SQUARED  "
LMATRIX =
REVRACE*AGECENT 1 0 0 0 -1 ;
REVRACE*AGECENT 1 0 0 -1 0  ;
REVRACE*AGECENT 1 0 -1 0 0 ;
REVRACE*AGECENT 1 -1 0 0 0  ;
REVRACE*AGESQ 1 0 0 0 -1 ;
REVRACE*AGESQ 1 0 0 -1 0  ;
REVRACE*AGESQ 1 0 -1 0 0 ;
REVRACE*AGESQ 1 -1 0 0 0
KMATRIX =0 ; 0 ; 0 ; 0; 0 ; 0 ; 0 ; 0.
```

Complex Samples: General Linear Model

Sample Design Information

		N
Unweighted Cases	Valid	4578
	Invalid	985
	Total	5563
Population Size		1.898E8
Subpopulation Size		1.898E8
Stage 1	Strata	15
	Units	30
Sampling Design Degrees of Freedom		15

a. Subpopulation: age18p = 1

Variable Information

		Mean
Dependent Variable	bpxdi1_1	70.61
Covariates	female	.51
	agec	-.0840
	agecsq	299.4190

Variable Information

		Mean
Dependent Variable	bpxdi1_1	70.61
Covariates	female	.51
	agec	-.0840
	agecsq	299.4190

Subpopulation: age18p = 1

Factor Information

		Weighted Count	Weighted Percent
revrace	1	9915591.033	5.2%
	2	2.047E7	10.8%
	3	1.379E8	72.6%
	4	6407271.118	3.4%
	5	1.519E7	8.0%
revmarcat	1	3.325E7	17.5%
	2	3.351E7	17.7%
	3	1.231E8	64.8%
Subpopulation Size		1.898E8	100.0%

Subpopulation: age18p = 1

Model Summary[a]

R Square	.135

Subpopulation: age18p

= 1

a. Model: bpxdi1_1 =
(Intercept) + revrace
+ revmarcat + female
+ agec + agecsq +
revrace * agec +
revrace * agecsq

Tests of Model Effects[a]

Source	df1	df2	Wald F	Sig.
(Corrected Model)	15.000	1.000	1056.581	.024
(Intercept)	1.000	15.000	17840.669	.000
revrace	4.000	12.000	3.393	.045
revmarcat	2.000	14.000	1.358	.289
female	1.000	15.000	63.287	.000
agec	1.000	15.000	26.904	.000
agecsq	1.000	15.000	127.314	.000
revrace * agec	4.000	12.000	1.856	.183
revrace * agecsq	4.000	12.000	.541	.708

Subpopulation: age18p = 1

a. Model: bpxdi1_1 = (Intercept) + revrace + revmarcat + female

+ agec + agecsq + revrace * agec + revrace * agecsq

Parameter Estimates[b]

Parameter	Estimate	Std. Error	95% Confidence Interval		Hypothesis Test			Design Effect
			Lower	Upper	t	df	Sig.	
(Intercept)	74.220	.466	73.228	75.212	159.419	15.000	.000	.267
[revrace=1.00]	.707	1.180	-1.808	3.222	.599	15.000	.558	.789
[revrace=2.00]	3.022	.917	1.067	4.977	3.295	15.000	.005	.605
[revrace=3.00]	1.424	.567	.216	2.631	2.513	15.000	.024	.380
[revrace=4.00]	.608	1.252	-2.060	3.277	.486	15.000	.634	.616
[revrace=5.00]	.000[a]	.	.	.	.	.	.	.
[revmarcat=1.00]	-.336	.586	-1.584	.913	-.573	15.000	.575	1.220
[revmarcat=2.00]	.990	.625	-.341	2.321	1.585	15.000	.134	1.471
[revmarcat=3.00]	.000[a]	.	.	.	.	.	.	.
female	-2.721	.342	-3.450	-1.992	-7.955	15.000	.000	1.012
agec	.134	.031	.068	.199	4.357	15.000	.001	.525
agecsq	-.014	.001	-.016	-.011	-11.993	15.000	.000	.218
[revrace=1.00] * agec	-.091	.053	-.205	.022	-1.709	15.000	.108	.516
[revrace=2.00] * agec	.041	.037	-.037	.119	1.124	15.000	.279	.451
[revrace=3.00] * agec	-.013	.040	-.098	.071	-.335	15.000	.742	.815
[revrace=4.00] * agec	.067	.078	-.098	.233	.867	15.000	.400	.962
[revrace=5.00] * agec	.000[a]	.	.	.	.	.	.	.
[revrace=1.00] * agecsq	.000	.003	-.006	.006	.070	15.000	.945	.451
[revrace=2.00] * agecsq	.002	.002	-.002	.006	1.172	15.000	.259	.292
[revrace=3.00] * agecsq	.001	.001	-.001	.004	.968	15.000	.348	.211
[revrace=4.00] * agecsq	.004	.003	-.003	.011	1.164	15.000	.262	.562
[revrace=5.00] * agecsq	.000[a]	.	.	.	.	.	.	.

Subpopulation: age18p = 1

a. Set to zero because this parameter is redundant.

b. Model: bpxdi1_1 = (Intercept) + revrace + revmarcat + female + agec + agecsq + revrace * agec + revrace * agecsq

Contrast Coefficients[a]

Parameter	Contrast							
	L1	L2	L3	L4	L5	L6	L7	L8
(Intercept)	.000	.000	.000	.000	.000	.000	.000	.000
[revrace=1.0000]	.000	.000	.000	.000	.000	.000	.000	.000
[revrace=2.0000]	.000	.000	.000	.000	.000	.000	.000	.000
[revrace=3.0000]	.000	.000	.000	.000	.000	.000	.000	.000
[revrace=4.0000]	.000	.000	.000	.000	.000	.000	.000	.000
[revrace=5.0000]	.000	.000	.000	.000	.000	.000	.000	.000
[revmarcat=1.0000]	.000	.000	.000	.000	.000	.000	.000	.000
[revmarcat=2.0000]	.000	.000	.000	.000	.000	.000	.000	.000
[revmarcat=3.0000]	.000	.000	.000	.000	.000	.000	.000	.000
agesq	.000	.000	.000	.000	.000	.000	.000	.000
agecent	.000	.000	.000	.000	.000	.000	.000	.000
female	.000	.000	.000	.000	.000	.000	.000	.000
[revrace=1.0000] * agecent	1.000	1.000	1.000	1.000	.000	.000	.000	.000
[revrace=2.0000] * agecent	.000	.000	.000	-1.000	.000	.000	.000	.000
[revrace=3.0000] * agecent	.000	.000	-1.000	.000	.000	.000	.000	.000
[revrace=4.0000] * agecent	.000	-1.000	.000	.000	.000	.000	.000	.000
[revrace=5.0000] * agecent	-1.000	.000	.000	.000	.000	.000	.000	.000
[revrace=1.0000] * agesq	.000	.000	.000	.000	1.000	1.000	1.000	1.000
[revrace=2.0000] * agesq	.000	.000	.000	.000	.000	.000	.000	-1.000
[revrace=3.0000] * agesq	.000	.000	.000	.000	.000	.000	-1.000	.000
[revrace=4.0000] * agesq	.000	.000	.000	.000	.000	-1.000	.000	.000
[revrace=5.0000] * agesq	.000	.000	.000	.000	-1.000	.000	.000	.000

Subpopulation: age18 = 1

a. The default display of this matrix is the transpose of the corresponding L matrix.

Overall Test Results

df1	df2	Wald F	Sig.
8.000	8.000	.983	.509

Subpopulation: age18 = 1

```
* ANALYSIS EXAMPLE 7.5: ALL PREDICTORS PLUS INTERACTION OF GENDER TIMES AGE AND AGE SQUARED

CSGLM  bpxdi1_1 BY revrace revmarcat WITH agesq agecent female
  /PLAN FILE='F:\applied_analysis_book\SPSS Analysis Examples Replication\Analysis Examples Replication Winter 2010
SPSSv18\csplan_nhanes.csaplan'
  /DOMAIN VARIABLE=age18(1)
  /MODEL revrace revmarcat agesq agecent female female*agecent female*agesq
  /INTERCEPT INCLUDE=YES SHOW=YES
 /STATISTICS PARAMETER SE CINTERVAL TTEST DEFF
  /PRINT SUMMARY VARIABLEINFO SAMPLEINFO
  /TEST TYPE=F PADJUST=LSD
  /MISSING CLASSMISSING=EXCLUDE
 /CRITERIA CILEVEL=95
/CUSTOM LABEL="GENDER TIMES AGE AND AGE SQUARED"
LMATRIX =
       FEMALE*AGECENT 1  ;
       FEMALE*AGESQ 1
       KMATRIX =0 ; 0 .
```

Complex Samples: General Linear Model

Sample Design Information

		N
Unweighted Cases	Valid	4578
	Invalid	985
	Total	5563
Population Size		1.898E8
Subpopulation Size		1.898E8
Stage 1	Strata	15
	Units	30
Sampling Design Degrees of Freedom		15

a. Subpopulation: age18p = 1

Variable Information

		Mean
Dependent Variable	bpxdi1_1	70.61
Covariates	female	.51
	agec	-.0840
	agecsq	299.4190

Subpopulation: age18p = 1

Factor Information

		Weighted Count	Weighted Percent
revrace	1	9915591.033	5.2%
	2	2.047E7	10.8%
	3	1.379E8	72.6%
	4	6407271.118	3.4%
	5	1.519E7	8.0%
revmarcat	1	3.325E7	17.5%
	2	3.351E7	17.7%
	3	1.231E8	64.8%
Subpopulation Size		1.898E8	100.0%

Subpopulation: age18p = 1

Model Summary[a]

R Square	.134

Subpopulation: age18p
= 1

a. Model: bpxdi1_1 =
(Intercept) + revrace
+ revmarcat + female
+ agec + agecsq +
female * agec +
female * agecsq

Tests of Model Effects[a]

Source	df1	df2	Wald F	Sig.
(Corrected Model)	11.000	5.000	95.966	.000
(Intercept)	1.000	15.000	18430.792	.000
revrace	4.000	12.000	5.159	.012
revmarcat	2.000	14.000	1.112	.356
female	1.000	15.000	20.588	.000
agec	1.000	15.000	36.428	.000
agecsq	1.000	15.000	109.548	.000
female * agec	1.000	15.000	.255	.621
female * agecsq	1.000	15.000	1.161	.298

Subpopulation: age18p = 1

a. Model: bpxdi1_1 = (Intercept) + revrace + revmarcat + female
+ agec + agecsq + female * agec + female * agecsq

Parameter Estimates[b]

Parameter	Estimate	Std. Error	95% Confidence Interval		Hypothesis Test			Design Effect
			Lower	Upper	t	df	Sig.	
(Intercept)	74.138	.567	72.929	75.347	130.696	15.000	.000	.793
[revrace=1.00]	1.208	.933	-.781	3.196	1.295	15.000	.215	1.014
[revrace=2.00]	3.492	.777	1.835	5.149	4.492	15.000	.000	.980
[revrace=3.00]	1.796	.632	.450	3.143	2.844	15.000	.012	1.113
[revrace=4.00]	1.201	1.096	-1.135	3.537	1.096	15.000	.291	1.045
[revrace=5.00]	.000[a]	.	.	.	.	.	.	.
[revmarcat=1.00]	-.346	.585	-1.593	.900	-.592	15.000	.563	1.228
[revmarcat=2.00]	.907	.653	-.484	2.298	1.390	15.000	.185	1.593
[revmarcat=3.00]	.000[a]	.	.	.	.	.	.	.
female	-3.237	.713	-4.758	-1.717	-4.537	15.000	.000	2.415
agec	.118	.020	.076	.159	6.036	15.000	.000	1.313
agecsq	-.013	.001	-.016	-.011	-10.466	15.000	.000	2.115
female * agec	.014	.028	-.045	.073	.505	15.000	.621	1.721
female * agecsq	.002	.002	-.002	.005	1.077	15.000	.298	1.993

Subpopulation: age18p = 1

a. Set to zero because this parameter is redundant.

b. Model: bpxdi1_1 = (Intercept) + revrace + revmarcat + female + agec + agecsq + female * agec + female * agecsq

Contrast Coefficients^a

Parameter	Contrast	
	L1	L2
(Intercept)	.000	.000
[revrace=1.0000]	.000	.000
[revrace=2.0000]	.000	.000
[revrace=3.0000]	.000	.000
[revrace=4.0000]	.000	.000
[revrace=5.0000]	.000	.000
[revmarcat=1.0000]	.000	.000
[revmarcat=2.0000]	.000	.000
[revmarcat=3.0000]	.000	.000
agesq	.000	.000
agecent	.000	.000
female	.000	.000
agecent * female	1.000	.000
agesq * female	.000	1.000

Subpopulation: age18 = 1

a. The default display of this matrix
is the transpose of the corresponding L
matrix.

Overall Test Results

df1	df2	Wald F	Sig.
2.000	14.000	1.732	.213

Subpopulation: age18 = 1

*Refit the final model without interactions and do regression diagnostics using saved variables residual and predicted

* Complex Samples General Linear Model.
CSGLM bpxdi1_1 WITH agec agecsq female prevmar nevmar othhis white black other
  /PLAN FILE='F:\applied_analysis_book\SPSS Analysis Examples Replication\Analysis Examples Replication Winter 2010
SPSSv18\csplan_nhanes.csaplan'
  /DOMAIN VARIABLE=age18(1)
  /MODEL agec agecsq female prevmar nevmar othhis white black other
  /INTERCEPT INCLUDE=YES SHOW=YES
  /STATISTICS PARAMETER SE CINTERVAL TTEST DEFF
  /PRINT SUMMARY SAMPLEINFO
  /TEST TYPE=ADJF PADJUST=LSD
  /SAVE PRED RESID
  /MISSING CLASSMISSING=EXCLUDE
  /CRITERIA CILEVEL=95.

Sample Design Information

		N
Unweighted Cases	Valid	5526
	Invalid	4822
	Total	10348
Population Size		2.054E8
Subpopulation Size		1.898E8
Stage 1	Strata	15
	Units	30
Sampling Design Degrees of Freedom		15

a. Subpopulation: age18 = 1

Model Summary[a]

R Square	.134

Subpopulation: age18

= 1

a. Model: bpxdi1_1 =

(Intercept) + agec +

agecsq + female +

prevmar + nevmar +

othhis + white +

black + other

Model Summary[a]

R Square	.134

Subpopulation: age18

= 1

Model Summary[a]

R Square	.134

Subpopulation: age18 = 1

a. Model: bpxdi1_1 = (Intercept) + agec + agecsq + female + prevmar + nevmar + othhis + white + black + other

Tests of Model Effects[a]

Source	df1	df2	Adjusted Wald F	Sig.
(Corrected Model)	5.506	82.593	60.592	.000
(Intercept)	1.000	15.000	26363.793	.000
agec	1.000	15.000	71.462	.000
agecsq	1.000	15.000	266.864	.000
female	1.000	15.000	64.985	.000
prevmar	1.000	15.000	2.801	.115
nevmar	1.000	15.000	.348	.564
othhis	1.000	15.000	1.197	.291
white	1.000	15.000	7.971	.013
black	1.000	15.000	19.774	.000
other	1.000	15.000	1.619	.223

Subpopulation: age18 = 1

a. Model: bpxdi1_1 = (Intercept) + agec + agecsq + female + prevmar + nevmar + othhis + white + black + other
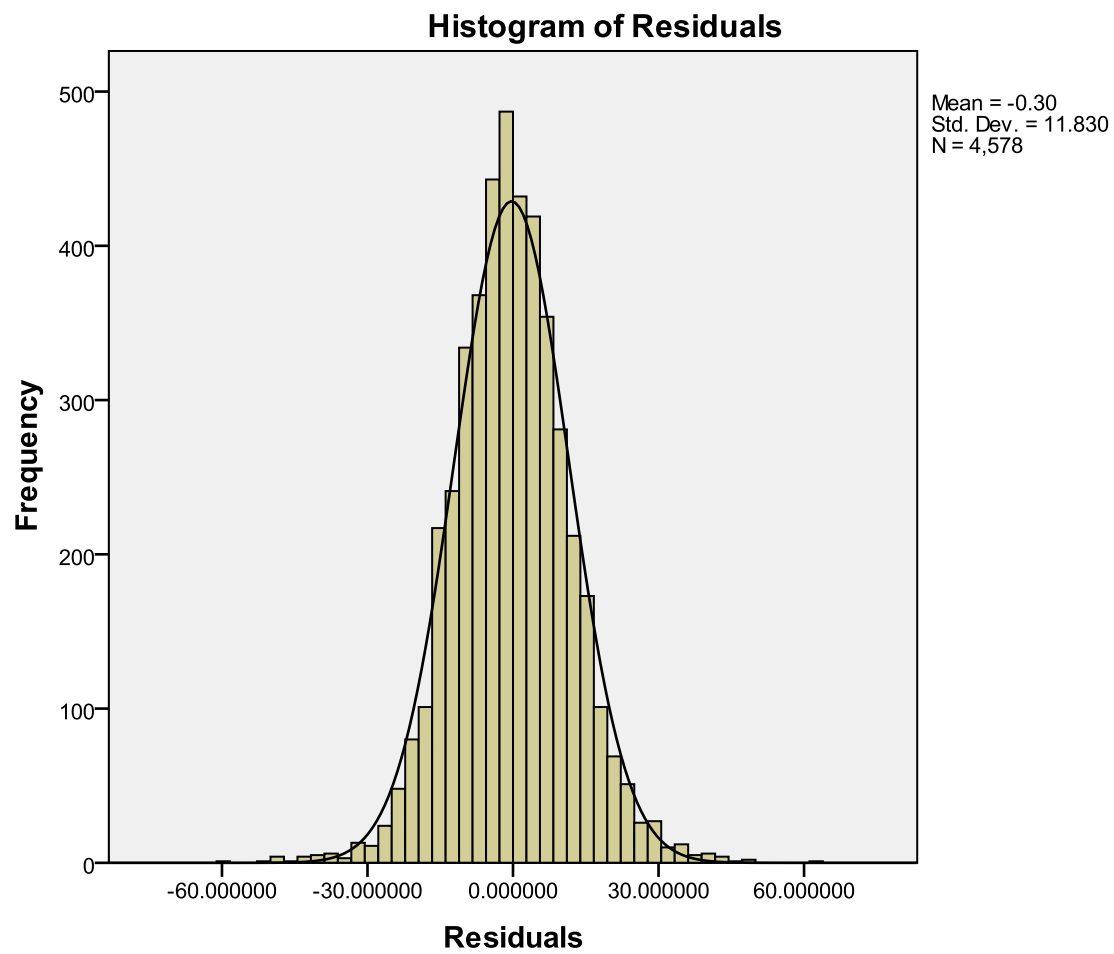
Parameter Estimates[a]

Parameter	Estimate	Std. Error	95% Confidence Interval		Hypothesis Test			Design Effect
			Lower	Upper	t	df	Sig.	
(Intercept)	73.859	.455	72.889	74.829	162.369	15.000	.000	.611
agec	.125	.015	.094	.157	8.454	15.000	.000	1.501
agecsq	-.012	.001	-.014	-.011	-16.336	15.000	.000	1.670
female	-2.721	.338	-3.441	-2.002	-8.061	15.000	.000	1.099
prevmar	1.040	.622	-.285	2.366	1.673	15.000	.115	1.634
nevmar	-.343	.582	-1.583	.897	-.590	15.000	.564	1.353
othhis	1.189	1.087	-1.127	3.505	1.094	15.000	.291	1.152
white	1.781	.631	.436	3.125	2.823	15.000	.013	1.238
black	3.465	.779	1.804	5.126	4.447	15.000	.000	1.096
other	1.189	.934	-.803	3.180	1.272	15.000	.223	1.137

Subpopulation: age18 = 1

a. Model: bpxdi1_1 = (Intercept) + agec + agecsq + female + prevmar + nevmar + othhis + white + black + other

```
GRAPH
  /HISTOGRAM(NORMAL)=Residual
  /TITLE='Histogram of Residuals '.
```


Histogram of Residuals

Mean = -0.30
Std. Dev. = 11.830
N = 4,578

```
GRAPH
  /SCATTERPLOT(BIVAR)=Predicted WITH Residual
  /MISSING=LISTWISE
  /TITLE='Residuals v. Predicted Y'.
```


Residuals v. Predicted Y