

CHAPTER 5 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 DESCRIPTIVE ANALYSES IN SPSS/PASW V18 COMPLEX SAMPLES MODULE

SPSS/PASW DESCRIPTIVE/TABULATE/RATIO commands can perform nearly all of the descriptive analyses presented in Chapter 5 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. The DESCRIPTIVES/TABULATE commands do not include the ability to perform linear contrasts for difference between means but this can be done through the use of the linear regression command CSGLM and are demonstrated for examples 5.12 and 5.13. Finally, complex sample survey corrected quantiles are not included in SPSS v18 therefore not included in this chapter.

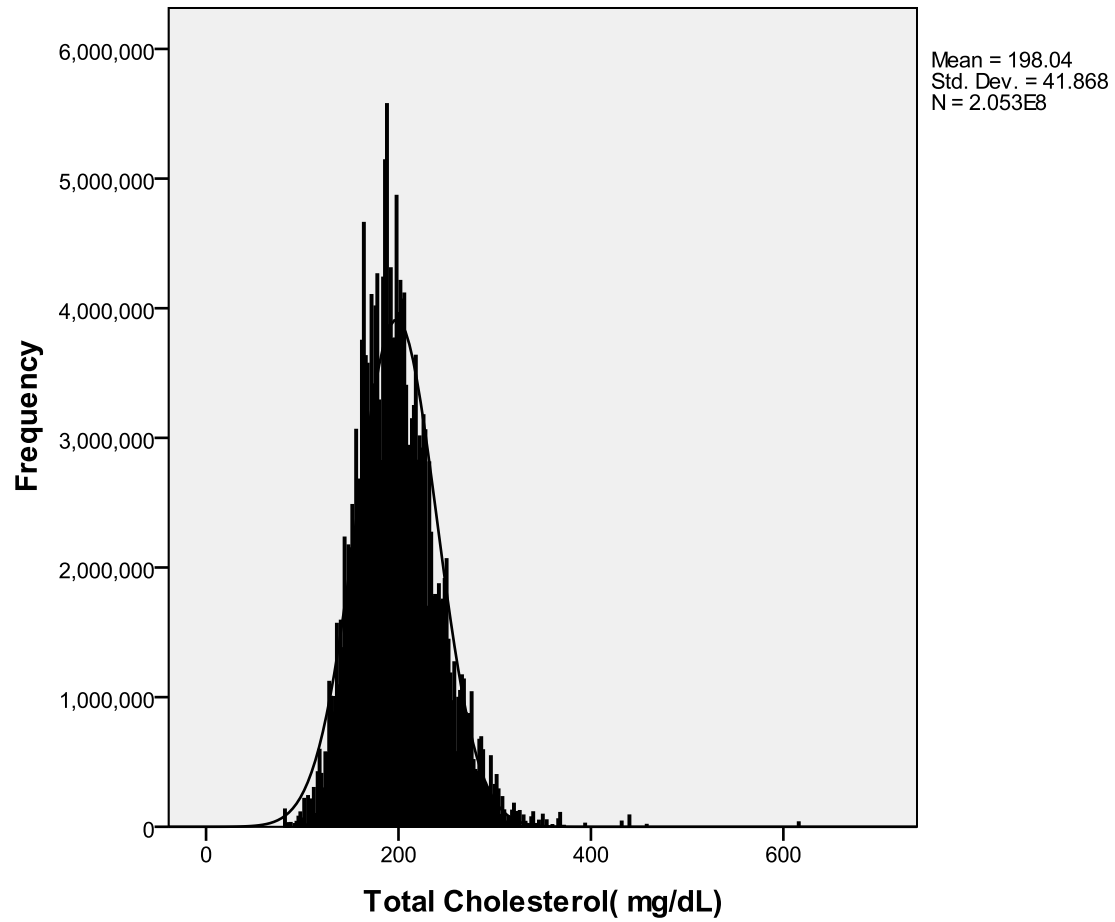
Analysis Example 5.1

Warning # 3211

On at least one case, the value of the weight variable was zero, negative, or missing. Such cases are invisible to statistical procedures and graphs which need positively weighted cases, but remain on the file and are processed by non-statistical facilities such as LIST and SAVE.

GRAPH

/HISTOGRAM(NORMAL)=LBXTC.



Cases weighted by Full Sample 2 Year MEC Exam Weight

*Analysis Example 5.2

Warning # 3211

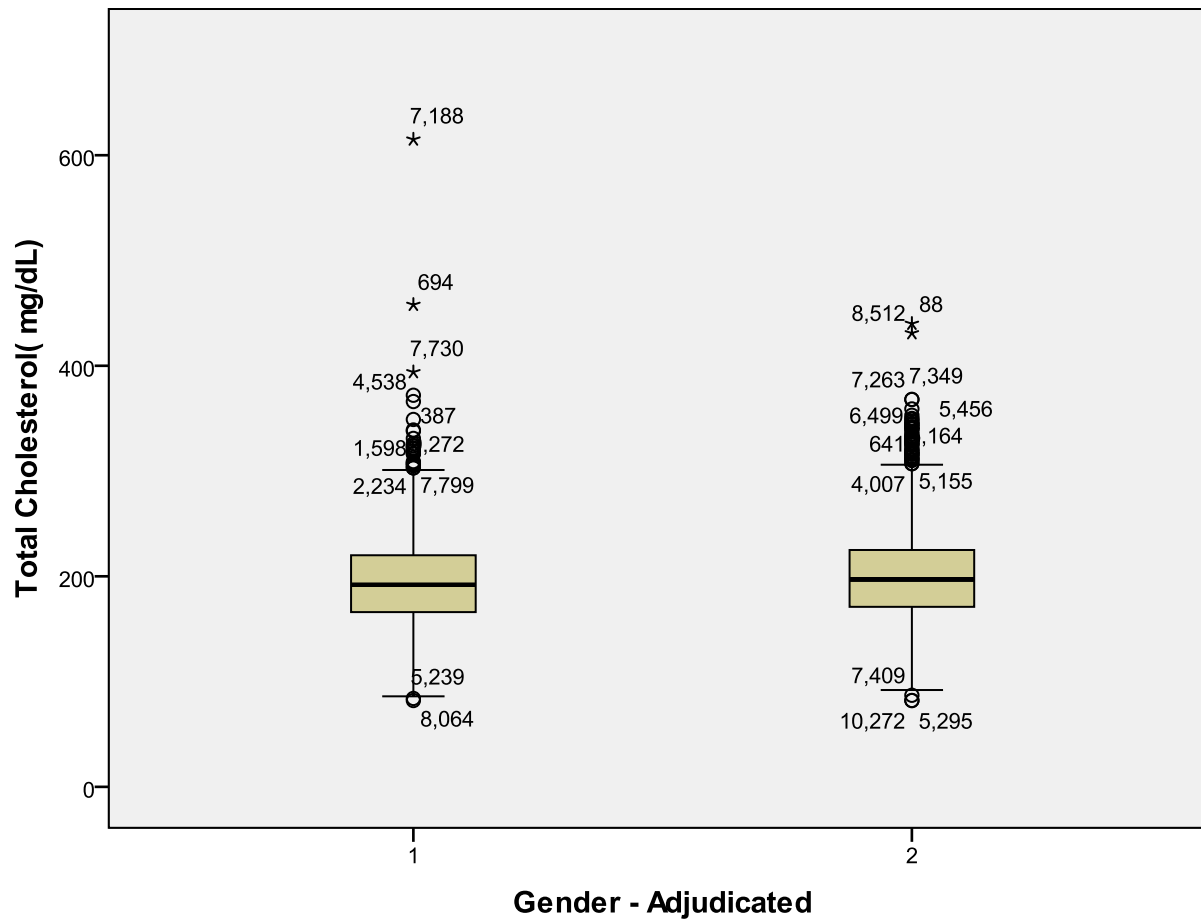
On at least one case, the value of the weight variable was zero, negative, or missing. Such cases are invisible to statistical procedures and graphs which need positively weighted cases, but remain on the file and are processed by non-statistical facilities such as LIST and SAVE.

EXAMINE VARIABLES=LBXTC BY RIAGENDR

/PLOT=BOXPLOT

/STATISTICS=NONE

/NOTOTAL.



Cases weighted by Full Sample 2 Year MEC Exam Weight

```

COMPUTE ncsrwtsh_pop=ncsrwtsh*(209128094/ 9282) .
* Analysis Example 5.3 total count of US Adults with MDE
* Complex Samples Frequencies.
CSTABULATE
  /PLAN FILE='F:\applied_analysis_book\ncsr_p1wt popwgt.csaplan'
  /TABLES VARIABLES=mde
  /CELLS POPSIZE
  /STATISTICS SE CIN(95)
  /MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.

```

Complex Samples: Tables

mde

	Estimate	Standard Error	95% Confidence Interval	
			Lower	Upper
Population Size 0	1.690E8	7876169.959	1.531E8	1.849E8
1	4.009E7	2567487.979	3.491E7	4.527E7
Total	2.091E8	1.022E7	1.885E8	2.298E8

```

* Analysis Example 5.3 Total Count of US Adults with MDE by Marital Status
* Complex Samples Frequencies.

```

```

CSTABULATE
  /PLAN FILE='F:\applied_analysis_book\ncsr_p1wt popwgt.csaplan'
  /TABLES VARIABLES=mde
  /SUBPOP TABLE=MAR3CAT DISPLAY=LAYERED
  /CELLS POPSIZE
  /STATISTICS SE CV CIN(95) DEFF
  /MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.

```

Complex Samples: Tables

mde

	Estimate	Standard Error	95% Confidence Interval		Coefficient of Variation	Design Effect
			Lower	Upper		
Population Size 0	1.690E8	7876169.959	1.531E8	1.849E8	.047	84.958
1	4.009E7	2567487.979	3.491E7	4.527E7	.064	9.028
Total	2.091E8	1.022E7	1.885E8	2.298E8	.049	.

Subpopulation Tables

mde

Marital Status-3 categories	Estimate	Standard Error	95% Confidence Interval		Coefficient of Variation	Design Effect	
			Lower	Upper			
1 Population Size	0	9.646E7	4705215.830	8.696E7	1.060E8	.049	18.907
	1	2.030E7	1584108.641	1.711E7	2.350E7	.078	6.075
	Total	1.168E8	6109331.527	1.044E8	1.291E8	.052	32.121
2 Population Size	0	3.239E7	1895123.123	2.856E7	3.621E7	.059	5.823
	1	1.036E7	702621.506	8942723.044	1.178E7	.068	2.225
	Total	4.275E7	2381848.331	3.794E7	4.755E7	.056	7.403
3 Population Size	0	4.019E7	2944180.685	3.425E7	4.613E7	.073	11.850
	1	9427345.371	773137.582	7867090.562	1.099E7	.082	2.947
	Total	4.962E7	3488233.477	4.258E7	5.666E7	.070	14.269

* Analysis Example 5.4 HRS Data to Estimate Total HH Assets

* Complex Samples Descriptives.

CSDSCRIPTIVES

/PLAN FILE='F:\applied_analysis_book\hrs_hhwgt.csaplan'

/SUMMARY VARIABLES=H8ATOTA

/SUBPOP TABLE=KFINR DISPLAY=LAYERED

/MEAN

/SUM

/STATISTICS SE DEFF CIN(95)

/MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.

Complex Samples: Descriptives

Univariate Statistics

		Estimate	Standard Error	95% Confidence Interval		Design Effect
				Lower	Upper	
Mean	h8atota:w8 total of all assets	591821.4151	32992.39190	525729.7123	657913.1180	2.721
Sum	h8atota:w8 total of all assets	4.87E13	2.86238E12	4.29E13	5.44E13	3.028

Subpopulation Descriptives

Univariate Statistics

2006 whether financial respondent		Estimate	Standard Error	95% Confidence Interval		Design Effect
				Lower	Upper	
1	Mean h8atota:w8 total of all assets	527313.1612	28012.78123	471196.8172	583429.5052	1.526
	Sum h8atota:w8 total of all assets	2.84E13	1.59559E12	2.52E13	3.16E13	1.680
5	Mean h8atota:w8 total of all assets	714161.1541	44891.86675	624231.9387	804090.3695	1.341
	Sum h8atota:w8 total of all assets	2.03E13	1.32310E12	1.76E13	2.29E13	1.395

*Analysis Example 5.5 Mean HH Income NCS-R Data

CSDESCRIPTIVES

/PLAN FILE='F:\applied_analysis_book\ncsr_p2wt.csaplan'

/SUMMARY VARIABLES=hhinc

/MEAN

/STATISTICS SE DEFF CIN(95)

/MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.

Complex Samples: Descriptives

Univariate Statistics

		Estimate	Standard Error	95% Confidence Interval		Design Effect
				Lower	Upper	
Mean	Household Income : Topcode	59277.0583	1596.34297	56055.5078	62498.6089	7.256E7

* ANALYSIS EXAMPLE 5.6 Mean Systolic Blood Pressure in US Adults using the NHANES Data
 * Complex Samples Frequencies.

CSDESCRIPTIVES

/PLAN FILE='F:\applied_analysis_book\csplan_nhanes.csplan'

/SUMMARY VARIABLES=BPXSY1

/SUBPOP TABLE=age18 DISPLAY=LAYERED

/MEAN

/STATISTICS SE DEFF CIN(95)

/MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.

Complex Samples: Descriptives

Univariate Statistics

	Estimate	Standard Error	95% Confidence Interval		Design Effect
			Lower	Upper	
Mean Systolic: Blood pres (1st rdg) mm Hg	123.11	.542	121.96	124.27	3.944

Subpopulation Descriptives

Univariate Statistics

age18	Estimate	Standard Error	95% Confidence Interval		Design Effect
			Lower	Upper	
1 Mean Systolic: Blood pres (1st rdg) mm Hg	123.11	.542	121.96	124.27	3.944

* ANALYSIS EXAMPLE 5.7 Mean HH Assets using the HRS Data
 CSDESCRIPTIVES
 /PLAN FILE='F:\applied_analysis_book\hrs_hhwgt.csaplan'
 /SUMMARY VARIABLES=H8ATOTA
 /SUBPOP TABLE=kfinr display=layered
 /MEAN
 /STATISTICS SE DEFF CIN(95)
 /MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.

Complex Samples: Descriptives

Univariate Statistics

		Estimate	Standard Error	95% Confidence Interval		Design Effect
				Lower	Upper	
Mean	h8atota:w8 total of all assets	591821.4151	32992.39190	525729.7123	657913.1180	2.721

Subpopulation Descriptives

Univariate Statistics

		Estimate	Standard Error	95% Confidence Interval		Design Effect
				Lower	Upper	
2006 whether financial respondent						
1	Mean h8atota:w8 total of all assets	527313.1612	28012.78123	471196.8172	583429.5052	1.526
5	Mean h8atota:w8 total of all assets	714161.1541	44891.86675	624231.9387	804090.3695	1.341

* ANALYSIS EXAMPLE 5.8 QUANTILES NOT AVAILABLE IN SPSS

* ANALYSIS EXAMPLE 5.9 RATIO of HDL to Total Cholesterol in US Adult Population using NHANES Data
 * Complex Samples Ratios.

CSDESCRIPTIVES

```

/PLAN FILE='F:\applied_analysis_book\csplan_nhanes.csplan'
/RATIO NUMERATOR=lbhdhd DENOMINATOR=lbxtc
/subpop table=age18 display=layered
/STATISTICS SE DEFF CIN(95)
/MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.
  
```

Complex Samples: Descriptives

Ratios 1

Numerator	Denominator	Ratio Estimate	Standard Error	95% Confidence Interval		Design Effect
				Lower	Upper	
Direct HDL-Cholesterol (mg/dL)	Total Cholesterol(mg/dL)	.284	.002	.280	.288	3.493

Subpopulation Descriptives

Ratios 1

age18	Numerator	Denominator	Ratio Estimate	Standard Error	95% Confidence Interval		Design Effect
					Lower	Upper	
0	Direct HDL-Cholesterol (mg/dL)	Total Cholesterol(mg/dL)	.336	.002	.330	.341	1.026
1	Direct HDL-Cholesterol (mg/dL)	Total Cholesterol(mg/dL)	.275	.002	.271	.280	3.646

* ANALYSIS EXAMPLE 5.10 Diabetes Proportions by Gender and Age > 70 Years HRS Data
 compute age71 = kage > 70.

* Complex Samples Descriptives.

CSDSCRIPTIVES

```

/PLAN FILE='F:\applied_analysis_book\hrs.csaplan'
/SUMMARY VARIABLES=diabetes
/SUBPOP TABLE=age71 BY GENDER DISPLAY=LAYERED
/MEAN
/STATISTICS SE DEFF CIN(95)
/MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.
  
```

Complex Samples: Descriptives

Univariate Statistics

		Estimate	Standard Error	95% Confidence Interval		Design Effect
				Lower	Upper	
Mean	DIABETES	.18	.004	.18	.19	1.513

Subpopulation Descriptives

Univariate Statistics

age71		gender		Estimate	Standard Error	95% Confidence Interval		Design Effect
						Lower	Upper	
0	1	Mean	DIABETES	.18	.007	.17	.19	1.595
	2	Mean	DIABETES	.17	.006	.15	.18	1.533
1	1	Mean	DIABETES	.24	.008	.22	.25	.879
	2	Mean	DIABETES	.18	.009	.17	.20	1.580

```

COMPUTE age45 = age > 45.
* Complex Samples Descriptives.
CSD DESCRIPTIVES
/PLAN FILE='F:\applied_analysis_book\csplan_nhanes.csaplan'
/SUMMARY VARIABLES=BPXSY1
/SUBPOP TABLE=female BY age45 DISPLAY=LAYERED
/MEAN
/STATISTICS SE DEFF CIN(95)
/MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.

```

Complex Samples: Descriptives

Univariate Statistics

	Estimate	Standard Error	95% Confidence Interval		Design Effect
			Lower	Upper	
Mean Systolic: Blood pres (1st rdg) mm Hg	120.66	.491	119.61	121.70	4.740

Subpopulation Descriptives

Univariate Statistics

female	age45	Estimate	Standard Error	95% Confidence Interval		Design Effect	
				Lower	Upper		
0	0	Mean Systolic: Blood pres (1st rdg) mm Hg	117.56	.553	116.39	118.74	3.607
	1	Mean Systolic: Blood pres (1st rdg) mm Hg	128.96	.757	127.35	130.58	2.145
1	0	Mean Systolic: Blood pres (1st rdg) mm Hg	110.84	.392	110.00	111.67	2.110
	1	Mean Systolic: Blood pres (1st rdg) mm Hg	132.09	1.065	129.82	134.36	2.984

*NOTE: LINEAR DIFFERENCES NOT AVAILABLE IN SPSS/PASW V18 DESCRIPTIVE COMMANDS BUT DEMONSTRATED HERE USING CSGLM INSTEAD, see CSGLM FOR DETAILS.

*EXAMPLE 5.12 HRS DATA

* Complex Samples General Linear Model.

CSGLM H8ATOTA BY EDCAT

```
/PLAN FILE='F:\applied_analysis_book\SPSS Analysis Examples Replication\Analysis Examples Replication Winter 2010 SPSSv18\hrs_hhwgt.csaplan'  
/DOMAIN VARIABLE=KFINR(1)  
/MODEL EDCAT  
/INTERCEPT INCLUDE=YES SHOW=YES  
/PRINT SUMMARY VARIABLEINFO SAMPLEINFO  
/TEST TYPE=F PADJUST=LSD  
/EMMEANS TABLES=EDCAT COMPARE CONTRAST=SIMPLE(4)  
/MISSING CLASSMISSING=EXCLUDE  
/CRITERIA CILEVEL=95.
```

Sample Design Information

		N
Unweighted Cases	Valid	18360
	Invalid	107
	Total	18467
Population Size		8.166E7
Subpopulation Size		5.365E7
Stage 1	Strata	56
	Units	112
Sampling Design Degrees of Freedom		56

a. Subpopulation: 2006 whether financial respondent = 1

Variable Information

		Mean
Dependent Variable	h8atota:w8 total of all assets	5.2874E5

Subpopulation: 2006 whether financial respondent = 1

Factor Information

		Weighted Count	Weighted Percent
EDCAT	1	1.042E7	19.4%
	2	1.722E7	32.1%
	3	1.219E7	22.7%
	4	1.382E7	25.8%
Subpopulation Size		5.365E7	100.0%

Subpopulation: 2006 whether financial respondent = 1

Model Summary^a

R Square	.021
----------	------

Subpopulation: 2006

whether financial

respondent = 1

a. Model: h8atota:w8

total of all assets =

(Intercept) + EDCAT

Tests of Model Effects^a

Source	df1	df2	Wald F	Sig.
(Corrected Model)	3.000	54.000	30.790	.000
(Intercept)	1.000	56.000	376.396	.000
EDCAT	3.000	54.000	30.790	.000

Subpopulation: 2006 whether financial respondent = 1

a. Model: h8atota:w8 total of all assets = (Intercept) + EDCAT

Estimated Marginal Means: EDCAT

Estimates

EDCAT	Mean	Std. Error	95% Confidence Interval	
			Lower	Upper
d 1	1.7839E5	2.45611E4	1.2918E5	2.2759E5
i 2	3.2839E5	1.70827E4	2.9417E5	3.6261E5
m 3	4.5546E5	2.70003E4	4.0137E5	5.0955E5
e 4	1.1072E6	1.02114E5	9.0265E5	1.3118E6
n				
s				
i				
o				
n				
0				

Subpopulation: 2006 whether financial respondent = 1

Note: codes for Edcat 1=0-11 2=12 3=13-15 4=16+ Years of Education.

Individual Test Results

EDCAT Simple Contrast ^a	Contrast Estimate	Hypothesized Value	Difference (Estimate - Hypothesized)	Std. Error	df1	df2	Wald F	Sig.
Level 1 vs. Level 4	-928818.049	.000	-928818.049	1.083E5	1.000	56.000	73.622	.000
Level 2 vs. Level 4	-778812.149	.000	-778812.149	1.024E5	1.000	56.000	57.893	.000
Level 3 vs. Level 4	-651746.496	.000	-651746.496	1.056E5	1.000	56.000	38.122	.000

Subpopulation: 2006 whether financial respondent = 1

a. Reference Category = 4

Overall Test Results

df1	df2	Wald F	Sig.
3.000	54.000	30.790	.000

Subpopulation: 2006 whether financial respondent = 1

*EXAMPLE 5.13 HRS DATA FOR 2004 AND 2006

* Complex Samples General Linear Model.

CSGLM totassets BY year

```

/PLAN FILE='F:\applied_analysis_book\SPSS Analysis Examples Replication\Analysis Examples Replication Winter 2010 SPSSv18\hrs_hhwgt_0406.csaplan'
/DOMAIN VARIABLE=finr0406(1)
/MODEL year
/INTERCEPT INCLUDE=YES SHOW=YES
/STATISTICS PARAMETER SE CINTERVAL
/PRINT VARIABLEINFO
/TEST TYPE=F PADJUST=LSD
/EMMEANS TABLES=year COMPARE CONTRAST=SIMPLE(2004)
/EMMEANS
/MISSING CLASSMISSING=EXCLUDE
/CRITERIA CILEVEL=95.

```

Variable Information

		Mean
Dependent Variable	totassets	4.7099E5

Subpopulation: finr0406 = 1

Factor Information

		Weighted Count	Weighted Percent
year	2004	5.123E7	48.8%
	2006	5.385E7	51.2%
Subpopulation Size		1.051E8	100.0%

Subpopulation: finr0406 = 1

Tests of Model Effects^a

Source	df1	df2	Wald F	Sig.
(Corrected Model)	1.000	56.000	33.281	.000
(Intercept)	1.000	56.000	436.499	.000
year	1.000	56.000	33.281	.000

Subpopulation: finr0406 = 1

a. Model: totassets = (Intercept) + year

Parameter Estimates^b

Parameter	Estimate	Std. Error	95% Confidence Interval	
			Lower	Upper
(Intercept)	5.273E5	2.801E4	4.712E5	5.834E5
[year=2004]	-1.155E5	2.003E4	-1.556E5	-7.541E4
[year=2006]	.000 ^a	.	.	.

Subpopulation: finr0406 = 1

a. Set to zero because this parameter is redundant.

b. Model: totassets = (Intercept) + year

Estimated Marginal Means 1: year

Estimates

year	Mean	Std. Error	95% Confidence Interval	
			Lower	Upper
2004	4.1179E5	2.06397E4	3.7044E5	4.5313E5
2006	5.2731E5	2.80128E4	4.7120E5	5.8343E5

Subpopulation: finr0406 = 1

Individual Test Results

year Simple Contrast ^a	Contrast Estimate	Hypothesized Value	Difference (Estimate - Hypothesized)	Std. Error	df1	df2	Wald F	Sig.
Level 2006 vs. Level 2004	115526.777	.000	115526.777	2.003E4	1.000	56.000	33.281	.000

Subpopulation: finr0406 = 1

a. Reference Category = 2004

Overall Test Results

df1	df2	Wald F	Sig.
1.000	56.000	33.281	.000

Subpopulation: finr0406 = 1