CHAPTER 9 ASDA ANALYSIS EXAMPLES REPLICATION-MPLUS 5.21

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 MULTINOMIAL LOGIT, ORDINAL AND POISSON REGRESSION ANALYSIS IN MPLUS 5.21

The analysis replication examples were all run using Mplus 5.21. Mplus is an advanced modeling tool and offers the ability to correctly account for complex sample survey data for all analytic techniques.

Mplus can perform nearly all of the modeling tasks presented in Chapter 9 of ASDA including multinomial logit, ordinal and Poisson regression but not negative binomial regression. Some of the fine points of this tool are use of a unique cluster variable with a different value for each person in the data set, use of a SUBPOPULATION statement for subpopulation analyses, use of TYPE=COMPLEX and ESTIMATOR=MLR on the ANALYSIS command, and a MODELTEST statement for linear contrasts providing a Wald ChiSq test for selected parameter estimates.

Because the MODEL TEST statement does not offer an easy way to do significance testing with multiple category outcomes, the bivariate testing process is omitted here for the multinomial model. Als, due to the difficulty of running multi-parameter tests in Mplus (requires a separate run with the use of MODEL TEST: for each group of variables tested), this technique is not demonstrated for each group of parameters in the multinomial model. See Chapter 8 logistic examples for how to do this and also see the Mplus User’s Guide for additional detail.

ANALYSIS EXAMPLE 9.2 MULTINOMIAL LOGIT NCS-R DATA

Mplus VERSION 5.21
MUTHEN & MUTHEN
04/07/2010 12:10 PM

INPUT INSTRUCTIONS

```
  TITLE: ANALYSIS EXAMPLE 9.2 NCSR DATA MULTINOMIAL LOGIT
  DATA:
    FILE IS "F:\applied_analysis_book\Mplus\ncsr.txt";

  VARIABLE:
    NAMES ARE MAR3CAT ED4CAT NCSRWTSH NCSRWTLG SEX HHINC WKSTAT3C
    SESTRAT SECLUSTR bmi mde sexf sexm ald
      racecat povcat agecentered age29 ag4cat ed011 ed12 ed1315 ed16 black
      white hispanic
      asianother age51 agecat revworkstatus numsecu ag29 ag44 ag59 ag60
      prevmar nevmar  ;
    USEVARIABLES ARE wkstat3c NCSRWTSH sexm sestrat numsecu ald mde
    ed011 ed12 ed1315 prevmar nevmar ag44 ag59 ag60 wkst3cr ;

    missing are . ;

    WEIGHT IS NCSRWTlg ;
    stratification is sestrat ;
    cluster is numsecu ;
    nominal are wkst3cr ;

  DEFINE:
      if wkstat3c==1 then wkst3cr=3 ;
      if wkstat3c==2 then wkst3cr=2 ;
      if wkstat3c==3 then wkst3cr=1 ;

  ANALYSIS:
    type is complex;
    estimator is mlr ;

  Model:
    wkst3cr on sexm ald mde ed011 ed12 ed1315 ag44 ag59 ag60 prevmar nevmar ;
```

SUMMARY OF ANALYSIS

Number of groups 1
Number of observations 9282

Number of dependent variables 3
Number of independent variables 11
Number of continuous latent variables 0

Observed dependent variables

  Continuous
   WKSTAT3C   NCSRWTSH

  Unordered categorical (nominal)
   WKST3CR

Observed independent variables
   SEXM   ALD   MDE   EDO11   ED12   ED1315
   PREVMAR   NEVMAR   AG44   AG59   AG60

Variables with special functions

  Stratification   SESTRAT
  Cluster variable   NUMSECU
  Weight variable   NCSRWTLG

Estimator MLR
Information matrix OBSERVED
Optimization Specifications for the Quasi-Newton Algorithm for
Continuous Outcomes
  Maximum number of iterations 100
  Convergence criterion 0.100D-05
Optimization Specifications for the EM Algorithm
  Maximum number of iterations 500
  Convergence criteria
    Loglikelihood change 0.100D-02
    Relative loglikelihood change 0.100D-05
    Derivative 0.100D-02
Optimization Specifications for the M step of the EM Algorithm for
Categorical Latent variables
  Number of M step iterations 1
  M step convergence criterion 0.100D-02
  Basis for M step termination ITERATION
Optimization Specifications for the M step of the EM Algorithm for
Censored, Binary or Ordered Categorical (Ordinal), Unordered
Categorical (Nominal) and Count Outcomes
  Number of M step iterations 1
  M step convergence criterion 0.100D-02
  Basis for M step termination ITERATION
  Maximum value for logit thresholds 15
  Minimum value for logit thresholds -15
  Minimum expected cell size for chi-square 0.100D-01
Maximum number of iterations for H1 2000
Convergence criterion for H1 0.100D-03
Optimization algorithm EMA
Integration Specifications
  Type STANDARD
  Number of integration points 15
  Dimensions of numerical integration 0
  Adaptive quadrature ON
Cholesky OFF

Input data file(s)
  F:\applied_analysis_book\Mplus\ncsr.txt
Input data format FREE

SUMMARY OF DATA

Number of missing data patterns	2
Number of strata	42
Number of clusters	84

COVARIANCE COVERAGE OF DATA

Minimum covariance coverage value 0.100

PROPORTION OF DATA PRESENT FOR Y

Covariance Coverage

	WKSTAT3C	NCSRWTSH	SEXM	ALD	MDE
WKSTAT3C	0.715				
NCSRWTSH	0.715	1.000			
SEXM	0.715	1.000	1.000		
ALD	0.715	1.000	1.000	1.000	
MDE	0.715	1.000	1.000	1.000	1.000
ED011	0.715	1.000	1.000	1.000	1.000
ED12	0.715	1.000	1.000	1.000	1.000
ED1315	0.715	1.000	1.000	1.000	1.000
PREVMAR	0.715	1.000	1.000	1.000	1.000
NEVMAR	0.715	1.000	1.000	1.000	1.000
AG44	0.715	1.000	1.000	1.000	1.000
AG59	0.715	1.000	1.000	1.000	1.000
AG60	0.715	1.000	1.000	1.000	1.000

Covariance Coverage

	ED011	ED12	ED1315	PREVMAR	NEVMAR
ED011	1.000				
ED12	1.000	1.000			
ED1315	1.000	1.000	1.000		
PREVMAR	1.000	1.000	1.000	1.000	
NEVMAR	1.000	1.000	1.000	1.000	1.000
AG44	1.000	1.000	1.000	1.000	1.000
AG59	1.000	1.000	1.000	1.000	1.000
AG60	1.000	1.000	1.000	1.000	1.000

Covariance Coverage

	AG44	AG59	AG60
AG44	1.000		
AG59	1.000	1.000	
AG60	1.000	1.000	1.000

THE MODEL ESTIMATION TERMINATED NORMALLY

TESTS OF MODEL FIT

Loglikelihood

H0 Value	-29231.654
H0 Scaling Correction Factor for MLR	30.822

Information Criteria

Number of Free Parameters	28
Akaike (AIC)	58519.307
Bayesian (BIC)	58719.111
Sample-Size Adjusted BIC (n* = (n + 2) / 24)	58630.131

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
WKST3CR#1 ON				
SEXM	-0.640	0.110	-5.818	0.000
ALD	0.333	0.130	2.559	0.011
MDE	0.099	0.088	1.120	0.263
ED011	1.230	0.160	7.704	0.000
ED12	0.578	0.139	4.162	0.000
ED1315	0.313	0.135	2.309	0.021
AG44	-0.316	0.129	-2.457	0.014
AG59	0.065	0.171	0.380	0.704
AG60	2.381	0.173	13.729	0.000
PREVMAR	-0.052	0.105	-0.498	0.619
NEVMAR	0.553	0.132	4.176	0.000
WKST3CR#2 ON				
SEXM	-1.393	0.198	-7.049	0.000
ALD	-0.164	0.357	-0.459	0.646
MDE	-0.140	0.157	-0.889	0.374
ED011	1.731	0.310	5.575	0.000
ED12	0.884	0.177	4.982	0.000
ED1315	0.366	0.241	1.517	0.129
AG44	-0.852	0.295	-2.894	0.004
AG59	-0.838	0.258	-3.246	0.001
AG60	1.828	0.295	6.204	0.000
PREVMAR	-0.590	0.225	-2.619	0.009
NEVMAR	-2.785	0.380	-7.323	0.000
Means				
WKSTAT3C	1.653	0.019	86.686	0.000
NCSRWTSH	1.261	0.107	11.827	0.000
Intercepts				
WKST3CR#1	-1.609	0.191	-8.434	0.000
WKST3CR#2	-2.375	0.323	-7.345	0.000
Variances				
WKSTAT3C	0.829	0.013	62.571	0.000
NCSRWTSH	0.626	0.229	2.728	0.006

LOGISTIC REGRESSION ODDS RATIO RESULTS

WKST3CR#1 ON	
SEXM	0.527
ALD	1.395
MDE	1.104
ED011	3.420
ED12	1.783
ED1315	1.367
AG44	0.729
AG59	1.067

AG60	10.811
PREVMAR	0.949
NEVMAR	1.738

WKST3CR#2 ON

SEXM	0.248
ALD	0.849
MDE	0.870
ED011	5.646
ED12	2.420
ED1315	1.441
AG44	0.426
AG59	0.433
AG60	6.224
PREVMAR	0.554
NEVMAR	0.062

QUALITY OF NUMERICAL RESULTS

Condition Number for the Information Matrix (ratio of smallest to largest eigenvalue) 0.618E-03

Beginning Time: 12:10:05
Ending Time: 12:10:08
Elapsed Time: 00:00:03

ANALYSIS EXAMPLE 9.3 ORDINAL LOGISTIC REGRESSION HRS DATA

Mplus VERSION 5.21
MUTHEN & MUTHEN
04/07/2010   4:00 PM

INPUT INSTRUCTIONS

```
  TITLE:  ANALYSIS EXAMPLE 9.3 HRS DATA ORDINAL REGRESSION
  DATA:
    FILE IS "F:\applied_analysis_book\Mplus\hrs.txt";

  VARIABLE:
    NAMES ARE numsecu kwgtr stratum secu selfrhealth kage male ;

   USEVARIABLES ARE kwgtr selfrhealth kage male ;
   Missing are . ;
   WEIGHT IS kwgtr ;
   stratification is stratum ;
   cluster is numsecu ;
   categorical are selfrhealth  ;

  ANALYSIS:
    type is complex;
    estimator is mlr ;

  Model:
    selfrhealth on kage male ;
```

```
SUMMARY OF ANALYSIS

Number of groups                                                 1
Number of observations                                       18442

Number of dependent variables                                    1
Number of independent variables                                  2
Number of continuous latent variables                            0

Observed dependent variables

  Binary and ordered categorical (ordinal)
   SELFRHEA

Observed independent variables
   KAGE        MALE

Variables with special functions

  Stratification        STRATUM
  Cluster variable      NUMSECU
  Weight variable       KWGTR

Estimator                                                      MLR
Information matrix                                        OBSERVED
Optimization Specifications for the Quasi-Newton Algorithm for
Continuous Outcomes
  Maximum number of iterations                                 100
  Convergence criterion                                0.100D-05
Optimization Specifications for the EM Algorithm
  Maximum number of iterations                                 500
  Convergence criteria
    Loglikelihood change                               0.100D-02
    Relative loglikelihood change                      0.100D-05
    Derivative                                         0.100D-02
Optimization Specifications for the M step of the EM Algorithm for
Categorical Latent variables
  Number of M step iterations                                    1
  M step convergence criterion                         0.100D-02
  Basis for M step termination                         ITERATION
Optimization Specifications for the M step of the EM Algorithm for
Censored, Binary or Ordered Categorical (Ordinal), Unordered
Categorical (Nominal) and Count Outcomes
  Number of M step iterations                                    1
  M step convergence criterion                         0.100D-02
  Basis for M step termination                         ITERATION
  Maximum value for logit thresholds                          15
  Minimum value for logit thresholds                         -15
  Minimum expected cell size for chi-square            0.100D-01
Maximum number of iterations for H1                         2000
Convergence criterion for H1                           0.100D-03
Optimization algorithm                                       EMA
Integration Specifications
  Type                                                  STANDARD
  Number of integration points                                15
  Dimensions of numerical integration                          0
  Adaptive quadrature                                         ON
Link                                                       LOGIT
Cholesky                                                     OFF

Input data file(s)
  F:\applied_analysis_book\Mplus\hrs.txt
Input data format  FREE


SUMMARY OF DATA
```

```
     Number of missing data patterns             1
     Number of strata                           56
     Number of clusters                        112

COVARIANCE COVERAGE OF DATA
Minimum covariance coverage value   0.100

SUMMARY OF CATEGORICAL DATA PROPORTIONS

    SELFRHEA
      Category 1    0.127
      Category 2    0.306
      Category 3    0.299
      Category 4    0.190
      Category 5    0.077

THE MODEL ESTIMATION TERMINATED NORMALLY

TESTS OF MODEL FIT

Loglikelihood

          H0 Value                       -27393.201
          H0 Scaling Correction Factor       2.248
            for MLR

Information Criteria

          Number of Free Parameters              6
          Akaike (AIC)                   54798.402
          Bayesian (BIC)                 54845.337
          Sample-Size Adjusted BIC       54826.269
            (n* = (n + 2) / 24)

MODEL RESULTS
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
 SELFRHEALT ON
    KAGE               0.029      0.002     13.228      0.000
    MALE              -0.071      0.032     -2.186      0.029
 Thresholds
    SELFRHEA$1        -0.071      0.153     -0.463      0.643
    SELFRHEA$2         1.614      0.153     10.560      0.000
    SELFRHEA$3         2.917      0.159     18.367      0.000
    SELFRHEA$4         4.405      0.165     26.650      0.000


LOGISTIC REGRESSION ODDS RATIO RESULTS
 SELFRHEALT ON
    KAGE               1.029
    MALE               0.932
QUALITY OF NUMERICAL RESULTS

     Condition Number for the Information Matrix              0.121E-05
       (ratio of smallest to largest eigenvalue)
```

ANALYSIS EXAMPLE 9.4 POISSON HRS DATA

Mplus VERSION 5.21
MUTHEN & MUTHEN
04/08/2010   1:52 PM

INPUT INSTRUCTIONS

```
  TITLE:  ANALYSIS EXAMPLE 9.4 HRS DATA POISSON REGRESSION
  DATA:
    FILE IS "F:\applied_analysis_book\Mplus\hrspoi.txt";

  VARIABLE:
    NAMES ARE kwgtr stratum numsecu numfalls24 male age75_84 age85 arthritis diabetes
      bodywgt totheight offset24 age65p ;

   USEVARIABLES ARE numfalls24 male age75_84 age85 arthritis diabetes
      bodywgt totheight ;
   Missing are . ;
   WEIGHT IS kwgtr ;
   stratification is stratum ;
   cluster is numsecu ;
   count = numfalls24  ;
   subpopulation =(age65p==1) ;

  ANALYSIS:
    type is complex;
    estimator is mlr ;

  Model:
    numfalls24 on male age75_84 age85 arthritis diabetes bodywgt totheight ;
```

ANALYSIS EXAMPLE 9.4 HRS DATA POISSON REGRESSION

SUMMARY OF ANALYSIS

```
Number of groups                                                 1
Number of observations                                       10891

Number of dependent variables                                    1
Number of independent variables                                  7
Number of continuous latent variables                            0

Observed dependent variables

  Count
   NUMFALLS

Observed independent variables
   MALE        AGE75_84    AGE85       ARTHRITI    DIABETES    BODYWGT
   TOTHEIGH

Variables with special functions

  Stratification        STRATUM
  Cluster variable      NUMSECU
  Weight variable       KWGTR

Estimator                                                      MLR
Information matrix                                        OBSERVED
Optimization Specifications for the Quasi-Newton Algorithm for
Continuous Outcomes
  Maximum number of iterations                                 100
  Convergence criterion                                0.100D-05
Optimization Specifications for the EM Algorithm
  Maximum number of iterations                                 500
  Convergence criteria
    Loglikelihood change                               0.100D-02
    Relative loglikelihood change                      0.100D-05
    Derivative                                         0.100D-02
Optimization Specifications for the M step of the EM Algorithm for
Categorical Latent variables
  Number of M step iterations                                    1
  M step convergence criterion                         0.100D-02
  Basis for M step termination                         ITERATION
Optimization Specifications for the M step of the EM Algorithm for
Censored, Binary or Ordered Categorical (Ordinal), Unordered
Categorical (Nominal) and Count Outcomes
  Number of M step iterations                                    1
  M step convergence criterion                         0.100D-02
  Basis for M step termination                         ITERATION
  Maximum value for logit thresholds                          15
  Minimum value for logit thresholds                         -15
  Minimum expected cell size for chi-square            0.100D-01
Maximum number of iterations for H1                         2000
Convergence criterion for H1                           0.100D-03
Optimization algorithm                                       EMA
Integration Specifications
  Type                                                  STANDARD
  Number of integration points                                15
  Dimensions of numerical integration                          0
  Adaptive quadrature                                         ON
Cholesky                                                      OFF

Input data file(s)
  F:\applied_analysis_book\Mplus\hrspoi.txt
Input data format  FREE
```

THE MODEL ESTIMATION TERMINATED NORMALLY

TESTS OF MODEL FIT

Loglikelihood

H0 Value	-18903.479
H0 Scaling Correction Factor for MLR	10.313

Information Criteria

Number of Free Parameters	8
Akaike (AIC)	37822.957
Bayesian (BIC)	37881.323
Sample-Size Adjusted BIC (n* = (n + 2) / 24)	37855.900

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
NUMFALLS24 ON				
MALE	0.183	0.107	1.706	0.088
AGE75_84	0.238	0.053	4.459	0.000
AGE85	0.584	0.090	6.490	0.000
ARTHRITIS	0.487	0.082	5.905	0.000
DIABETES	0.260	0.069	3.766	0.000
BODYWGT	0.001	0.001	1.044	0.297
TOTHEIGHT	-0.022	0.011	-2.034	0.042
Intercepts				
NUMFALLS24	0.494	0.636	0.777	0.437

QUALITY OF NUMERICAL RESULTS

Condition Number for the Information Matrix (ratio of smallest to largest eigenvalue)	0.549E-07

Beginning Time:	13:52:00
Ending Time:	13:52:02
Elapsed Time:	00:00:02

NOTE: NO NEGATIVE BINOMIAL IN MPLUS (NO ZERO-INFLATED EITHER)