CHAPTER 9 ASDA ANALYSIS EXAMPLES REPLICATION-SUDAAN 10.0.1

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 LOGISTIC/POISSON REGRESSION ANALYSIS IN SUDAAN 10.0.1

The analysis replication examples were all run using SAS-callable SUDAAN version 10.0.1. There are very few differences between SAS-callable and stand-alone SUDAAN with the exception of the names of the procedures are sometimes slightly different as to avoid confusion with SAS procedures.

Sudaan does not offer the ability to perform graphical analyses within the program therefore are not included in this output however output data sets can be saved and used in other software packages, see Chapter 7 for an example of doing this.

Sudaan commands MULTILOG/LOGLINK can perform some of the analyses presented in Chapter 9 of ASDA with the exception of negative binomial regression and the zero-inflated versions of Poisson and negative binomial regression. Some of the fine points of these procedures are the use of a SUBPOPN statement for subpopulation analyses, a CLASS statement for declaration of categorical variables, RFORMAT and REFLEVEL for use with formatted variables and optional reference level changes, and an EFFECTS/CONTRAST statement for hypothesis tests and many other options for analysis/output. Please see the Sudaan 10.0.1 Language and Examples Guides for additional detail.

```
title "Analysis Example 9.2: Multinomial Logistic Regression : NCSR" ;
proc multilog data=ncsr filetype=sas deft1 ;
nest sestrat seclustr ;
weight ncsrwtlg ;
class revworkstatus ed4cat ag4cat mar3cat / nofreq ;
rformat ag4cat agf. ; rformat sex sf. ; rformat ed4cat edf. ; rformat mar3cat marf. ; rformat mde mdef. ;
rformat revworkstatus wkr. ;  rformat ald aldf. ;
reflevel ag4cat=1 mar3cat=1 ed4cat=1 ;
model revworkstatus = sexm ald mde ed4cat ag4cat mar3cat / genlogit ;
test adjwaldf ;
print / style=nchs    ;
run ;
```

Analysis Example 9.2: Multinomial Logistic Regression : NCSR

S U D A A N
Software for the Statistical Analysis of Correlated Data
Copyright Research Triangle Institute January 2009
Release 10.0.1

DESIGN SUMMARY: Variances will be computed using the Taylor Linearization
Method, Assuming a With Replacement (WR) Design
Sample Weight: NCSRWTLG
Stratification Variables(s): SESTRAT
Primary Sampling Unit: SECLUSTR

Independence parameters have converged in 7 iterations

Number of observations read : 5692 Weighted count: 5692
Number of observations skipped : 3590
(WEIGHT variable nonpositive)
Observations used in the analysis : 5679 Weighted count: 5667
Denominator degrees of freedom : 42

Maximum number of estimable parameters for the model is 24

File NCSR contains 84 Clusters
84 clusters were used to fit the model
Maximum cluster size is 142 records
Minimum cluster size is 18 records

Sample and Population Counts for Response Variable REVWORKSTATUS
Based on observations used in the analysis

NLF	:	Sample Count	1630	Population Count	1706
Unemployed	:	Sample Count	283	Population Count	290
Employed	:	Sample Count	3766	Population Count	3671

-2 * Normalized Log-Likelihood with Intercepts Only : 9025.92
-2 * Normalized Log-Likelihood Full Model : 7367.23
Approximate Chi-Square (-2 * Log-L Ratio) : 1658.69
Degrees of Freedom : 22

Note: The approximate Chi-Square is not adjusted for clustering.
Refer to hypothesis test table for adjusted test.

Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Independent
Link Function: Generalized Logit
Response variable REVWORKSTATUS: REVWORKSTATUS
by: REVWORKSTATUS (log-odds), Independent Variables and Effects.

REVWORKSTATUS (log-odds) Independent Variables and Effects	Beta Coeff.	DEFF Beta #1	SE Beta	Lower 95% Limit Beta
NLF vs Employed				
Intercept	-0.38	1.80	0.17	-0.73
SEXM	-0.64	2.45	0.11	-0.86
ALD	0.33	0.77	0.13	0.07
MDE	0.10	1.00	0.09	-0.08
Years of education-4 categories				
0-11 Yrs	0.00	.	0.00	0.00
12 Yrs	-0.65	2.08	0.14	-0.94
13-15 Yrs	-0.92	2.01	0.15	-1.21
16+ Yrs	-1.23	2.10	0.16	-1.55
AG4CAT				
18-29	0.00	.	0.00	0.00
30-44	-0.32	1.49	0.13	-0.58
45-59	0.06	2.43	0.17	-0.28
60+	2.38	1.98	0.17	2.03
Marital Status-3 categories				
Married	0.00	.	0.00	0.00
Previously Married	-0.05	1.35	0.11	-0.26
Never Married	0.55	1.83	0.13	0.29

Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Independent
Link Function: Generalized Logit
Response variable REVWORKSTATUS: REVWORKSTATUS
by: REVWORKSTATUS (log-odds), Independent Variables and Effects.

REVWORKSTATUS (log-odds) Independent Variables and Effects	Upper 95% Limit Beta	T-Test B=0	P-value T-Test B=0
NLF vs Employed			
Intercept	-0.03	-2.19	0.0338
SEXM	-0.42	-5.82	0.0000
ALD	0.60	2.56	0.0142
MDE	0.28	1.12	0.2691
Years of education-4 categories			
0-11 Yrs	0.00	.	.
12 Yrs	-0.37	-4.62	0.0000
13-15 Yrs	-0.62	-6.26	0.0000
16+ Yrs	-0.91	-7.70	0.0000
AG4CAT			
18-29	0.00	.	.
30-44	-0.06	-2.46	0.0182
45-59	0.41	0.38	0.7056
60+	2.73	13.73	0.0000
Marital Status-3 categories			
Married	0.00	.	.
Previously Married	0.16	-0.50	0.6213
Never Married	0.82	4.18	0.0001

Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Independent
Link Function: Generalized Logit
Response variable REVWORKSTATUS: REVWORKSTATUS
by: REVWORKSTATUS (log-odds), Independent Variables and Effects.

REVWORKSTATUS (log-odds) Independent Variables and Effects	Beta Coeff.	DEFF Beta #1	SE Beta	Lower 95% Limit Beta
Unemployed vs Employed				
Intercept	-0.64	1.58	0.30	-1.24
SEXM	-1.39	1.78	0.20	-1.79
ALD	-0.16	0.78	0.36	-0.88
MDE	-0.14	0.70	0.16	-0.46
Years of education-4 categories				
0-11 Yrs	0.00	.	0.00	0.00
12 Yrs	-0.85	1.94	0.24	-1.32
13-15 Yrs	-1.37	1.69	0.26	-1.88
16+ Yrs	-1.73	1.94	0.31	-2.36
AG4CAT				
18-29	0.00	.	0.00	0.00
30-44	-0.85	1.61	0.29	-1.45
45-59	-0.84	1.16	0.26	-1.36
60+	1.83	1.87	0.29	1.23
Marital Status-3 categories				
Married	0.00	.	0.00	0.00
Previously Married	-0.59	2.01	0.23	-1.04
Never Married	-2.78	0.67	0.38	-3.55

Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Independent
Link Function: Generalized Logit
Response variable REVWORKSTATUS: REVWORKSTATUS
by: REVWORKSTATUS (log-odds), Independent Variables and Effects.

REVWORKSTATUS (log-odds) Independent Variables and Effects	Upper 95% Limit Beta	T-Test B=0	P-value T-Test B=0
Unemployed vs Employed			
Intercept	-0.05	-2.17	0.0354
SEXM	-0.99	-7.05	0.0000
ALD	0.56	-0.46	0.6487
MDE	0.18	-0.89	0.3792
Years of education-4 categories			
0-11 Yrs	0.00	.	.
12 Yrs	-0.37	-3.60	0.0008
13-15 Yrs	-0.85	-5.30	0.0000
16+ Yrs	-1.10	-5.57	0.0000
AG4CAT			
18-29	0.00	.	.
30-44	-0.26	-2.89	0.0060
45-59	-0.32	-3.25	0.0023
60+	2.42	6.20	0.0000
Marital Status-3 categories			
Married	0.00	.	.
Previously Married	-0.14	-2.62	0.0122
Never Married	-2.02	-7.32	0.0000

Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Independent
Link Function: Generalized Logit
Response variable REVWORKSTATUS: REVWORKSTATUS
by: Contrast.

Contrast	Degrees of Freedom	Adj Wald F	P-value Adj Wald F
OVERALL MODEL	24	82.36	0.0000
MODEL MINUS INTERCEPT	22	73.91	0.0000
INTERCEPT	.	.	.
SEXM	2	35.75	0.0000
ALD	2	5.05	0.0110
MDE	2	1.14	0.3302
ED4CAT	6	13.68	0.0000
AG4CAT	6	83.59	0.0000
MAR3CAT	4	24.81	0.0000

Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Independent
Link Function: Generalized Logit
Response variable REVWORKSTATUS: REVWORKSTATUS
by: REVWORKSTATUS (log-odds), Independent Variables and Effects.

REVWORKSTATUS (log-odds) Independent Variables and Effects	Odds Ratio	Lower 95% Limit OR	Upper 95% Limit OR
NLF vs Employed			
Intercept	0.68	0.48	0.97
SEXM	0.53	0.42	0.66
ALD	1.40	1.07	1.82
MDE	1.10	0.92	1.32
Years of education-4 categories			
0-11 Yrs	1.00	1.00	1.00
12 Yrs	0.52	0.39	0.69
13-15 Yrs	0.40	0.30	0.54
16+ Yrs	0.29	0.21	0.40
AG4CAT			
18-29	1.00	1.00	1.00
30-44	0.73	0.56	0.95
45-59	1.07	0.76	1.51
60+	10.81	7.62	15.34
Marital Status-3 categories			
Married	1.00	1.00	1.00
Previously Married	0.95	0.77	1.17
Never Married	1.74	1.33	2.27

Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Independent
Link Function: Generalized Logit
Response variable REVWORKSTATUS: REVWORKSTATUS
by: REVWORKSTATUS (log-odds), Independent Variables and Effects.

REVWORKSTATUS (log-odds) Independent Variables and Effects	Odds Ratio	Lower 95% Limit OR	Upper 95% Limit OR
Unemployed vs Employed			
Intercept	0.53	0.29	0.95
SEXM	0.25	0.17	0.37
ALD	0.85	0.41	1.74
MDE	0.87	0.63	1.19
Years of education-4 categories			
0-11 Yrs	1.00	1.00	1.00
12 Yrs	0.43	0.27	0.69
13-15 Yrs	0.26	0.15	0.43
16+ Yrs	0.18	0.09	0.33
AG4CAT			
18-29	1.00	1.00	1.00
30-44	0.43	0.24	0.77
45-59	0.43	0.26	0.73
60+	6.22	3.43	11.28
Marital Status-3 categories			
Married	1.00	1.00	1.00
Previously Married	0.55	0.35	0.87
Never Married	0.06	0.03	0.13

```
title "Analysis Example 9.3: Ordinal Logistic Regression : HRS " ;
proc multilog data=hrs filetype=sas deft1 ;
nest stratum secu  ;
weight kwgtr ;
class revselfrhealth gender / nofreq ;
reflevel gender=2 ;
rformat gender gf. ;
rformat revselfrhealth revsrh. ;
model revselfrhealth = kage gender / cumlogit ;
setenv decwidth=3 ;
test adjwaldf ;
print / style=nchs  ;
run ;
```

```
Analysis Example 9.3: Ordinal Logistic Regression : HRS

                                S U D A A N
           Software for the Statistical Analysis of Correlated Data
         Copyright      Research Triangle Institute      January 2009
                               Release 10.0.1


DESIGN SUMMARY: Variances will be computed using the Taylor Linearization
Method, Assuming a With Replacement (WR) Design
    Sample Weight: KWGTR
    Stratification Variables(s): STRATUM
    Primary Sampling Unit: SECU


Independence parameters have converged in 4 iterations

Number of observations read       :  16954    Weighted count: 76540667
Number of observations skipped    :   1513
(WEIGHT variable nonpositive)
Observations used in the analysis :  16930    Weighted count: 76444941
Denominator degrees of freedom    :     56


Maximum number of estimable parameters for the model is  6

File HRS contains  112 Clusters
 112 clusters were used to fit the model
Maximum cluster size is 400 records
Minimum cluster size is  10 records


Sample and Population Counts for Response Variable REVSELFRHEALTH
Based on observations used in the analysis
  Poor     :  Sample Count     1422    Population Count   5917389
  Fair     :  Sample Count     3594    Population Count  14551146
  Good     :  Sample Count     5225    Population Count  22848636
  Very Good:  Sample Count     4856    Population Count  23387921
  Excellent:  Sample Count     1833    Population Count   9739849

-2 * Normalized Log-Likelihood with Intercepts Only : 50778.02
-2 * Normalized Log-Likelihood Full Model           : 50294.64
Approximate Chi-Square (-2 * Log-L Ratio)           :   483.38
Degrees of Freedom                                  :        2

Note: The approximate Chi-Square is not adjusted for clustering.
      Refer to hypothesis test table for adjusted test.
```

Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Independent
Link Function: Cumulative Logit
Response variable REVSELFRHEALTH: REVSELFRHEALTH
by: REVSELFRHEALTH (cum-logit), Independent Variables and Effects.

REVSELFRHEALTH (cum-logit), Independent Variables and Effects	Beta Coeff.	DEFF Beta #1	SE Beta	Lower 95% Limit Beta
REVSELFRHEALTH (cum-logit)				
Intercept 1: Poor	-4.405	2.791	0.165	-4.736
Intercept 2: Fair	-2.917	2.865	0.159	-3.235
Intercept 3: Good	-1.614	2.783	0.153	-1.920
Intercept 4: Very Good	0.071	2.747	0.153	-0.236
age at 2006 interview	0.029	2.680	0.002	0.024
gender				
Male	-0.071	1.359	0.032	-0.135
Female	0.000	.	0.000	0.000

Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Independent
Link Function: Cumulative Logit
Response variable REVSELFRHEALTH: REVSELFRHEALTH
by: REVSELFRHEALTH (cum-logit), Independent Variables and Effects.

REVSELFRHEALTH (cum-logit), Independent Variables and Effects	Upper 95% Limit Beta	T-Test B=0	P-value T-Test B=0
REVSELFRHEALTH (cum-logit)			
Intercept 1: Poor	-4.074	-26.650	0.000
Intercept 2: Fair	-2.599	-18.367	0.000
Intercept 3: Good	-1.308	-10.560	0.000
Intercept 4: Very Good	0.377	0.463	0.645
age at 2006 interview	0.033	13.228	0.000
gender			
Male	-0.006	-2.186	0.033
Female	0.000	.	.

Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Independent
Link Function: Cumulative Logit
Response variable REVSELFRHEALTH: REVSELFRHEALTH
by: Contrast.

Contrast	Degrees of Freedom	Adj Wald F	P-value Adj Wald F
OVERALL MODEL	6.000	1459.062	0.000
MODEL MINUS INTERCEPT	2.000	90.206	0.000
KAGE	1.000	174.992	0.000
GENDER	1.000	4.780	0.033

Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Independent
Link Function: Cumulative Logit
Response variable REVSELFRHEALTH: REVSELFRHEALTH
by: REVSELFRHEALTH (cum-logit), Independent Variables and Effects.

REVSELFRHEALTH (cum-logit), Independent Variables and Effects	Odds Ratio	Lower 95% Limit OR	Upper 95% Limit OR
REVSELFRHEALTH (cum-logit)			
Intercept 1: Poor	0.012	0.009	0.017
Intercept 2: Fair	0.054	0.039	0.074
Intercept 3: Good	0.199	0.147	0.270
Intercept 4: Very Good	1.073	0.790	1.459
age at 2006 interview	1.029	1.025	1.034
gender			
Male	0.932	0.873	0.994
Female	1.000	1.000	1.000

```
title "Analysis Example 9.4: Poisson Regression : HRS" ;
proc loglink data=hrs filetype=sas ;
nest stratum secu  ;
weight kwgtr ;
subpopn age65p=1 ;
class age3cat  gender / nofreq ;
rformat age3cat agf. ; rformat gender gf. ;
reflevel age3cat=1 gender=2;
model numfalls24 = gender age3cat arthritis diabetes bodywgt totheight / offset=offset24 ;
setenv decwidth=4 ;
print idr lowidr upidr / betas=all tests=all style=nchs ;
run ;
```

```
Analysis Example 9.4: Poisson Regression : HRS



                              S U D A A N
           Software for the Statistical Analysis of Correlated Data
         Copyright      Research Triangle Institute      January 2009
                               Release 10.0.1


DESIGN SUMMARY: Variances will be computed using the Taylor Linearization Method, Assuming a With Replacement (WR) Design
    Sample Weight: KWGTR
    Stratification Variables(s): STRATUM
    Primary Sampling Unit: SECU


Independence parameters have converged in 10 iterations

Number of observations read       :  16954    Weighted count: 76540667
Number of observations skipped    :   1513
(WEIGHT variable nonpositive)
Observations in subpopulation     :  11206    Weighted count: 37449807
Observations used in the analysis :  10440    Weighted count: 35017430
Denominator degrees of freedom    :     56


Maximum number of estimable parameters for the model is  8

File HRS contains  112 Clusters
 104 clusters were used to fit the model
Maximum cluster size is 279 records
Minimum cluster size is  25 records

Weighted mean response is 0.868776


-2 * Normalized Log-Likelihood with Intercepts Only : -380262.85
-2 * Normalized Log-Likelihood Full Model           : -19677.36
Approximate Chi-Square (-2 * Log-L Ratio)          : 360585.49
Degrees of Freedom                                 :        7

Note: The approximate Chi-Square is not adjusted for clustering.
      Refer to hypothesis test table for adjusted test.
```

Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Independent
Link Function: Log
Response variable NUMFALLS24: NUMFALLS24
Offset variable OFFSET24: OFFSET24
For Subpopulation: AGE65P = 1
by: Independent Variables and Effects.

Independent Variables and Effects	Beta Coeff.	DEFF Beta #4	SE Beta	Lower 95% Limit Beta	Upper 95% Limit Beta	T-Test B=0	P-value T-Test B=0	Var Beta
Intercept	-2.6842	6.9614	0.6359	-3.9582	-1.4103	-4.2209	0.0001	0.4044
gender								
Male	0.1831	12.1602	0.1073	-0.0319	0.3982	1.7059	0.0936	0.0115
Female	0.0000	.	0.0000	0.0000	0.0000	.	.	0.0000
AGE3CAT								
65-74	0.0000	.	0.0000	0.0000	0.0000	.	.	0.0000
75-84	0.2384	4.9548	0.0535	0.1313	0.3455	4.4592	0.0000	0.0029
85+	0.5839	9.0511	0.0900	0.4036	0.7641	6.4895	0.0000	0.0081
ARTHRITIS	0.4867	9.9019	0.0824	0.3216	0.6518	5.9055	0.0000	0.0068
DIABETES	0.2596	7.7047	0.0689	0.1215	0.3977	3.7664	0.0004	0.0048
BODYWGT	0.0009	6.7457	0.0009	-0.0008	0.0027	1.0437	0.3011	0.0000
TOTHEIGHT	-0.0224	7.7515	0.0110	-0.0445	-0.0003	-2.0336	0.0467	0.0001

Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Independent
Link Function: Log
Response variable NUMFALLS24: NUMFALLS24
Offset variable OFFSET24: OFFSET24
For Subpopulation: AGE65P = 1
by: Contrast.

Contrast	Degrees of Freedom	S_waite Adj DF	S_waite Adj F	P-value S_waite Adj F	S_waite Adj ChiSq	P-value S_waite ChiSq	Wald F	P-value Wald F
OVERALL MODEL	8.0000	5.4616	1233.5408	0.0000	6737.0882	0.0000	711.5800	0.0000
MODEL MINUS INTERCEPT	7.0000	5.2234	17.3084	0.0000	90.4084	0.0000	17.4344	0.0000
INTERCEPT	.	.	.	.	.	.	.	.
GENDER	1.0000	1.0000	2.9102	0.0936	2.9102	0.0881	2.9102	0.0936
AGE3CAT	2.0000	1.6962	23.6297	0.0000	40.0811	0.0000	27.4625	0.0000
ARTHRITIS	1.0000	1.0000	34.8747	0.0000	34.8747	0.0000	34.8747	0.0000
DIABETES	1.0000	1.0000	14.1861	0.0004	14.1861	0.0002	14.1861	0.0004
BODYWGT	1.0000	1.0000	1.0893	0.3011	1.0893	0.2967	1.0893	0.3011
TOTHEIGHT	1.0000	1.0000	4.1355	0.0467	4.1355	0.0420	4.1355	0.0467

Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Independent
Link Function: Log
Response variable NUMFALLS24: NUMFALLS24
Offset variable OFFSET24: OFFSET24
For Subpopulation: AGE65P = 1
by: Contrast.

Contrast	Adj Wald F	P-value Adj Wald F	Wald ChiSq	P-value Wald ChiSq
OVERALL MODEL	622.6325	0.0000	5692.6398	0.0000
MODEL MINUS INTERCEPT	15.5664	0.0000	122.0407	0.0000
INTERCEPT	.	.	.	.
GENDER	2.9102	0.0936	2.9102	0.0880
AGE3CAT	26.9721	0.0000	54.9250	0.0000
ARTHRITIS	34.8747	0.0000	34.8747	0.0000
DIABETES	14.1861	0.0004	14.1861	0.0002
BODYWGT	1.0893	0.3011	1.0893	0.2966
TOTHEIGHT	4.1355	0.0467	4.1355	0.0420

Variance Estimation Method: Taylor Series (WR)
SE Method: Robust (Binder, 1983)
Working Correlations: Independent
Link Function: Log
Response variable NUMFALLS24: NUMFALLS24
Offset variable OFFSET24: OFFSET24
For Subpopulation: AGE65P = 1
by: Independent Variables and Effects.

Independent Variables and Effects	Incidence Density Ratio	Lower 95% Limit IDR	Upper 95% Limit IDR
Intercept	0.0683	0.0191	0.2441
gender			
Male	1.2010	0.9686	1.4891
Female	1.0000	1.0000	1.0000
AGE3CAT			
65-74	1.0000	1.0000	1.0000
75-84	1.2692	1.1403	1.4127
85+	1.7930	1.4973	2.1471
ARTHRITIS	1.6270	1.3794	1.9190
DIABETES	1.2964	1.1292	1.4884
BODYWGT	1.0009	0.9992	1.0027
TOTHEIGHT	0.9778	0.9564	0.9997