CHAPTER 6 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 commands can perform nearly all of the categorical data analysis examples presented in Chapter 6 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.

A special note: for analyses using CSTabulate for two way crosstabulations with a test of independence of row/columns and a subpopulation indicator, SPSS requires that the analyst declare the subpopulation variable in the CSPlan file (in the strata specification). This is demonstrated in Example 6.8 where we provide syntax for the adjusted CSPlan file and then the CSTabulate syntax too. Please refer to this example and the SPSS documentation for details.

```
SPSS Analysis Examples Replication Chapter 6
* ANALYSIS EXAMPLE 6.1 PROPORTIONS OF IRREGULAR HEART BEAT US ADULT POPULATION NHANES DATA
* Complex Samples Frequencies.
CSTABULATE
  /PLAN FILE='F:\applied_analysis_book\csplan_nhanes.csaplan'
  /TABLES VARIABLES=irregular
  /SUBPOP TABLE=age18p DISPLAY=LAYERED
  /CELLS TABLEPCT
  /STATISTICS SE CIN(95) DEFF
  /MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

1=yes 0=no

		Estimate	Standard Error	95% Confidence Interval		Design Effect
				Lower	Upper	
% of Total	0	97.5%	.6%	96.0%	98.5%	11.518
	1	2.5%	.6%	1.5%	4.0%	11.518
	Total	100.0%	.0%	100.0%	100.0%	.

1=yes 0=no

age18p			Estimate	Standard Error	95% Confidence Interval		Design Effect
					Lower	Upper	
0	% of Total	0	99.1%	.3%	98.3%	99.6%	1.827
		1	.9%	.3%	.4%	1.7%	1.827
		Total	100.0%	.0%	100.0%	100.0%	.
1	% of Total	0	97.0%	.7%	95.2%	98.2%	10.352
		1	3.0%	.7%	1.8%	4.8%	10.352
		Total	100.0%	.0%	100.0%	100.0%	.

```
* Complex Samples Descriptives.
CSDESCRIPTIVES
  /PLAN FILE='F:\applied_analysis_book\SPSS Analysis Examples Replication\Analysis Examples Replication Winter 2010 SPSSv18\csplan_nhanes.csaplan'
  /SUMMARY VARIABLES=irregular
  /SUBPOP TABLE=age18 DISPLAY=LAYERED
  /MEAN
  /STATISTICS SE DEFF CIN(95)
  /MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.
```

Univariate Statistics

		Estimate	Standard Error	95% Confidence Interval		Design Effect
				Lower	Upper	
Mean	1=yes 0=no	.03	.007	.02	.04	7.912

Univariate Statistics

age18			Estimate	Standard Error	95% Confidence Interval		Design Effect
					Lower	Upper	
1	Mean	1=yes 0=no	.03	.007	.02	.04	7.912

```
* ANALYSIS EXAMPLE 6.2 RACE/ETHNICITY PROPORTIONS US ADULT POPULATION NHANES DATA
* Complex Samples Frequencies.
CSTABULATE
  /PLAN FILE='F:\applied_analysis_book\csplan_nhanes.csaplan'
  /TABLES VARIABLES=RIDRETH1
  /SUBPOP TABLE=age18p DISPLAY=LAYERED
  /CELLS TABLEPCT
  /STATISTICS SE CIN(95) DEFF
  /MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

1=mex 2=oth hisp 3=white 4=black 5=other

		Estimate	Standard Error	95% Confidence Interval		Design Effect
				Lower	Upper	
% of Total	1	9.4%	1.1%	7.3%	11.9%	12.274
	2	3.7%	.8%	2.3%	5.8%	15.943
	3	68.7%	2.9%	62.2%	74.5%	35.625
	4	12.3%	2.0%	8.7%	17.3%	34.000
	5	5.9%	.7%	4.6%	7.6%	7.686
	Total	100.0%	.0%	100.0%	100.0%	.

1=mex 2=oth hisp 3=white 4=black 5=other

age18p			Estimate	Standard Error	95% Confidence Interval		Design Effect
					Lower	Upper	
0	% of Total	1	13.7%	1.4%	10.9%	17.0%	3.609
		2	4.6%	1.1%	2.8%	7.6%	5.714
		3	59.6%	3.5%	52.0%	66.8%	10.653
		4	14.4%	2.2%	10.3%	19.8%	8.403
		5	7.7%	1.3%	5.4%	11.0%	4.976
		Total	100.0%	.0%	100.0%	100.0%	.
1	% of Total	1	8.1%	1.0%	6.2%	10.5%	9.523
		2	3.4%	.7%	2.1%	5.4%	11.805
		3	71.4%	2.8%	65.2%	76.9%	26.294
		4	11.7%	2.0%	8.1%	16.7%	26.631
		5	5.4%	.6%	4.3%	6.8%	4.645
		Total	100.0%	.0%	100.0%	100.0%	.

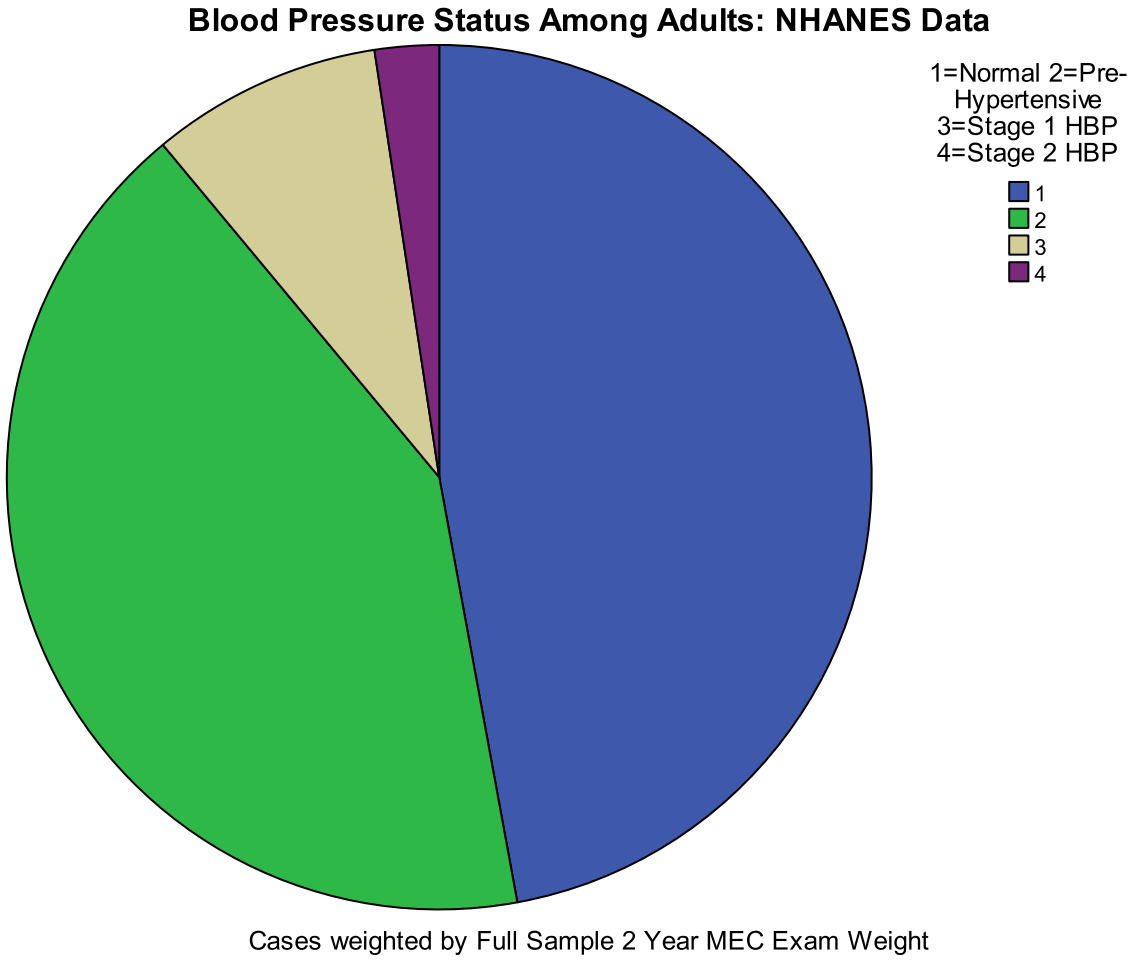
```
* ANALYSIS EXAMPLE 6.3 BLOOD PRESSURE CATEGORY IN THE US ADULT POPULATION: NHANES DATA
* Complex Samples Frequencies.
CSTABULATE
  /PLAN FILE='F:\applied_analysis_book\csplan_nhanes.csaplan'
  /TABLES VARIABLES=BP_CAT
  /SUBPOP TABLE=AGE18P DISPLAY=LAYERED
  /CELLS TABLEPCT
  /STATISTICS SE CIN(95) DEFF
  /MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

1=Normal 2=Pre-Hypertensive 3=Stage 1 HBP 4=Stage 2 HBP

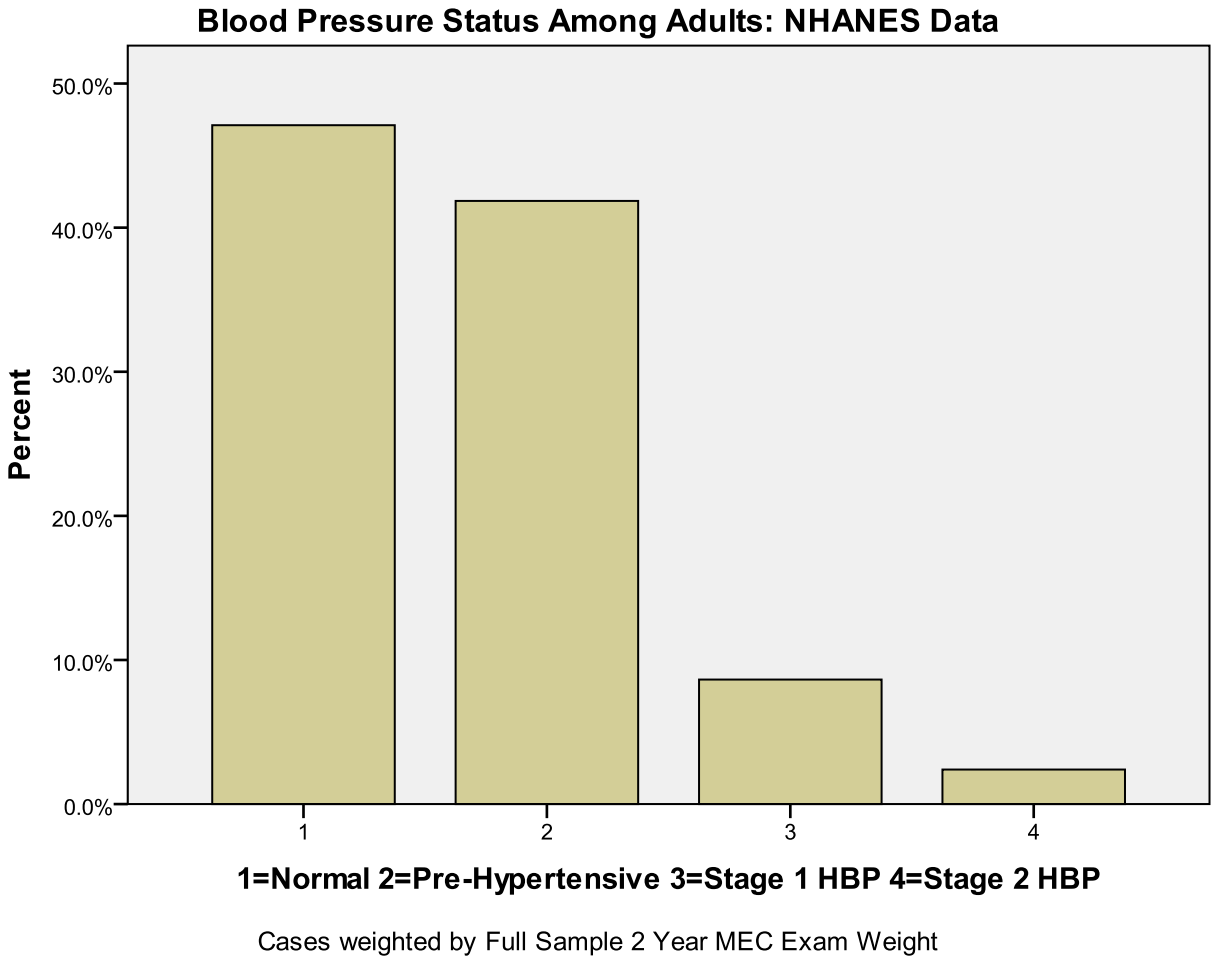
age18			Estimate	Standard Error	95% Confidence Interval		Design Effect
					Lower	Upper	
0	% of Total	1	89.1%	1.2%	86.2%	91.5%	1.881
		2	10.7%	1.2%	8.4%	13.4%	1.690
		3	.2%	.1%	.1%	.7%	.713
		Total	100.0%	.0%	100.0%	100.0%	.
1	% of Total	1	47.1%	1.1%	44.8%	49.5%	3.037
		2	41.9%	1.2%	39.4%	44.4%	3.515
		3	8.6%	.6%	7.4%	10.1%	3.006
		4	2.4%	.2%	1.9%	3.0%	1.522
		Total	100.0%	.0%	100.0%	100.0%	.

```
*ANALYSIS EXAMPLE 6.4 GOODNESS OF FIT IS NOT AVAILABLE IN SPSS/PASW V18
```

```
* ANALYSIS EXAMPLE 6.5 PIE CHART OF BLOOD PRESSURE CATEGORY IN THE US ADULT POPULATION: NHANES DATA


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
  /PIE=PCT BY bp_cat
  /TITLE='Blood Pressure Status Among Adults: NHANES Data'.
```



```
* ANALYSIS EXAMPLE 6.5 BAR CHART OF BLOOD PRESSURE CATEGORY IN THE US ADULT POPULATION: NHANES DATA


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
  /BAR(SIMPLE)=PCT BY bp_cat
  /TITLE='Blood Pressure Status Among Adults: NHANES Data '.
```



```
* Complex Samples Crosstabs.
CSTABULATE
  /PLAN FILE='F:\applied_analysis_book\SPSS Analysis Examples Replication\Analysis Examples Replication Winter 2010 SPSSv18\ncsr_p1wt.csaplan'
  /TABLES VARIABLES=SEX BY mde
  /CELLS ROWPCT TABLEPCT
  /STATISTICS SE CIN(95) DEFF
  /TEST INDEPENDENCE
  /MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

NOTE: CODES FOR SEX 1=MALE 2=FEMALE, CODES FOR MDE 0=NO 1=YES.

Sex * mde

Sex				mde 0	mde 1	mde Total
1	% within Sex	Estimate		84.9%	15.1%	100.0%
		Standard Error		.8%	.8%	.0%
		95% Confidence Interval	Lower	83.3%	13.6%	100.0%
			Upper	86.4%	16.7%	100.0%
		Design Effect		2.083	2.083	.
	% of Total	Estimate		40.7%	7.2%	47.9%
		Standard Error		.7%	.3%	.5%
		95% Confidence Interval	Lower	39.3%	6.6%	46.8%
			Upper	42.1%	7.9%	49.0%
		Design Effect		1.874	1.637	1.051

2	% within Sex	Estimate	77.1%	22.9%	100.0%
		Standard Error	.6%	.6%	.0%
		95% Confidence Interval Lower	75.9%	21.8%	100.0%
		Upper	78.2%	24.1%	100.0%
		Design Effect	.873	.873	.
	% of Total	Estimate	40.2%	12.0%	52.1%
		Standard Error	.5%	.3%	.5%
		95% Confidence Interval Lower	39.1%	11.4%	51.0%
		Upper	41.3%	12.6%	53.2%
		Design Effect	1.110	.809	1.051
Total	% within Sex	Estimate	80.8%	19.2%	100.0%
		Standard Error	.5%	.5%	.0%
		95% Confidence Interval Lower	79.8%	18.2%	100.0%
		Upper	81.8%	20.2%	100.0%
		Design Effect	1.424	1.424	.
	% of Total	Estimate	80.8%	19.2%	100.0%
		Standard Error	.5%	.5%	.0%
		95% Confidence Interval Lower	79.8%	18.2%	100.0%
		Upper	81.8%	20.2%	100.0%
		Design Effect	1.424	1.424	.

Tests of Independence

		Chi-Square	Adjusted F	df1	df2	Sig.
Sex * mde	Pearson	92.150	57.978	1	42	.000
	Likelihood Ratio	93.106	58.580	1	42	.000

The adjusted F is a variant of the second-order Rao-Scott adjusted chi-square statistic. Significance is based on the adjusted F and its degrees of freedom.

```
*ANALYSIS EXAMPLE 6.7 COMPARING PROPORTIONS OF US ADULTS SEX BY MDE NCSR DATA
* Complex Samples Crosstabs.
CSTABULATE
  /PLAN FILE='F:\applied_analysis_book\ncsr_p1wt.csaplan'
  /TABLES VARIABLES=SEX BY mde
  /CELLS ROWPCT TABLEPCT
  /STATISTICS SE CIN(95) DEFF
  /TEST ODDSRATIO RELRISK RISKDIFF INDEPENDENCE
  /MISSING SCOPE=LISTWISE CLASSMISSING=EXCLUDE.
```

NOTE: CODES FOR SEX 1=MALE 2=FEMALE, CODES FOR MDE 0=NO 1=YES.

Sex * mde

				mde	
Sex			0	1	Total
1	% within Sex	Estimate	84.9%	15.1%	100.0%
		Standard Error	.8%	.8%	.0%
		95% Confidence Interval Lower	83.3%	13.6%	100.0%
		Upper	86.4%	16.7%	100.0%
		Design Effect	2.083	2.083	.
	% of Total	Estimate	40.7%	7.2%	47.9%
		Standard Error	.7%	.3%	.5%
		95% Confidence Interval Lower	39.3%	6.6%	46.8%
		Upper	42.1%	7.9%	49.0%
		Design Effect	1.874	1.637	1.051

2	% within Sex	Estimate	77.1%	22.9%	100.0%
		Standard Error	.6%	.6%	.0%
		95% Confidence Interval Lower	75.9%	21.8%	100.0%
		Upper	78.2%	24.1%	100.0%
		Design Effect	.873	.873	.
	% of Total	Estimate	40.2%	12.0%	52.1%
		Standard Error	.5%	.3%	.5%
		95% Confidence Interval Lower	39.1%	11.4%	51.0%
		Upper	41.3%	12.6%	53.2%
		Design Effect	1.110	.809	1.051
Total	% within Sex	Estimate	80.8%	19.2%	100.0%
		Standard Error	.5%	.5%	.0%
		95% Confidence Interval Lower	79.8%	18.2%	100.0%
		Upper	81.8%	20.2%	100.0%
		Design Effect	1.424	1.424	.
	% of Total	Estimate	80.8%	19.2%	100.0%
		Standard Error	.5%	.5%	.0%
		95% Confidence Interval Lower	79.8%	18.2%	100.0%
		Upper	81.8%	20.2%	100.0%
		Design Effect	1.424	1.424	.

Tests of Independence

		Chi-Square	Adjusted F	df1	df2	Sig.
Sex * mde	Pearson	92.150	57.978	1	42	.000
	Likelihood Ratio	93.106	58.580	1	42	.000

The adjusted F is a variant of the second-order Rao-Scott adjusted chi-square statistic.

Significance is based on the adjusted F and its degrees of freedom.

Measures of Association

			Estimate	95% Confidence Interval	
				Lower	Upper
Sex * mde	Odds Ratio		1.676	1.460	1.923
	Relative Risk	For cohort mde = 0	1.102	1.076	1.128
		For cohort mde = 1	.658	.586	.737
	Risk Difference	For cohort mde = 0	.079	.059	.098
		For cohort mde = 1	-.079	-.098	-.059

Statistics are computed only for 2-by-2 tables with all cells observed.

```
*ANALYSIS EXAMPLE 6.8 EDUCATION BY ALCOHOL DEPENDENCE IN THE SUBPOPULATION OF THOSE 18-28: NCSR DATA


* Analysis Preparation Wizard.
CSPLAN ANALYSIS
  /PLAN FILE='F:\brahms\applied_analysis_book\SPSS Analysis Examples Replication\Analysis Examples Replication Winter 2010
SPSSv18\ncsr_p2wt_subpop29_cluster.csaplan'
  /PLANVARS ANALYSISWEIGHT=NCSRWTLG
  /SRSESTIMATOR TYPE=WOR
  /PRINT PLAN
  /DESIGN STRATA=SESTRAT CLUSTER=SECLUSTR age29
  /ESTIMATOR TYPE=WR.

* Complex Samples Crosstabs.
CSTABULATE
  /PLAN FILE='F:\brahms\applied_analysis_book\SPSS Analysis Examples Replication\Analysis Examples Replication Winter 2010
SPSSv18\ncsr_p2wt_subpop29.csaplan'
  /TABLES VARIABLES=ED4CAT BY ald
  /SUBPOP TABLE=age29 DISPLAY=LAYERED
  /CELLS ROWPCT
  /STATISTICS SE
  /TEST INDEPENDENCE
  /MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

Years of education-4 categories * ald

Years of education-4 categories			ald		
			0	1	Total
1	% within Years of education-4 categories	Estimate	93.3%	6.7%	100.0%
		Standard Error	.9%	.9%	.0%
2	% within Years of education-4 categories	Estimate	94.4%	5.6%	100.0%
		Standard Error	.6%	.6%	.0%
3	% within Years of education-4 categories	Estimate	94.2%	5.8%	100.0%
		Standard Error	.5%	.5%	.0%
4	% within Years of education-4 categories	Estimate	96.3%	3.7%	100.0%
		Standard Error	.5%	.5%	.0%
Total	% within Years of education-4 categories	Estimate	94.6%	5.4%	100.0%
		Standard Error	.3%	.3%	.0%

Tests of Independence

		Chi-Square	Adjusted F	df1	df2	Sig.
Years of education-4 categories * ald	Pearson	10.793	3.757	2.520	209.145	.017
	Likelihood Ratio	11.401	3.968	2.520	209.145	.013

The adjusted F is a variant of the second-order Rao-Scott adjusted chi-square statistic.

Significance is based on the adjusted F and its degrees of freedom.

Years of education-4 categories 1=0-11 2=12 3=13-15 4=16+ * ald

age29	Years of education-4 categories 1=0-11 2=12 3=13-15 4=16+			ald		
				0	1	Total
0	1	% within Years of education-4 categories 1=0-11 2=12 3=13-15 4=16+	Estimate	94.0%	6.0%	100.0%
			Standard Error	.8%	.8%	.0%
			95% Confidence Interval Lower	92.2%	4.5%	100.0%
			95% Confidence Interval Upper	95.5%	7.8%	100.0%
			Design Effect	9.912E6	9.912E6	.
	2	% within Years of education-4 categories 1=0-11 2=12 3=13-15 4=16+	Estimate	94.2%	5.8%	100.0%
			Standard Error	.6%	.6%	.0%
			95% Confidence Interval Lower	92.8%	4.7%	100.0%
			95% Confidence Interval Upper	95.3%	7.2%	100.0%
			Design Effect	1.224E7	1.224E7	.
	3	% within Years of education-4 categories 1=0-11 2=12 3=13-15 4=16+	Estimate	93.9%	6.1%	100.0%
			Standard Error	.5%	.5%	.0%
			95% Confidence Interval Lower	92.7%	5.1%	100.0%
			95% Confidence Interval Upper	94.9%	7.3%	100.0%
			Design Effect	6.471E6	6.471E6	.
	4	% within Years of education-4 categories 1=0-11 2=12 3=13-15 4=16+	Estimate	96.8%	3.2%	100.0%
			Standard Error	.5%	.5%	.0%
			95% Confidence Interval Lower	95.7%	2.3%	100.0%
			95% Confidence Interval Upper	97.7%	4.3%	100.0%
			Design Effect	1.011E7	1.011E7	.
	Total	% within Years of education-4 categories 1=0-11 2=12 3=13-15 4=16+	Estimate	94.7%	5.3%	100.0%
			Standard Error	.3%	.3%	.0%
			95% Confidence Interval Lower	94.1%	4.7%	100.0%
			95% Confidence Interval Upper	95.3%	5.9%	100.0%
			Design Effect	1.019E7	1.019E7	.

1	1	% within Years of education-4 categories 1=0-11 2=12 3=13-15 4=16+	Estimate		90.9%	9.1%	100.0%
			Standard Error		2.9%	2.9%	.0%
			95% Confidence Interval	Lower	83.0%	4.7%	100.0%
				Upper	95.3%	17.0%	100.0%
			Design Effect		2.738E7	2.738E7	.
	2	% within Years of education-4 categories 1=0-11 2=12 3=13-15 4=16+	Estimate		95.1%	4.9%	100.0%
			Standard Error		1.3%	1.3%	.0%
			95% Confidence Interval	Lower	91.6%	2.8%	100.0%
				Upper	97.2%	8.4%	100.0%
			Design Effect		1.888E7	1.888E7	.
	3	% within Years of education-4 categories 1=0-11 2=12 3=13-15 4=16+	Estimate		95.1%	4.9%	100.0%
			Standard Error		1.0%	1.0%	.0%
			95% Confidence Interval	Lower	92.6%	3.2%	100.0%
				Upper	96.8%	7.4%	100.0%
			Design Effect		1.120E7	1.120E7	.
	4	% within Years of education-4 categories 1=0-11 2=12 3=13-15 4=16+	Estimate		93.1%	6.9%	100.0%
			Standard Error		1.4%	1.4%	.0%
			95% Confidence Interval	Lower	89.8%	4.6%	100.0%
				Upper	95.4%	10.2%	100.0%
			Design Effect		7.124E6	7.124E6	.
	Total	% within Years of education-4 categories 1=0-11 2=12 3=13-15 4=16+	Estimate		94.1%	5.9%	100.0%
			Standard Error		.9%	.9%	.0%
			95% Confidence Interval	Lower	92.0%	4.4%	100.0%
				Upper	95.6%	8.0%	100.0%
			Design Effect		2.091E7	2.091E7	.

Tests of Independence

age29			Chi-Square	Adjusted F	df1	df2	Sig.
0	Years of education-4 categories * ald	Pearson	13.192	5.315	2.472	103.824	.004
		Likelihood Ratio	14.614	5.888	2.472	103.824	.002
1	Years of education-4 categories * ald	Pearson	6.096	1.659	2.742	112.432	.184
		Likelihood Ratio	5.654	1.539	2.742	112.432	.212

The adjusted F is a variant of the second-order Rao-Scott adjusted chi-square statistic. Significance is based on the adjusted F and its degrees of freedom.

```
*ANALYSIS EXAMPLE 6.9 SIMPLE LOGISTIC REGRESSION: MALE/FEMALE ODDS RATIO OF LT MDE NCSR DATA

* Complex Samples Logistic Regression.
CSLOGISTIC  mde(LOW) WITH sexm
  /PLAN FILE='F:\applied_analysis_book\ncsr_p1wt.csaplan'
  /MODEL sexm
  /INTERCEPT INCLUDE=YES SHOW=YES
  /STATISTICS EXP SE CINTERVAL TTEST DEFF
  /TEST TYPE=ADJCHISQUARE PADJUST=LSD
  /MISSING CLASSMISSING=EXCLUDE
  /CRITERIA MXITER=100 MXSTEP=5 PCONVERGE=[1e-006 RELATIVE] LCONVERGE=[0] CHKSEP=20 CILEVEL=95
  /PRINT SUMMARY VARIABLEINFO SAMPLEINFO.
```

Tests of Model Effects

Source	df	Adjusted Wald Chi-Square	Sig.
(Corrected Model)	1.000	57.277	.000
(Intercept)	1.000	1439.096	.000
sexm	1.000	57.277	.000

Dependent Variable: mde (reference category = 0)

Model: (Intercept), sexm

Parameter Estimates

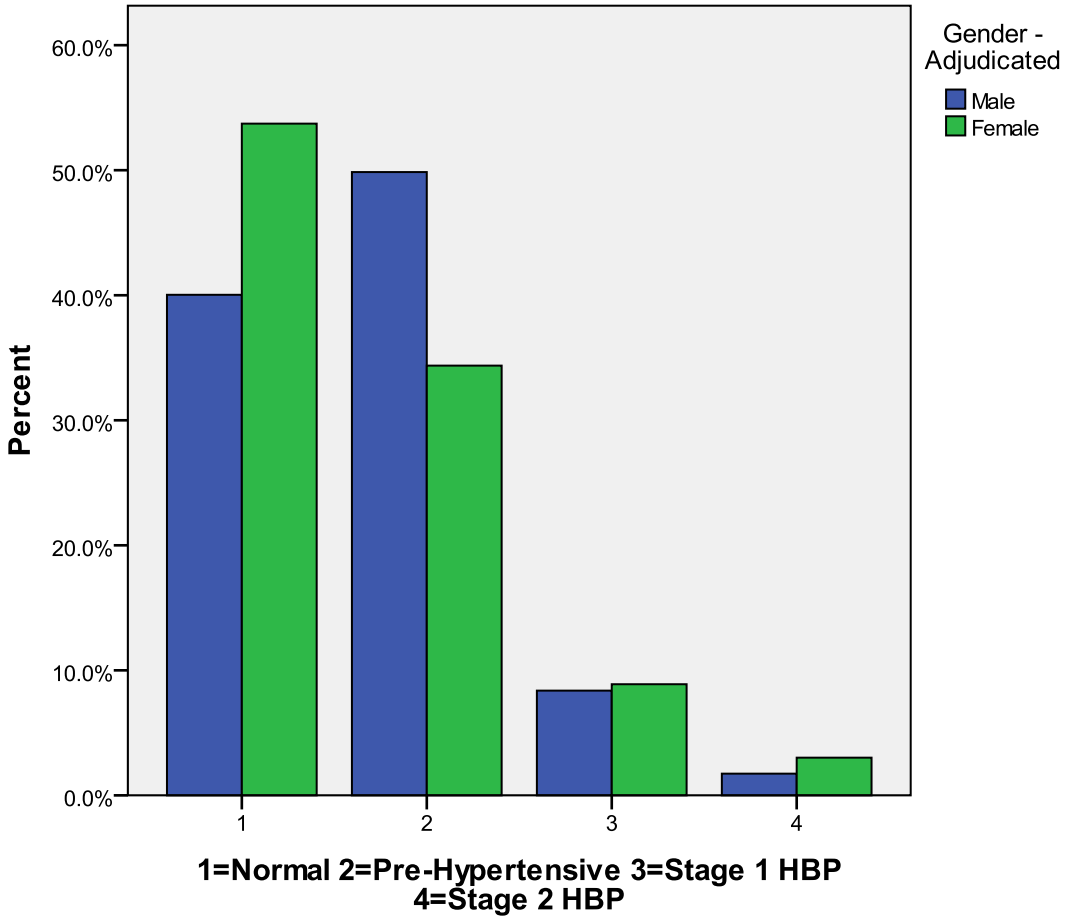
			95% Confidence Interval		Hypothesis Test					95% Confidence Interval for Exp(B)	
mde	Parameter	Std. Error	Lower	Upper	t	df	Sig.	Design Effect	Exp(B)	Lower	Upper
1	(Intercept)	.032	-1.277	-1.148	-37.935	42.000	.000	.873	.298	.279	.317
	sexm	.068	-.654	-.379	-7.568	42.000	.000	1.589	.597	.520	.685

Dependent Variable: mde (reference category = 0)

Model: (Intercept), sexm

Figure 6.8 Bar Graph of Blood Pressure by Gender NHANES Data

```
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
  /BAR(GROUPED)=PCT BY bp_cat BY RIAGENDR.
```




```
*EXAMPLE 6.10 ASSOCIATION BETWEEN GENDER AND DEPRESSION CONTROLLING FOR AGE CATEGORIES: NCSR DATA NOT AVAILABLE IN SPSS/PASW V18
```