

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.

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 SURVIVAL ANALYSES USING WesVar 4.3

WesVar uses repeated replication variance estimation methods exclusively and as a result does not offer the Taylor Series Linearization approach.

WesVar is a point and click tool with log and output files that echo the options and variables selected for the particular analysis. As a result the output presented for WesVar examples consists of the log file and the output file. The exact syntax is not presented since it is not generated by the program nor is it possible to run WesVar with just user-written syntax but "Workbook" files can be created for a record of the analysis session. The workbook files will be posted on the ASDA web site in the near future and would enhance this output. From the output provided, you can determine the data used, output options, variables analyzed and other details of the analysis.

WesVar Regression menus can perform only some of the analysis examples in Chapter 10: the discrete time logistic regression is available (given the correct data structure of a person year file) while the Cox Proportional Hazards model, Kaplan-Meier survival curves and other graphics are not included in this software package.

Some of the fine points of this tool are the use of the subpopulation filter in the regression request statement, specification of the "success" or outcome value for the dependent variable, various output options to specify the statistics of interest and a number of Repeated Replication variance estimation methods (JK1, JK2, BRR, etc.). For these examples, the JK2 method was used throughout but other methods are available. For ease of use in the model, use of "reverse coded" class variables is used for matching the omitted category default of the lowest category in Stata. See the WesVar User's Guide for details.

ANALYSIS EXAMPLE: DISCRETE TIME LOGIT REGRESSION NCS-R DATA (TABLE 10.5)

WESVAR VERSION NUMBER : 4.3
 TIME THE JOB EXECUTED : 14:15:45 03/28/2010
 INPUT DATASET NAME : C:\Program Files\Westat\WesVar\Data\ncsrpersonyear_JK2_P1WT_INT LE AGEONSETMDE.var
 TIME THE INPUT DATASET CREATED : 14:02:19 03/28/2010
 FULL SAMPLE WEIGHT : NCSRWTSH
 REPLICATE WEIGHTS : RPL01...RPL42
 VARIANCE ESTIMATION METHOD : JK2

TYPE OF ANALYSIS : LOGISTIC
 CONVERGENCE CRITERION : 1e-06
 MAXIMUM NUMBER OF ITERATIONS : 25
 VALUE OF ALPHA (CONFIDENCE LEVEL %) : 0.05000 (95.00000 %)
 OPTION OUTPUT REPLICATE COEFFICIENTS : OFF
 OPTION OUTPUT ITERATION HISTORY : OFF

MODEL(S): mdtv = int AGE sexm ED_REV[4] race_rev[4] MAR_REV[3]

NUMBER OF REPLICATES : 42
 NUMBER OF OBSERVATIONS READ : 385696
 WEIGHTED NUMBER OF OBSERVATIONS READ : 386866.047

MODEL : mdtv = int AGE sexm ED_REV[4] race_rev[4] MAR_REV[3]

Class Variable Index :

ED_REV.1 : 1
 ED_REV.2 : 2
 ED_REV.3 : 3
 ED_REV.4 : 4
 RACE_REV.1 : 1
 RACE_REV.2 : 2
 RACE_REV.3 : 3
 RACE_REV.4 : 4
 MAR_REV.1 : 1
 MAR_REV.2 : 2
 MAR_REV.3 : 3

OPTIONS : Intercept,
 No Standardized Coefficient,
 Degrees of Freedom = 42
 t VALUE : 2.018

STARTING VALUES : INTERCEPT : 0.0000

int : 0.0000
 AGE : 0.0000
 sexm : 0.0000
 ED_REV.1 : 0.0000
 ED_REV.2 : 0.0000
 ED_REV.3 : 0.0000
 race_rev.1 : 0.0000
 race_rev.2 : 0.0000
 race_rev.3 : 0.0000
 MAR_REV.1 : 0.0000
 MAR_REV.2 : 0.0000

BY : None Specified.

MISSING : 0 (UNWEIGHTED)
 0.000000 (WEIGHTED)

NONMISSING : 385696 (UNWEIGHTED)
 386866.046859 (WEIGHTED)

Success = records with dependent value equal to 1 : 1829 (UNWEIGHTED)
 1779.463743 (WEIGHTED)

Failure = records with dependent value equal to 0 : 383867 (UNWEIGHTED)
 385086.583116 (WEIGHTED)

ITERATIONS REQUIRED FOR FULL SAMPLE : 10

MAXIMUM ITERATIONS FOR REPLICATE SAMPLE : 10

-2 LOG LIKELIHOOD FOR FULL SAMPLE : 21657.91151

-2 LOG LIKELIHOOD FOR MODEL CONTAINING INTERCEPT ONLY : 22704.04647

Negative log-likelihood: 0.046
 Likelihood ratio (Cox-Snell): 0.003 Maximum possible value: 0.057
 Likelihood ratio (Estrella): 0.003

PARAMETER	PARAMETER	STANDARD ERROR	TEST FOR H0:		LOWER 95%	UPPER 95%
	ESTIMATE	OF ESTIMATE	PARAMETER=0	PROB> T		
INTERCEPT	-3.44	0.161	-21.314	0.000	-3.761	-3.110
int	0.03	0.002	15.880	0.000	0.029	0.037
AGE	-0.06	0.002	-23.836	0.000	-0.063	-0.053
sexm	-0.44	0.061	-7.244	0.000	-0.569	-0.321
ED_REV.1	-0.02	0.063	-0.307	0.760	-0.147	0.108
ED_REV.2	0.09	0.057	1.631	0.110	-0.022	0.208
ED_REV.3	-0.02	0.065	-0.308	0.759	-0.152	0.112
race_rev.1	0.07	0.119	0.623	0.536	-0.166	0.314
race_rev.2	-0.46	0.149	-3.059	0.004	-0.758	-0.156
race_rev.3	-0.25	0.135	-1.844	0.072	-0.520	0.023
MAR_REV.1	-0.04	0.088	-0.401	0.691	-0.213	0.143
MAR_REV.2	0.49	0.060	8.187	0.000	0.372	0.616

TEST	F VALUE	NUM. DF	DENOM. DF	PROB>F
OVERALL FIT	52.625	11	32	0.000
int	252.161	1	42	0.000
AGE	568.149	1	42	0.000
sexm	52.469	1	42	0.000
ED_REV[4]	1.872	3	40	0.150
race_rev[4]	12.093	3	40	0.000
MAR_REV[3]	34.915	2	41	0.000

PARAMETER	ESTIMATE	LOWER 95%	UPPER 95%
int	1.03	1.029	1.038
AGE	0.94	0.939	0.948
sexm	0.64	0.566	0.725
ED_REV.1	0.98	0.863	1.114
ED_REV.2	1.10	0.978	1.231
ED_REV.3	0.98	0.859	1.118
race_rev.1	1.08	0.847	1.368
race_rev.2	0.63	0.468	0.856
race_rev.3	0.78	0.594	1.024
MAR_REV.1	0.97	0.808	1.153
MAR_REV.2	1.64	1.451	1.852

NOTE: CODES FOR EDUCATION 1=0-11 2=12 3=13-15 4=16+ YEARS OF EDUCATION, CODES FOR MARITAL STATUS 1=MARRIED 2=PREVIOUSLY MARRIED 3=NEVER MARRIED, RACE 1=OTHER 2=HISPANIC 3=BLACK 4=WHITE. REVERSE CODING USED IN THE MODEL IS SIMPLY THE REVERSE OF THE CODES ABOVE.