CHAPTER 9 ANALYSIS EXAMPLES REPLICATION SAS V9.2

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.

GENERAL NOTES ON MULTINOMIAL AND ORDINAL LOGISTIC REGRESSION AND PROC SURVEYLOGISTIC

PROC SURVEYLOGISTIC is the general purpose tool for survey data logistic regression. This procedure is a multi-purpose tool that can perform correct subpopulation analyses and offers a number of output options such as a class statement for categorical predictors, and a test statement for custom hypothesis testing of parameters.

This chapter focuses on multinomial and ordinal logit regression with nominal and ordinal outcomes. Use of PROC SURVEYLOGISTIC with the appropriate link option is shown. For example, for multinomial logit regression use of the glogit link is shown along with the default logit link for ordinal logistic regression. This chapter also includes analytic examples of count models such as Poisson or negative binomial but SAS v9.2 does not have the ability to perform these analyses with appropriate complex sample survey adjustments. Therefore, examples of Poisson, negative binomial and zero-inflated Poisson and negative binomial models are omitted here.

Some options to note: use of the class statement requires the / param=ref specification if you want to use a reference group parameterization instead of the default effects coding approach, use of the (ref=first) allows specification of the omitted category for the class variables, and use of the (event='1') syntax declares the probability modeled for the outcome variable. There are examples of these options in this chapter. PROC SURVEYLOGISTIC also allows the use of the test statement and use of the crossing operator (*) for interaction variables in the model statement.

```
* CHAPTER 9 GENERALIZED LINEAR MODELS;
*note this example does not use formats for the predictors as it is difficult to then use formatted variables in the test
statement;

title "Analysis Example 9.2: Multinomial Logistic Regression : NCSR" ;
proc surveylogistic data=ncsr ;
strata sestrat ;
cluster seclustr ;
weight ncsrwtlg ;
class sex (ref='2') ed4cat (ref='1') ag4cat (ref='1') mar3cat (ref='1') / param=ref ;
model wkstat3c (descending)  =  ag4cat sex ald mde ed4cat mar3cat / link=glogit ;
testeduc: test ed4cat2_2=ed4cat2_3, ed4cat3_2=ed4cat3_3, ed4cat4_2=ed4cat4_3 ;
run ;
```

Analysis Example 9.2: Multinomial Logistic Regression : NCSR

The SURVEYLOGISTIC Procedure

Model Information

Data Set	WORK.NCSR	
Response Variable	WKSTAT3C	Work Status 3 categories
Number of Response Levels	3	
Stratum Variable	SESTRAT	SAMPLING ERROR STRATUM
Number of Strata	42	
Cluster Variable	SECLUSTR	SAMPLING ERROR CLUSTER
Number of Clusters	84	
Weight Variable	NCSRWTLG	NCSR sample part 2 weight
Model	Generalized Logit	
Optimization Technique	Newton-Raphson	
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	5679
Sum of Weights Read	5692
Sum of Weights Used	5667.185

Response Profile

Ordered Value	WKSTAT3C	Total Frequency	Total Weight
1	3	1630	1705.8959
2	2	283	289.8166
3	1	3766	3671.4725

Logits modeled use WKSTAT3C=1 as the reference category.

NOTE: 2649 observations were deleted due to missing values for the response or explanatory variables.

NOTE: 954 observations having nonpositive frequencies or weights were excluded since they do not contribute to the analysis.

Class Level Information

Class	Value	Design Variables		
SEX	1	1		
	2	0		
ED4CAT	1	0	0	0
	2	1	0	0
	3	0	1	0
	4	0	0	1

ag4cat	1	0	0	0
	2	1	0	0
	3	0	1	0
	4	0	0	1
MAR3CAT	1	0	0	
	2	1	0	
	3	0	1	

Model Convergence Status

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

Analysis Example 9.2: Multinomial Logistic Regression : NCSR

The SURVEYLOGISTIC Procedure

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	9011.140	7399.903
SC	9024.429	7559.372
-2 Log L	9007.140	7351.903

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	1655.2366	22	<.0001
Score	1561.6117	22	<.0001
Wald	3239.2937	22	<.0001

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
ag4cat	6	567.0168	<.0001
SEX	2	72.9590	<.0001
ald	2	10.3013	0.0058
mde	2	2.3233	0.3130
ED4CAT	6	92.8002	<.0001
MAR3CAT	4	106.5049	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	WKSTAT3C	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	3	1	-0.3795	0.1733	4.7963	0.0285
Intercept	2	1	-0.6438	0.2967	4.7093	0.0300
ag4cat 2	3	1	-0.3164	0.1290	6.0134	0.0142
ag4cat 2	2	1	-0.8524	0.2951	8.3414	0.0039
ag4cat 3	3	1	0.0650	0.1712	0.1441	0.7043
ag4cat 3	2	1	-0.8377	0.2586	10.4918	0.0012
ag4cat 4	3	1	2.3806	0.1738	187.7225	<.0001
ag4cat 4	2	1	1.8284	0.2953	38.3406	<.0001
SEX 1	3	1	-0.6403	0.1103	33.7167	<.0001
SEX 1	2	1	-1.3932	0.1980	49.4922	<.0001
ald	3	1	0.3332	0.1305	6.5203	0.0107
ald	2	1	-0.1638	0.3577	0.2097	0.6470
mde	3	1	0.0985	0.0882	1.2491	0.2637
mde	2	1	-0.1398	0.1576	0.7868	0.3751
ED4CAT 2	3	1	-0.6514	0.1413	21.2510	<.0001

ED4CAT	2	2	1	-0.8470	0.2359	12.8962	0.0003
ED4CAT	3	3	1	-0.9169	0.1468	39.0196	<.0001
ED4CAT	3	2	1	-1.3653	0.2580	27.9968	<.0001
ED4CAT	4	3	1	-1.2295	0.1599	59.1139	<.0001
ED4CAT	4	2	1	-1.7310	0.3111	30.9556	<.0001
MAR3CAT	2	3	1	-0.0523	0.1052	0.2467	0.6194
MAR3CAT	2	2	1	-0.5899	0.2257	6.8322	0.0090
MAR3CAT	3	3	1	0.5528	0.1326	17.3709	<.0001
MAR3CAT	3	2	1	-2.7834	0.3807	53.4429	<.0001

Analysis Example 9.2: Multinomial Logistic Regression : NCSR

The SURVEYLOGISTIC Procedure

Odds Ratio Estimates

Effect		WKSTAT3C	Point Estimate	95% Wald Confidence Limits	
ag4cat	2 vs 1	3	0.729	0.566	0.938
ag4cat	2 vs 1	2	0.426	0.239	0.760
ag4cat	3 vs 1	3	1.067	0.763	1.493
ag4cat	3 vs 1	2	0.433	0.261	0.718
ag4cat	4 vs 1	3	10.811	7.691	15.198
ag4cat	4 vs 1	2	6.224	3.489	11.102
SEX	1 vs 2	3	0.527	0.425	0.654
SEX	1 vs 2	2	0.248	0.168	0.366
ald		3	1.395	1.081	1.802
ald		2	0.849	0.421	1.711
mde		3	1.104	0.928	1.312
mde		2	0.870	0.639	1.184
ED4CAT	2 vs 1	3	0.521	0.395	0.688
ED4CAT	2 vs 1	2	0.429	0.270	0.681
ED4CAT	3 vs 1	3	0.400	0.300	0.533
ED4CAT	3 vs 1	2	0.255	0.154	0.423
ED4CAT	4 vs 1	3	0.292	0.214	0.400
ED4CAT	4 vs 1	2	0.177	0.096	0.326
MAR3CAT	2 vs 1	3	0.949	0.772	1.166
MAR3CAT	2 vs 1	2	0.554	0.356	0.863
MAR3CAT	3 vs 1	3	1.738	1.340	2.254
MAR3CAT	3 vs 1	2	0.062	0.029	0.130

Linear Hypotheses Testing Results

Label	Wald Chi-Square	DF	Pr > ChiSq
testeduc	3.9353	3	0.2685

```
title "Analysis Example 9.3: Ordinal Logistic Regression: HRS Data" ;
* note: no deff option for design effects in PROC SURVEYLOGISTIC ;
proc surveylogistic data=hrs ;
strata stratum;
cluster secu;
weight kwgtr;
class gender (ref='2') / param=ref ;
model selfrhealth (descending)= kage gender ;
run ;
```

```
NOTE: CODES FOR GENDER 1=MALE 2=FEMALE.
```

Analysis Example 9.3: Ordinal Logistic Regression: HRS

The SURVEYLOGISTIC Procedure

Model Information

Data Set	WORK.HRS	
Response Variable	selfrhealth	
Number of Response Levels	5	
Stratum Variable	STRATUM	stratum id
Number of Strata	56	
Cluster Variable	SECU	sampling error computation unit
Number of Clusters	112	
Weight Variable	KWGTR	2006 weight: respondent level
Model	Cumulative 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	18467
Number of Observations Used	16930
Sum of Weights Read	76540667
Sum of Weights Used	76444941

Response Profile

Ordered Value	selfrhealth	Total Frequency	Total Weight
1	5	1422	5917389
2	4	3594	14551146
3	3	5225	22848636
4	2	4856	23387921
5	1	1833	9739849

Probabilities modeled are cumulated over the lower Ordered Values.

NOTE: 25 observations were deleted due to missing values for the response or explanatory variables.

NOTE: 1512 observations having nonpositive frequencies or weights were excluded since they do not contribute to the analysis.

Class Level Information

Class	Value	Design Variables
GENDER	1	1
	2	0

Model Convergence Status

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

Score Test for the Proportional Odds Assumption

Chi-Square	DF	Pr > ChiSq
199756.999	6	<.0001

Analysis Example 9.3: Ordinal Logistic Regression : HRS

The SURVEYLOGISTIC Procedure

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	229280756	227098119
SC	229280787	227098165
-2 Log L	229280748	227098107

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	2182640.85	2	<.0001
Score	2177931.73	2	<.0001
Wald	180.7676	2	<.0001

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
KAGE	1	172.8261	<.0001
GENDER	1	4.7741	0.0289

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept 5	1	-4.4053	0.1653	710.5710	<.0001
Intercept 4	1	-2.9167	0.1585	338.5922	<.0001
Intercept 3	1	-1.6142	0.1526	111.8772	<.0001
Intercept 2	1	0.0709	0.1531	0.2143	0.6434
KAGE	1	0.0288	0.00219	172.8261	<.0001
GENDER 1	1	-0.0707	0.0323	4.7741	0.0289

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
KAGE	1.029	1.025	1.034
GENDER 1 vs 2	0.932	0.875	0.993

Association of Predicted Probabilities and Observed Responses

Percent Concordant	55.5	Somers' D	0.125
Percent Discordant	43.0	Gamma	0.127
Percent Tied	1.4	Tau-a	0.095
Pairs	108722365	c	0.562