```
* SAS Analysis Examples Replication for ASDA 3rd Edition
* Berglund Winter 2025
* Chapter 6 ;

libname d "P:\ASDA3\Data Sets for Analysis Examples and Stata R Code" ;
ods listing ;
ods graphics off ;
options nodate nonumber ls=125 ps=68 ;

ods rtf style=minimal bodytitle ;
title ;
data c6_nhanes ;
  set d.nhanes1112 ;
  if age >= 18 then age18p=1 ; else if age >= 0 and age <18 then age18p=0 ;
run ;

title "Example 6.1: Estimating the Proportion of the U.S. Adult Population with an Irregular Heart Beat. " ;
* wald confidence limit is default ;
proc surveyfreq data=c6_nhanes ;
 strata sdmvstra ; cluster sdmvpsu ; weight wtmec2yr ;
 tables age18p*irregular / deff row cl ;
 run ;
* logit confidence limit ;
proc surveyfreq data=c6_nhanes ;
 strata sdmvstra ; cluster sdmvpsu ; weight wtmec2yr ;
 tables age18p*irregular / deff row cl(logit)  ;
run ;
* means for proportion of 0/1 variable ;
proc surveymeans data=c6_nhanes ;
 strata sdmvstra ; cluster sdmvpsu ; weight wtmec2yr ;
 domain age18p ;
 var irregular  ;
run ;
* Note: Franco Confidence Intervals not available in SAS proc surveymeans ;

title " Example 6.2: Estimating the Proportion of U.S. Adults by Race and Ethnicity using NHANES data. " ;
proc surveyfreq data=c6_nhanes ;
 strata sdmvstra ; cluster sdmvpsu ; weight wtmec2yr ;
 tables age18p*ridreth1 / nowt nocellpercent row(deff cl ) ;
run ;

* Note: Bayesian approach not available in SAS ;

title " Example 6.3: Estimating the Proportions of U.S. Adults by Blood Pressure Category using the 2011-2012 NHANES
Data." ;
proc surveyfreq data=c6_nhanes ;
 strata sdmvstra ; cluster sdmvpsu ; weight wtmec2yr ;
 tables age18p*bp_cat / nowt nocellpercent row(deff cl) ;
run ;

libname russia "P:\ASDA3\Data Sets for Analysis Examples and Stata R Code" ;
data c6_russia ;
  set russia.ess6_russia ;
run ;

title "Example 6.4: A Goodness of Fit Test for Proportions of Russians age 15+ by Marital Status." ;
proc surveyfreq data=c6_russia ;
 strata stratify ; cluster psu ; weight pspwght ;
 tables marcat / row(deff cl) lrchisq(secondorder) testp=(.5 .25 .25) ;
run ;

* Note PROC SGPLOT allows use a weight statement ;
proc format ; value marcatf 1='Currently Married' 2='Previously Married' 3='Never Married' ;run ;
proc freq data=c6_russia ;
 table marcat / out=outfreq ;
  weight pspwght ;
  format marcat marcatf. ;
run ;

* Bar chart ;

title "Example 6.5: Bar Chart of the Estimated Proportions of Russians age 15+ by Marital Status." ;
proc sgplot data=outfreq ;
 format marcat marcatf. percent 6.2 ;
  styleattrs backcolor=white datacolors=(orange green blue);
 vbar marcat / response=percent datalabel ;
run ;

*Use PROC SGPIE (pre-production) in SAS STAT 15.3, note no weight statement so use output from proc freq with weight
variable used;
```

```
title "Example 6.5: Pie Chart of the Estimated Proportions of Russians age 15+ by Marital Status." ;
proc sgpie data=outfreq ;
 format marcat marcatf. percent 6.2 ;
 styleattrs backcolor=white datacolors=(orange green blue);
 pie marcat / response=percent ;
run ;

libname ncsr "P:\ASDA3\Data Sets for Analysis Examples and Stata R Code" ;
data c6_ncsr ;
 set ncsr.ncsr ;
run ;

* include formats for tables ;
proc format ;
 value sexf 1='Male' 2='Female' ; value mdef 1='MDE' 0='No MDE' ; run ;

title " Example 6.6: Estimation of Total and Row Proportions for the Crosstabulation of Gender and Lifetime Major
Depression Status (Source: NCS-R)." ;
proc surveyfreq data=c6_ncsr ;
 strata sestrat ; cluster seclustr ; weight ncsrwtsh ;
 tables sex*mde / deff chisq (secondorder) ;
 format sex sexf. mde mdef. ;
run ;

proc surveyfreq data=c6_ncsr ;
 strata sestrat ; cluster seclustr ; weight ncsrwtsh ;
 tables sex*mde / row(deff cl) chisq(secondorder) ;
 format sex sexf. mde mdef. ;
run ;

title " Example 6.7: Comparing the Proportions of U.S. Adult Men and Women with Lifetime Major Depression.  " ;
* use proc surveymeans for difference in means ;
proc surveymeans data=c6_ncsr ;
 strata sestrat ; cluster seclustr ; weight ncsrwtsh ;
 domain sex / diffmeans ;
 var mde ;
run ;

* alternate method: linear contrast of male v. female mde done with LSMEANS / DIFF option in PROC SURVEYREG ;
proc surveyreg data=c6_ncsr ;
 strata sestrat ; cluster seclustr ; weight ncsrwtsh ;
 class sex ;
 model mde = sex / solution ;
 lsmeans sex /diff ;
run ;
* Note: no Bayesian method available in SAS ;

title " Example 6.8: Testing the Independence of MDE and Gender in U.S. Adults Using the NCS-R data." ;
proc surveyfreq data=c6_ncsr ;
 strata sestrat ; cluster seclustr ; weight ncsrwtsh ;
 tables sex*mde / chisq(secondorder) ;
run ;

* data management for next analysis ;
data c6_ncsr1 ;
 set c6_ncsr ;
 * create indicator for subpopulation of interest ;
 age18_28=0 ;
 if 18<=age<=28 then age18_28=1 ;
run ;

title "Example 6.9: Testing the Independence of Alcohol Dependence and Education Level in Young Adults (Ages 18-28)
using the NCS-R data. " ;
proc surveyfreq data=c6_ncsr1 ;
 strata sestrat ; cluster seclustr ; weight ncsrwtlg ;
 tables age18_28*ed4cat*ald / nocellpercent row chisq chisq(secondorder) ;
 tables age18_28*ald / row chisq chisq(secondorder) ;
run ;

* compare to chisq from proc freq without design correction;
proc freq data=c6_ncsr1 ;
 weight ncsrwtlg ;
 where age18_28=1 ;
 tables ed4cat*ald / chisq  ;
run ;

title "Example 6.10: Simple Logistic Regression to Estimate the NCS-R Male/Female Odds Ratio for Lifetime Major
Depressive Episode. " ;
proc surveylogistic data=c6_ncsr ;
 strata sestrat ; cluster seclustr ; weight ncsrwtsh ;
```

```
 model mde(event='1') = sexm ;
run ;

proc format ; value gndrf 1='Male' 2='Female' ; run ;
title "Figure 6.8: Weighted Bar Chart of Marital Status by Gender for Russians age 15+" ;
proc sgplot data=c6_russia ;
 vbar gndr / group=marcat stat=percent weight=pspwght groupdisplay=cluster datalabel ;
 format marcat marcatf. gndr gndrf.  ;
run ;

* Example 6.11: Using the NCS-R Data to Estimate and Test the Association between Gender and Depression in the U.S.
Adult Population when controlling for Age.
* NOTE: this is done using SUDAAN in book as SAS a SURVEY procedure for this test

* Example 6.12: A Simple Log-linear Model to Test the Association between Lifetime Major Depression Episode and Sex.
* NOTE: done in IVEware and R in book since SAS does not offer a SURVEY procedure for Log-Linear models ;
ods rtf close ;
```

Example 6.1: Estimating the Proportion of the U.S. Adult Population with an Irregular Heart Beat.

The SURVEYFREQ Procedure

Data Summary	
Number of Strata	14
Number of Clusters	31
Number of Observations	9756
Number of Observations Used	9338
Number of Obs with Nonpositive Weights	418
Sum of Weights	306590681

Table of age18p by irregular									
age18p	irregular	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent	95% Confidence Limits for Percent		Design Effect
0	0	3642	72706140	5745470	24.5373	0.7485	22.9582	26.1164	2.7306
	1	11	305580	116221	0.1031	0.0385	0.0220	0.1843	1.2958
	Total	3653	73011720	5775234	24.6404	0.7560	23.0454	26.2355	2.7783
1	0	5264	219630508	13382261	74.1222	0.7705	72.4966	75.7479	2.7938
	1	110	3666308	459959	1.2373	0.1254	0.9727	1.5019	1.1617
	Total	5374	223296816	13647988	75.3596	0.7560	73.7645	76.9546	2.7783
Total	0	8906	292336649	18562605	98.6595	0.1460	98.3515	98.9676	1.4547
	1	121	3971887	522982	1.3405	0.1460	1.0324	1.6485	1.4547
	Total	9027	296308536	18854873	100.0000				
Frequency Missing = 311									

Table of age18p by irregular					
age18p	irregular	Row Percent	Std Err of Row Percent	95% Confidence Limits for Row Percent	
0	0	99.5815	0.1537	99.2571	99.9058
	1	0.4185	0.1537	0.0942	0.7429
	Total	100.0000			
1	0	98.3581	0.1678	98.0041	98.7121
	1	1.6419	0.1678	1.2879	1.9959
	Total	100.0000			
Total	0				
	1				
	Total				
Frequency Missing = 311					

Example 6.1: Estimating the Proportion of the U.S. Adult Population with an Irregular Heart Beat.

The SURVEYFREQ Procedure

Data Summary	
Number of Strata	14
Number of Clusters	31
Number of Observations	9756
Number of Observations Used	9338
Number of Obs with Nonpositive Weights	418
Sum of Weights	306590681

Table of age18p by irregular									
age18p	irregular	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent	95% Confidence Limits for Percent		Design Effect
0	0	3642	72706140	5745470	24.5373	0.7485	22.9927	26.1505	2.7306
	1	11	305580	116221	0.1031	0.0385	0.0469	0.2264	1.2958
	Total	3653	73011720	5775234	24.6404	0.7560	23.0803	26.2700	2.7783
1	0	5264	219630508	13382261	74.1222	0.7705	72.4636	75.7144	2.7938
	1	110	3666308	459959	1.2373	0.1254	0.9988	1.5319	1.1617
	Total	5374	223296816	13647988	75.3596	0.7560	73.7300	76.9197	2.7783
Total	0	8906	292336649	18562605	98.6595	0.1460	98.3139	98.9351	1.4547
	1	121	3971887	522982	1.3405	0.1460	1.0649	1.6861	1.4547
	Total	9027	296308536	18854873	100.0000				
Logit confidence limits are computed for percents.									
Frequency Missing = 311									

Table of age18p by irregular					
age18p	irregular	Row Percent	Std Err of Row Percent	95% Confidence Limits for Row Percent	
0	0	99.5815	0.1537	99.0930	99.8074
	1	0.4185	0.1537	0.1926	0.9070
	Total	100.0000			
1	0	98.3581	0.1678	97.9639	98.6770
	1	1.6419	0.1678	1.3230	2.0361
	Total	100.0000			
Total	0				
	1				
	Total				
Logit confidence limits are computed for percents.					
Frequency Missing = 311					

Example 6.1: Estimating the Proportion of the U.S. Adult Population with an Irregular Heart Beat.

The SURVEYMEANS Procedure

Data Summary	
Number of Strata	14
Number of Clusters	31
Number of Observations	9756
Number of Observations Used	9338
Number of Obs with Nonpositive Weights	418
Sum of Weights	306590681

Statistics						
Variable	Label	N	Mean	Std Error of Mean	95% CL for Mean	
irregular	1=yes 0=no	9027	0.013405	0.001460	0.01032441	0.01648472

Example 6.1: Estimating the Proportion of the U.S. Adult Population with an Irregular Heart Beat.

The SURVEYMEANS Procedure

Statistics for age18p Domains							
age18p	Variable	Label	N	Mean	Std Error of Mean	95% CL for Mean	
0	irregular	1=yes 0=no	3653	0.004185	0.001537	0.00094168	0.00742902
1	irregular	1=yes 0=no	5374	0.016419	0.001678	0.01287902	0.01995896

Example 6.2: Estimating the Proportion of U.S. Adults by Race and Ethnicity using NHANES data.

The SURVEYFREQ Procedure

Data Summary	
Number of Strata	14
Number of Clusters	31
Number of Observations	9756
Number of Observations Used	9338
Number of Obs with Nonpositive Weights	418
Sum of Weights	306590681

Table of age18p by RIDRETH1							
age18p	RIDRETH1	Frequency	Row Percent	Std Err of Row Percent	95% Confidence Limits for Row Percent		Design Effect of Row Percent
0	1	747	15.3445	3.1660	8.6649	22.0241	28.7200
	2	434	8.1138	1.6928	4.5423	11.6854	14.3062
	3	827	53.2818	4.9603	42.8165	63.7472	36.7898
	4	1077	14.6731	2.8088	8.7470	20.5992	23.4540
	5	638	8.5867	1.1462	6.1684	11.0051	6.2301
	Total	3723	100.0000				
1	1	569	7.9168	1.7251	4.2772	11.5563	22.9166
	2	577	6.6224	1.5193	3.4171	9.8278	20.9546
	3	2014	65.9386	3.8892	57.7332	74.1440	37.8077
	4	1505	11.7185	2.3370	6.7878	16.6491	29.6379
	5	950	7.8037	1.0917	5.5004	10.1070	9.2998
	Total	5615	100.0000				
Total	1	1316					
	2	1011					
	3	2841					
	4	2582					
	5	1588					
	Total	9338					

Example 6.3: Estimating the Proportions of U.S. Adults by Blood Pressure Category using the 2011-2012 NHANES Data.

The SURVEYFREQ Procedure

Data Summary	
Number of Strata	14
Number of Clusters	31
Number of Observations	9756
Number of Observations Used	9338
Number of Obs with Nonpositive Weights	418
Sum of Weights	306590681

Table of age18p by bp_cat

age18p	bp_cat	Frequency	Row Percent	Std Err of Row Percent	95% Confidence Limits for Row Percent		Design Effect of Row Percent
0	1	1543	90.3833	0.9724	88.3317	92.4350	1.8474
	2	154	9.5266	0.9553	7.5110	11.5422	1.7980
	3	2	0.0901	0.0658	0.0000	0.2289	0.8163
	4	0	.	.	.	.	.
	Total	1699	100.0000				
1	1	2438	47.2223	1.5521	43.9477	50.4970	5.1760
	2	2284	42.7985	1.2035	40.2593	45.3378	3.1684
	3	489	7.9778	0.5815	6.7509	9.2047	2.4669
	4	145	2.0013	0.4385	1.0762	2.9264	5.2493
	Total	5356	100.0000				
Total	1	3981					
	2	2438					
	3	491					
	4	145					
	Total	7055					

Frequency Missing = 2283

Example 6.4: A Goodness of Fit Test for Proportions of Russians age 15+ by Marital Status.

The SURVEYFREQ Procedure

Data Summary	
Number of Strata	8
Number of Clusters	184
Number of Observations	2484
Sum of Weights	2484

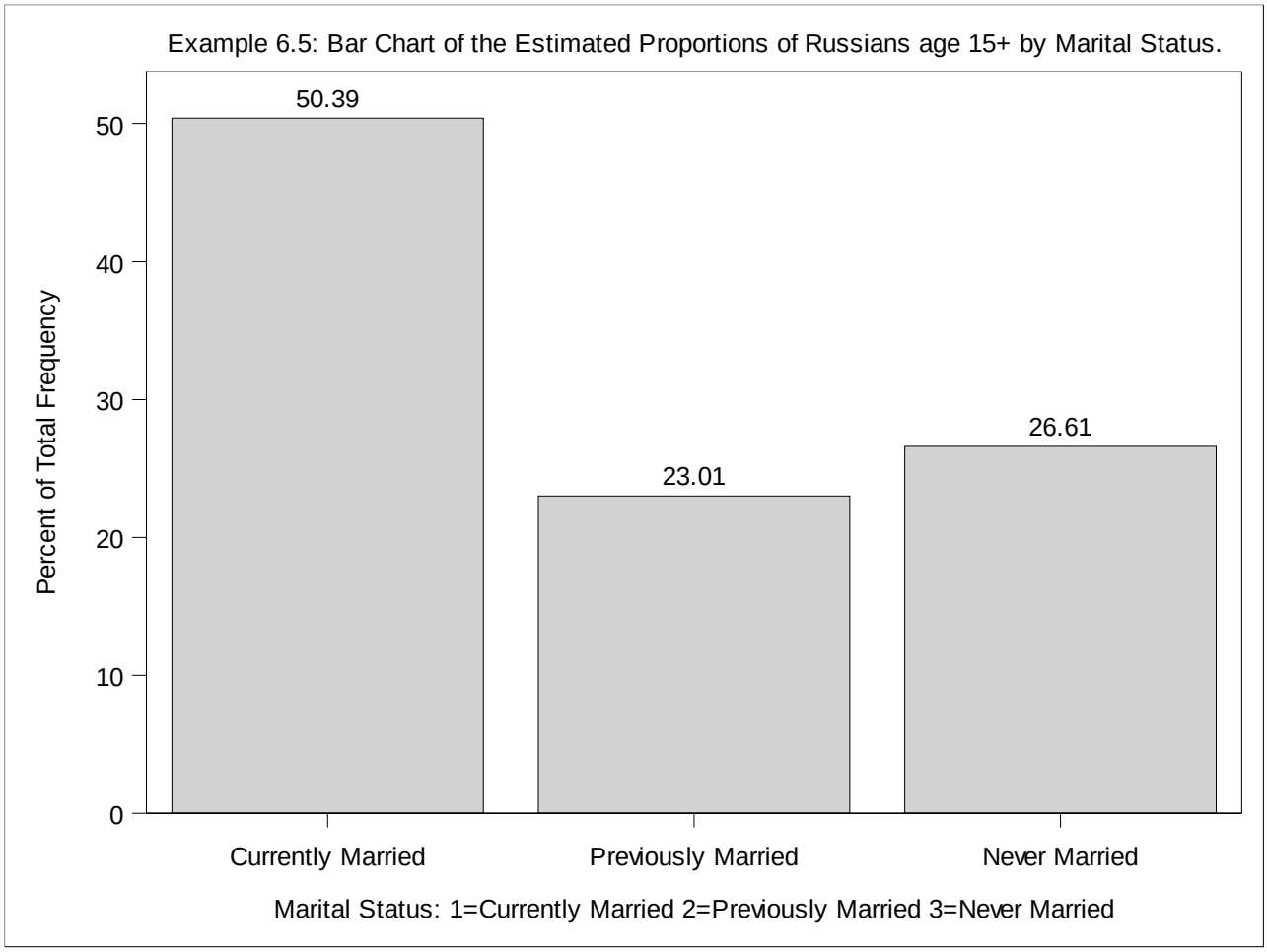
Marital Status: 1=Currently Married 2=Previously Married 3=Never Married						
marcat	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Test Percent	Std Err of Percent
1	1066	1236	61.07696	50.3860	50.00	1.2878
2	791	564.34716	36.12788	23.0066	25.00	1.1536
3	587	652.67253	37.60300	26.6074	25.00	1.3401
Total	2444	2453	94.17226	100.0000		
Frequency Missing = 40						

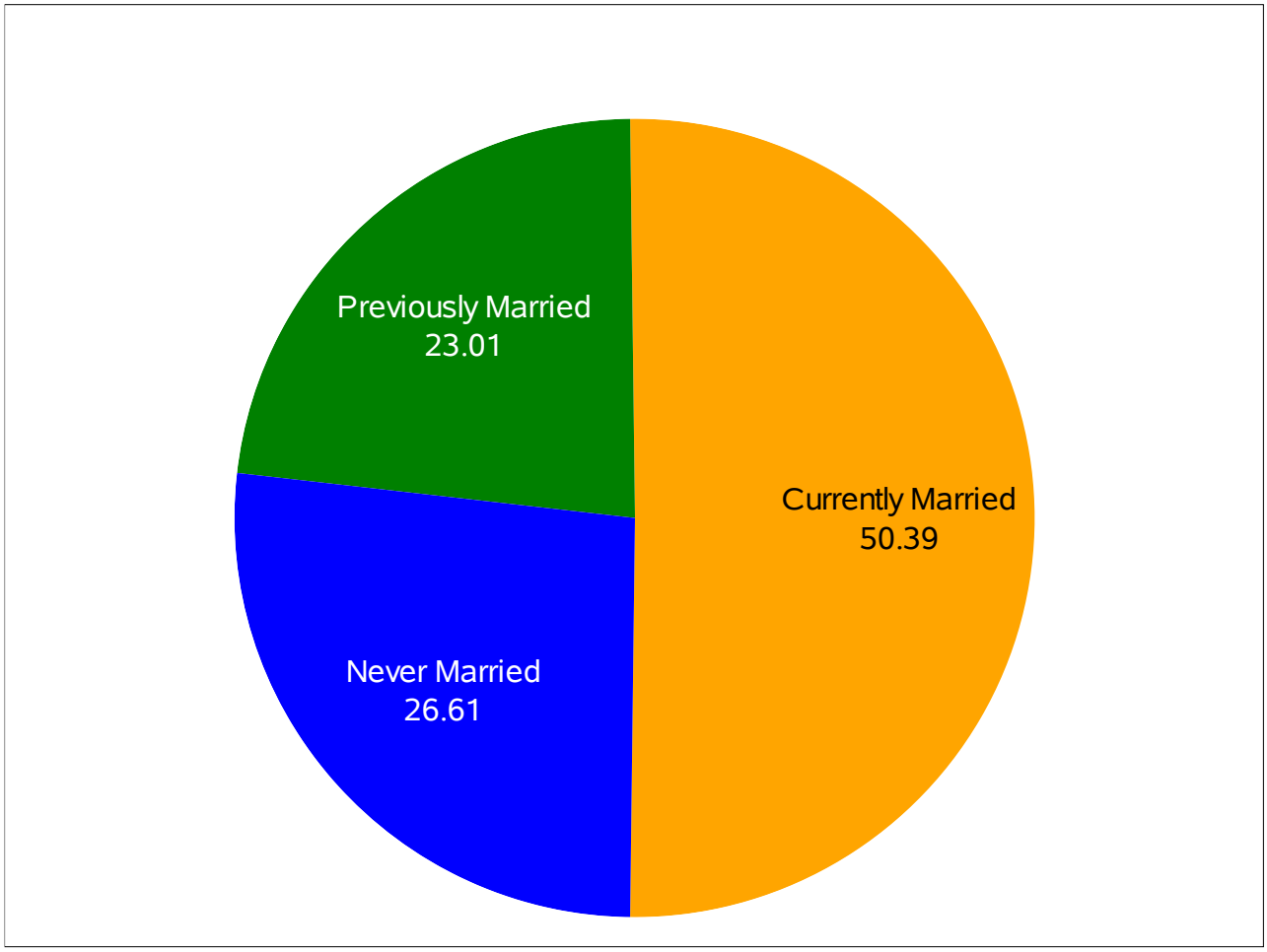
Rao-Scott Likelihood Ratio Test	
Likelihood Ratio Chi-Square	6.5381
Design Correction	1.9332
First-Order Chi-Square	3.3821
Second-Order Chi-Square	3.2472
DF	1.92
Pr > ChiSq	0.1856
F Value	1.6910
Num DF	1.92
Den DF	337.97
Pr > F	0.1871
Sample Size = 2444	

Example 6.4: A Goodness of Fit Test for Proportions of Russians age 15+ by Marital Status.

The FREQ Procedure

Marital Status: 1=Currently Married 2=Previously Married 3=Never Married				
marcat	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Currently Married	1235.957	50.39	1235.957	50.39
Previously Married	564.3472	23.01	1800.304	73.39
Never Married	652.6725	26.61	2452.977	100.00
Frequency Missing = 31.023132863				


Example 6.5: Bar Chart of the Estimated Proportions of Russians age 15+ by Marital Status.
50.39
23.01
26.61
Percent of Total Frequency
0
10
20
30
40
50
Currently Married
Previously Married
Never Married
Marital Status: 1=Currently Married 2=Previously Married 3=Never Married


Previously Married
23.01
Currently Married
50.39
Never Married
26.61

Example 6.6: Estimation of Total and Row Proportions for the Crosstabulation of Gender and Lifetime Major Depression Status (Source: NCS-R).

The SURVEYFREQ Procedure

Data Summary	
Number of Strata	42
Number of Clusters	84
Number of Observations	9282
Sum of Weights	9282.00015

Table of sex by mde							
sex	mde	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent	Design Effect
Male	No MDE	3522	3774	169.19112	40.6644	0.6980	1.8741
	MDE	617	670.23208	57.70029	7.2208	0.3438	1.6372
	Total	4139	4445	215.70025	47.8852	0.5315	1.0508
Female	No MDE	3931	3728	195.07524	40.1644	0.5361	1.1097
	MDE	1212	1109	61.50166	11.9504	0.3028	0.8086
	Total	5143	4837	248.29286	52.1148	0.5315	1.0508
Total	No MDE	7453	7503	349.57814	80.8289	0.4877	1.4245
	MDE	1829	1779	113.95611	19.1711	0.4877	1.4245
	Total	9282	9282	453.54554	100.0000		

Rao-Scott Chi-Square Test	
Pearson Chi-Square	92.1499
Design Correction	1.3725
First-Order Chi-Square	67.1387
Second-Order Chi-Square	67.1387
DF	1
Pr > ChiSq	<.0001
F Value	67.1387
Num DF	1
Den DF	42
Pr > F	<.0001
Sample Size = 9282	

Example 6.6: Estimation of Total and Row Proportions for the Crosstabulation of Gender and Lifetime Major Depression Status (Source: NCS-R).

The SURVEYFREQ Procedure

Data Summary	
Number of Strata	42
Number of Clusters	84
Number of Observations	9282
Sum of Weights	9282.00015

Table of sex by mde								
sex	mde	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent	Row Percent	Std Err of Row Percent
Male	No MDE	3522	3774	169.19112	40.6644	0.6980	84.9207	0.7748
	MDE	617	670.23208	57.70029	7.2208	0.3438	15.0793	0.7748
	Total	4139	4445	215.70025	47.8852	0.5315	100.0000	
Female	No MDE	3931	3728	195.07524	40.1644	0.5361	77.0692	0.5647
	MDE	1212	1109	61.50166	11.9504	0.3028	22.9308	0.5647
	Total	5143	4837	248.29286	52.1148	0.5315	100.0000	
Total	No MDE	7453	7503	349.57814	80.8289	0.4877		
	MDE	1829	1779	113.95611	19.1711	0.4877		
	Total	9282	9282	453.54554	100.0000			

Table of sex by mde				
sex	mde	95% Confidence Limits for Row Percent		Design Effect of Row Percent
Male	No MDE	83.3571	86.4842	1.9398
	MDE	13.5158	16.6429	1.9398
	Total			
Female	No MDE	75.9295	78.2088	0.9279
	MDE	21.7912	24.0705	0.9279
	Total			
Total	No MDE			
	MDE			
	Total			

Rao-Scott Chi-Square Test	
Pearson Chi-Square	92.1499
Design Correction	1.3725
First-Order Chi-Square	67.1387
Second-Order Chi-Square	67.1387
DF	1
Pr > ChiSq	<.0001

Rao-Scott Chi-Square Test	
F Value	67.1387
Num DF	1
Den DF	42
Pr > F	<.0001
Sample Size = 9282	

Example 6.7: Comparing the Proportions of U.S. Adult Men and Women with Lifetime Major Depression.

The SURVEYMEANS Procedure

Data Summary	
Number of Strata	42
Number of Clusters	84
Number of Observations	9282
Sum of Weights	9282.00015

Statistics						
Variable	Label	N	Mean	Std Error of Mean	95% CL for Mean	
mde	Major Depressive Episode 1=Yes 0=No	9282	0.191711	0.004877	0.18186944	0.20155304

Example 6.7: Comparing the Proportions of U.S. Adult Men and Women with Lifetime Major Depression.

The SURVEYMEANS Procedure

Statistics for sex Domains							
sex	Variable	Label	N	Mean	Std Error of Mean	95% CL for Mean	
1	mde	Major Depressive Episode 1=Yes 0=No	4139	0.150793	0.007748	0.13515762	0.16642905
2	mde	Major Depressive Episode 1=Yes 0=No	5143	0.229308	0.005647	0.21791168	0.24070492

Differences of mde (Major Depressive Episode 1=Yes 0=No) Means for sex Domains						
sex	-sex	Diff Estimate	Std Error	DF	t Value	Pr > |t|
1	2	-0.078515	0.009552	42	-8.22	<.0001

Example 6.7: Comparing the Proportions of U.S. Adult Men and Women with Lifetime Major Depression.

The SURVEYREG Procedure

Regression Analysis for Dependent Variable mde

Data Summary	
Number of Observations	9282
Sum of Weights	9282.0
Weighted Mean of mde	0.19171
Weighted Sum of mde	1779.5

Design Summary	
Number of Strata	42
Number of Clusters	84

Fit Statistics	
R-Square	0.009928
Root MSE	0.3917
Denominator DF	42

Class Level Information			
CLASS Variable	Label	Levels	Values
sex	Sex 1=Male 2=Female	2	1 2

Tests of Model Effects			
Effect	Num DF	F Value	Pr > F
Model	1	67.56	<.0001
Intercept	1	1560.02	<.0001
sex	1	67.56	<.0001

The denominator degrees of freedom for the F tests is 42.

Estimated Regression Coefficients				
Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.2293083	0.00564756	40.60	<.0001
sex 1	-0.0785150	0.00955235	-8.22	<.0001
sex 2	0.0000000	0.00000000	.	.

The degrees of freedom for the t tests is 42.
Matrix X'WX is singular and a generalized inverse was used to solve the normal equations. Estimates are not unique.

sex Least Squares Means					
Sex 1=Male 2=Female	Estimate	Standard Error	DF	t Value	Pr > |t|
1	0.1508	0.007748	42	19.46	<.0001

sex Least Squares Means					
Sex 1=Male 2=Female	Estimate	Standard Error	DF	t Value	Pr > |t|
2	0.2293	0.005648	42	40.60	<.0001

Differences of sex Least Squares Means						
Sex 1=Male 2=Female	Sex 1=Male 2=Female	Estimate	Standard Error	DF	t Value	Pr > |t|
1	2	-0.07851	0.009552	42	-8.22	<.0001

Example 6.8: Testing the Independence of MDE and Gender in U.S. Adults Using the NCS-R data.

The SURVEYFREQ Procedure

Data Summary	
Number of Strata	42
Number of Clusters	84
Number of Observations	9282
Sum of Weights	9282.00015

Table of sex by mde

sex	mde	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent
1	0	3522	3774	169.19112	40.6644	0.6980
	1	617	670.23208	57.70029	7.2208	0.3438
	Total	4139	4445	215.70025	47.8852	0.5315
2	0	3931	3728	195.07524	40.1644	0.5361
	1	1212	1109	61.50166	11.9504	0.3028
	Total	5143	4837	248.29286	52.1148	0.5315
Total	0	7453	7503	349.57814	80.8289	0.4877
	1	1829	1779	113.95611	19.1711	0.4877
	Total	9282	9282	453.54554	100.0000	

Rao-Scott Chi-Square Test	
Pearson Chi-Square	92.1499
Design Correction	1.3725
First-Order Chi-Square	67.1387
Second-Order Chi-Square	67.1387
DF	1
Pr > ChiSq	<.0001
F Value	67.1387
Num DF	1
Den DF	42
Pr > F	<.0001
Sample Size = 9282	

Example 6.9: Testing the Independence of Alcohol Dependence and Education Level in Young Adults (Ages 18-28) using the NCS-R data.

The SURVEYFREQ Procedure

Data Summary	
Number of Strata	42
Number of Clusters	84
Number of Observations	9282
Number of Observations Used	5692
Number of Obs with Nonpositive Weights	3590
Sum of Weights	5692.00048

Table of ED4CAT by ald						
Controlling for age18_28=0						
ED4CAT	ald	Frequency	Weighted Frequency	Std Err of Wgt Freq	Row Percent	Std Err of Row Percent
1	0	570	689.26106	48.65927	94.0454	0.7976
	1	71	43.64177	5.63734	5.9546	0.7976
	Total	641	732.90283	49.40849	100.0000	
2	0	1223	1362	89.50257	94.1603	0.6250
	1	107	84.49599	11.21382	5.8397	0.6250
	Total	1330	1447	95.77820	100.0000	
3	0	1145	1064	50.18931	93.8975	0.5242
	1	106	69.16919	7.39750	6.1025	0.5242
	Total	1251	1133	54.43718	100.0000	
4	0	1138	1077	70.55346	96.8458	0.4829
	1	57	35.07967	4.91665	3.1542	0.4829
	Total	1195	1112	70.44264	100.0000	
Total	0	4076	4193	184.54693		
	1	341	232.38662	17.38587		
	Total	4417	4425	194.71240		

Rao-Scott Chi-Square Test	
Pearson Chi-Square	13.1918
Design Correction	0.8128
First-Order Chi-Square	16.2305
Second-Order Chi-Square	12.1218
DF	2.24
Pr > ChiSq	0.0031
F Value	5.4102
Num DF	2.24
Den DF	94.10
Pr > F	0.0044

Rao-Scott Chi-Square Test
Sample Size = 5692

Table of ED4CAT by ald						
Controlling for age18_28=1						
ED4CAT	ald	Frequency	Weighted Frequency	Std Err of Wgt Freq	Row Percent	Std Err of Row Percent
1	0	186	200.86597	21.26023	90.8714	2.9380
	1	22	20.17818	6.69593	9.1286	2.9380
	Total	208	221.04416	21.91184	100.0000	
2	0	356	384.76486	26.45399	95.1442	1.3460
	1	26	19.63715	5.96079	4.8558	1.3460
	Total	382	404.40202	28.73956	100.0000	
3	0	424	413.14106	46.87778	95.1042	1.0042
	1	34	21.26768	4.69579	4.8958	1.0042
	Total	458	434.40873	48.50085	100.0000	
4	0	207	192.43145	30.56315	93.0962	1.3640
	1	20	14.27020	3.30762	6.9038	1.3640
	Total	227	206.70164	32.22052	100.0000	
Total	0	1173	1191	87.85136		
	1	102	75.35321	13.36129		
	Total	1275	1267	94.77459		

Rao-Scott Chi-Square Test	
Pearson Chi-Square	6.0957
Design Correction	1.4828
First-Order Chi-Square	4.1109
Second-Order Chi-Square	3.1310
DF	2.28
Pr > ChiSq	0.2536
F Value	1.3703
Num DF	2.28
Den DF	95.97
Pr > F	0.2591
Sample Size = 5692	

Table of age18_28 by ald								
age18_28	ald	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent	Row Percent	Std Err of Row Percent
0	0	4076	4193	184.54693	73.6658	1.1439	94.7489	0.3102
	1	341	232.38662	17.38587	4.0827	0.2490	5.2511	0.3102
	Total	4417	4425	194.71240	77.7485	1.1801	100.0000	

Table of age18_28 by ald								
age18_28	ald	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent	Row Percent	Std Err of Row Percent
1	0	1173	1191	87.85136	20.9277	1.1271	94.0505	0.8809
	1	102	75.35321	13.36129	1.3238	0.2082	5.9495	0.8809
	Total	1275	1267	94.77459	22.2515	1.1801	100.0000	
Total	0	5249	5384	234.33376	94.5935	0.3248		
	1	443	307.73983	25.12875	5.4065	0.3248		
	Total	5692	5692	251.09560	100.0000			

Rao-Scott Chi-Square Test	
Pearson Chi-Square	0.9389
Design Correction	1.5825
First-Order Chi-Square	0.5933
Second-Order Chi-Square	0.5933
DF	1
Pr > ChiSq	0.4411
F Value	0.5933
Num DF	1
Den DF	42
Pr > F	0.4454
Sample Size = 5692	

Example 6.9: Testing the Independence of Alcohol Dependence and Education Level in Young Adults (Ages 18-28) using the NCS-R data.

The FREQ Procedure

Table of ED4CAT by ald			
ED4CAT(Education 1=0-11 2=12 3=13-15 4=16+ Yrs)	ald(Alcohol Dependence 1=Yes 0=No)		
Frequency Percent Row Pct Col Pct	0	1	Total
1	200.866 15.86 90.87 16.86	20.1782 1.59 9.13 26.78	221.044 17.45
2	384.765 30.38 95.14 32.30	19.6372 1.55 4.86 26.06	404.402 31.93
3	413.141 32.62 95.10 34.68	21.2677 1.68 4.90 28.22	434.409 34.30
4	192.431 15.19 93.10 16.15	14.2702 1.13 6.90 18.94	206.702 16.32
Total	1191.2 94.05	75.3532 5.95	1266.56 100.00

Statistics for Table of ED4CAT by ald

Statistic	DF	Value	Prob
Chi-Square	3	6.0553	0.1089
Likelihood Ratio Chi-Square	3	5.6170	0.1318
Mantel-Haenszel Chi-Square	1	0.8938	0.3444
Phi Coefficient		0.0691	
Contingency Coefficient		0.0690	
Cramer's V		0.0691	

Sample Size = 1266.556545

Example 6.10: Simple Logistic Regression to Estimate the NCS-R Male/Female Odds Ratio for Lifetime Major Depressive Episode.

The SURVEYLOGISTIC Procedure

Model Information		
Data Set	WORK.C6_NCSR	
Response Variable	mde	Major Depressive Episode 1=Yes 0=No
Number of Response Levels	2	
Stratum Variable	sestrat	SAMPLING ERROR STRATUM
Number of Strata	42	
Cluster Variable	seclustr	SAMPLING ERROR CLUSTER
Number of Clusters	84	
Weight Variable	ncsrwtsh	NCSR sample part 1 weight
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	9282
Number of Observations Used	9282
Sum of Weights Read	9282
Sum of Weights Used	9282

Response Profile			
Ordered Value	mde	Total Frequency	Total Weight
1	0	7453	7502.5364
2	1	1829	1779.4637

Probability modeled is mde=1.

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	9074.130	8983.024
SC	9081.266	8997.296
-2 Log L	9072.130	8979.024

Testing Global Null Hypothesis: BETA=0				
Test	F Value	Num DF	Den DF	Pr > F
Likelihood Ratio	76.79	1	42	<.0001
Score	89.70	1	42	<.0001
Wald	57.26	1	42	<.0001
NOTE: First-order Rao-Scott design correction 1.2124 applied to the likelihood ratio test.				

Analysis of Maximum Likelihood Estimates				
Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	-1.2122	0.0320	-37.93	<.0001
sexm	-0.5160	0.0682	-7.57	<.0001
NOTE: The degrees of freedom for the t tests is 42.				

Odds Ratio Estimates			
Effect	Point Estimate	95% Confidence Limits	
sexm	0.597	0.520	0.685
NOTE: The degrees of freedom in computing the confidence limits is 42.			

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	31.3	Somers' D	0.135
Percent Discordant	17.8	Gamma	0.275
Percent Tied	50.9	Tau-a	0.043
Pairs	13631537	c	0.568

Figure 6.8: Weighted Bar Chart of Marital Status by Gender for Russians age 15+