

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 LINEAR REGRESSION ANALYSIS 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 Tables and Regression menus can perform nearly all of the examples in Chapter 7 through use of the Regression menus including weighted and complex sample corrected linear regression analyses. Some of the examples are omitted here due to the program's inability to perform un-weighted analyses as well as graphics and diagnostic plots.

Some of the fine points of this tool are the use of the subpopulation filter in the regression request statement, creation of variables used in the analyses (means, ratios, differences, etc.), 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. In order to match the omitted reference category to Stata (lowest category omitted) use of reversed classification variables is used in this chapter's examples. In the modeling commands of WesVar, variable labels are not applied in the output therefore codes for the variables used in the models are included in this output. See the WesVar User's Guide for details.

TABLES OF CLASS VARIABLES WITH LABELS FOR IDENTIFICATION IN THE MODELS

The WesVar output in this chapter uses the name of the variable plus the numeric code in the output (racecat_rev.1 for the first category of the reversed race variable) and so the following variable labels provide a way to identify the number with the variable labels.

All examples in this chapter use NHANES Adult data (Age 18+)

GENDER_R	STATISTIC	EST_TYPE	ESTIMATE	STDERROR	LOWER 95%	UPPER 95%	DEFF
1Female	SUM_WTS	PERCENT	51.77	0.513	50.68	52.87	0.562
2Male	SUM_WTS	PERCENT	48.23	0.513	47.13	49.32	0.562
MARGINAL	SUM_WTS	PERCENT	100.00

MARCAT_R	STATISTIC	EST_TYPE	ESTIMATE	STDERROR	LOWER 95%	UPPER 95%	DEFF
1Never Married	SUM_WTS PERCENT	PERCENT	17.83	1.214	15.24	20.41	5.359
2Previously Married	SUM_WTS PERCENT	PERCENT	18.25	0.744	16.66	19.84	1.976
3Married	SUM_WTS PERCENT	PERCENT	63.92	1.516	60.69	67.15	5.308
MARGINAL	SUM_WTS PERCENT	PERCENT	100.00

RACE_REV	STATISTIC	EST_TYPE	ESTIMATE	STDERROR	LOWER 95%	UPPER 95%	DEFF
1Other	SUM_WTS	PERCENT	5.40	0.588	4.15	6.66	3.612
2Black	SUM_WTS	PERCENT	11.73	1.986	7.49	15.96	20.320
3White	SUM_WTS	PERCENT	71.41	2.772	65.51	77.32	20.083
4Other Hispanic	SUM_WTS	PERCENT	3.38	0.745	1.79	4.97	9.059
5Mexican	SUM_WTS	PERCENT	8.08	1.006	5.93	10.22	7.265
MARGINAL	SUM_WTS	PERCENT	100.00

ANALYSIS EXAMPLE 7.5: BIVARIATE TESTS OF PREDICTORS FOR MODEL FITTING

MODEL : BPXDI1_1 = RACE_REV[5]

Class Variable Index :

RACE_REV.1 : 1
 RACE_REV.2 : 2
 RACE_REV.3 : 3
 RACE_REV.4 : 4
 RACE_REV.5 : 5

PARAMETER	PARAMETER ESTIMATE	STANDARD ERROR OF ESTIMATE	TEST FOR H0: PARAMETER=0	PROB> T	LOWER 95%	UPPER 95%
INTERCEPT	68.30	0.408	167.468	0.000	67.430	69.169
RACE_REV.1	1.78	1.036	1.723	0.105	-0.423	3.992
RACE_REV.2	3.73	0.741	5.034	0.000	2.149	5.306
RACE_REV.3	2.43	0.551	4.404	0.001	1.253	3.602
RACE_REV.4	1.59	1.079	1.476	0.161	-0.708	3.893

TEST	F VALUE	NUM. DF	DENOM. DF	PROB>F
OVERALL FIT	6.441	4	12	0.005
RACE_REV[5]	6.441	4	12	0.005

PARAMETER	PARAMETER ESTIMATE	STANDARD ERROR OF ESTIMATE	TEST FOR H0: PARAMETER=0	PROB> T	LOWER 95%	UPPER 95%
INTERCEPT	71.39	0.474	150.515	0.000	70.381	72.403
MARCAT_R.1	-4.39	0.578	-7.583	0.000	-5.619	-3.153
MARCAT_R.2	-0.07	0.682	-0.107	0.916	-1.527	1.381

TEST	F VALUE	NUM. DF	DENOM. DF	PROB>F	NOTE
OVERALL FIT	36.940	2	14	0.000	
MARCAT_R[3]	36.940	2	14	0.000	

PARAMETER	PARAMETER ESTIMATE	STANDARD ERROR OF ESTIMATE	TEST FOR H0: PARAMETER=0	PROB> T	LOWER 95%	UPPER 95%
INTERCEPT	72.07	0.423	170.529	0.000	71.169	72.970
GENDER_R.1	-2.84	0.378	-7.527	0.000	-3.650	-2.039

TEST	F VALUE	NUM. DF	DENOM. DF	PROB>F
OVERALL FIT	56.655	1	15	0.000
GENDER_R[2]	56.655	1	15	0.000

PARAMETER	PARAMETER ESTIMATE	STANDARD ERROR OF ESTIMATE	TEST FOR H0: PARAMETER=0	PROB> T	LOWER 95%	UPPER 95%
INTERCEPT	70.62	0.355	198.986	0.000	69.859	71.372
AGECENT	0.06	0.021	2.694	0.017	0.012	0.103

TEST	F VALUE	NUM. DF	DENOM. DF	PROB>F
OVERALL FIT	7.256	1	15	0.017
AGECENT	7.256	1	15	0.017

ANALYSIS EXAMPLE: UNWEIGHTED AND NON - COMPLEX SAMPLE SURVEY CORRECTED REGRESSION
NOT AVAILABLE IN WesVar

ANALYSIS EXAMPLE: WEIGHTED WITHOUT COMPLEX SAMPLE ADJUSTMENT
NOT AVAILABLE IN WesVar

ANALYSIS EXAMPLE 7.5: WEIGHTED AND DESIGN CORRECTED WITH ALL PREDICTORS IN MODEL

OPTIONS : Intercept,
 No Standardized Coefficient,
 Degrees of Freedom = 15
 t VALUE : 2.131

BY : AGE18P = 0
 MISSING : 3837 (UNWEIGHTED)
 58380235.173233 (WEIGHTED)

NONMISSING : 948 (UNWEIGHTED)
 15536185.383222 (WEIGHTED)

BY : AGE18P = 1
 MISSING : 985 (UNWEIGHTED)
 27852349.583949 (WEIGHTED)

NONMISSING : 4578 (UNWEIGHTED)
 189848121.699553 (WEIGHTED)

PARAMETER	PARAMETER ESTIMATE	STANDARD ERROR OF ESTIMATE	TEST FOR H0: PARAMETER=0	PROB> T	LOWER 95%	UPPER 95%
INTERCEPT	70.68	0.502	140.802	0.000	69.608	71.748
RACE_REV.1	1.96	0.993	1.972	0.067	-0.159	4.076
RACE_REV.2	4.41	0.754	5.850	0.000	2.802	6.015
RACE_REV.3	2.19	0.604	3.629	0.002	0.904	3.479
RACE_REV.4	1.79	1.117	1.599	0.131	-0.595	4.168
MARCAT_R.1	-4.36	0.576	-7.566	0.000	-5.583	-3.129
MARCAT_R.2	0.02	0.726	0.024	0.981	-1.530	1.565
GENDER_R.1	-3.00	0.328	-9.141	0.000	-3.696	-2.298
AGECENT	0.02	0.023	0.755	0.462	-0.031	0.065

TEST	F VALUE	NUM. DF	DENOM. DF	PROB>F	NOTE
OVERALL FIT	12.635	8	8	0.001	
RACE_REV[5]	7.769	4	12	0.002	
MARCAT_R[3]	29.390	2	14	0.000	
GENDER_R[2]	83.556	1	15	0.000	
AGECENT	0.570	1	15	0.462	

ANALYSIS EXAMPLE 7.5: ADD AGE SQUARED TO MODEL FOR FIT EVALUATION

OPTIONS : Intercept,
 No Standardized Coefficient,
 Degrees of Freedom = 15
 t VALUE : 2.131
 BY : AGE18P = 0
 MISSING : 3837 (UNWEIGHTED)
 58380235.173233 (WEIGHTED)
 NONMISSING : 948 (UNWEIGHTED)
 15536185.383222 (WEIGHTED)

 BY : AGE18P = 1
 MISSING : 985 (UNWEIGHTED)
 27852349.583949 (WEIGHTED)
 NONMISSING : 4578 (UNWEIGHTED)
 189848121.699553 (WEIGHTED)

PARAMETER	PARAMETER ESTIMATE	STANDARD ERROR OF ESTIMATE	TEST FOR H0: PARAMETER=0	PROB> T	LOWER 95%	UPPER 95%
INTERCEPT	73.86	0.454	162.766	0.000	72.892	74.826
RACE_REV.1	1.19	0.940	1.264	0.226	-0.816	3.193
RACE_REV.2	3.47	0.774	4.475	0.000	1.815	5.116
RACE_REV.3	1.78	0.631	2.823	0.013	0.436	3.125
RACE_REV.4	1.19	1.072	1.110	0.285	-1.095	3.474
MARCAT_R.1	-0.34	0.599	-0.573	0.575	-1.620	0.933
MARCAT_R.2	1.04	0.629	1.654	0.119	-0.301	2.382
GENDER_R.1	-2.72	0.336	-8.099	0.000	-3.437	-2.005
AGECENT	0.13	0.015	8.358	0.000	0.093	0.157
AGECENT**2	-0.01	0.001	-15.835	0.000	-0.014	-0.011

TEST	F VALUE	NUM. DF	DENOM. DF	PROB>F	NOTE
OVERALL FIT	89.573	9	7	0.000	
RACE_REV[5]	5.169	4	12	0.012	
MARCAT_R[3]	1.483	2	14	0.260	
GENDER_R[2]	65.590	1	15	0.000	
AGECENT	69.848	1	15	0.000	
AGECENT**2	250.741	1	15	0.000	

ANALYSIS EXAMPLE 7.5: ADD INTERACTION OF AGE*RACE/ETHNICITY

NOTE THAT "TEST1" IS THE TEST OF THE INTERACTION OF AGE*RACE/ETHNICITY

OPTIONS : Intercept,
 No Standardized Coefficient,
 Degrees of Freedom = 15
 t VALUE : 2.131
 TEST(S) : TEST1 : agecent*race_rev.1=0, agecent*race_rev.2=0, agecent*race_rev.3=0, agecent*race_rev.4=0,
 agesq*race_rev.1=0, agesq*race_rev.2=0, agesq*race_rev.3=0, agesq*race_rev.4=0
 BY : AGE18P = 0
 MISSING : 3837 (UNWEIGHTED)
 58380235.173233 (WEIGHTED)
 NONMISSING : 948 (UNWEIGHTED)
 15536185.383222 (WEIGHTED)
 BY : AGE18P = 1
 MISSING : 985 (UNWEIGHTED)
 27852349.583949 (WEIGHTED)
 NONMISSING : 4578 (UNWEIGHTED)
 189848121.699553 (WEIGHTED)

PARAMETER	PARAMETER ESTIMATE	STANDARD ERROR OF ESTIMATE	TEST FOR H0: PARAMETER=0	PROB> T	LOWER 95%	UPPER 95%
INTERCEPT	74.22	0.474	156.419	0.000	73.209	75.231
RACE_REV.1	0.71	1.194	0.592	0.563	-1.839	3.252
RACE_REV.2	3.02	0.911	3.317	0.005	1.080	4.964
RACE_REV.3	1.42	0.573	2.486	0.025	0.203	2.645
RACE_REV.4	0.61	1.275	0.477	0.640	-2.108	3.325
MARCAT_R.1	-0.34	0.604	-0.556	0.587	-1.623	0.951
MARCAT_R.2	0.99	0.635	1.558	0.140	-0.364	2.345
GENDER_R.1	-2.72	0.342	-7.953	0.000	-3.450	-1.992
AGECENT	0.13	0.031	4.282	0.001	0.067	0.200
AGESQ	-0.01	0.001	-12.187	0.000	-0.016	-0.011
agecent*race_rev.1	-0.09	0.059	-1.536	0.145	-0.217	0.035
agecent*race_rev.2	0.04	0.037	1.105	0.287	-0.038	0.120
agecent*race_rev.3	-0.01	0.040	-0.331	0.745	-0.099	0.072
agecent*race_rev.4	0.07	0.082	0.819	0.426	-0.108	0.243
agesq*race_rev.1	0.00	0.003	0.065	0.949	-0.006	0.007
agesq*race_rev.2	0.00	0.002	1.206	0.247	-0.002	0.005
agesq*race_rev.3	0.00	0.001	0.974	0.345	-0.001	0.004
agesq*race_rev.4	0.00	0.004	1.032	0.319	-0.004	0.012

TEST	F VALUE	NUM. DF	DENOM. DF	PROB>F
OVERALL FIT	* not testable	--	insufficient degrees of freedom.	
RACE_REV[5]	3.535	4	12	0.040
MARCAT_R[3]	1.327	2	14	0.297
GENDER_R[2]	63.243	1	15	0.000
AGECENT	25.890	1	15	0.000
AGESQ	107.376	1	15	0.000
agecent*race_rev[5]	1.802	4	12	0.193
agesq*race_rev[5]	0.469	4	12	0.758
TEST1	0.945	8	8	0.531

ANALYSIS EXAMPLE 7.5: ADD INTERACTION OF AGE AND GENDER

NOTE THAT "TEST1" IS THE AGE*GENDER INTERACTION

```

OPTIONS :      Intercept,
             No Standardized Coefficient,
             Degrees of Freedom = 15
             t VALUE : 2.131
TEST(S) :      TEST1 : agecent*GENDER_R.1=0, AGESQ*GENDER_R.1=0
BY :      AGE18P = 0
MISSING :      3837          (UNWEIGHTED)
             58380235.173233 (WEIGHTED)
NONMISSING :    948          (UNWEIGHTED)
             15536185.383222 (WEIGHTED)

BY :      AGE18P = 1
MISSING :      985          (UNWEIGHTED)
             27852349.583949 (WEIGHTED)
NONMISSING :    4578        (UNWEIGHTED)
             189848121.699553 (WEIGHTED)
    
```

PARAMETER	PARAMETER ESTIMATE	STANDARD ERROR OF ESTIMATE	TEST FOR H0: PARAMETER=0	PROB> T	LOWER 95%	UPPER 95%
INTERCEPT	74.14	0.571	129.861	0.000	72.921	75.355
RACE_REV.1	1.21	0.941	1.284	0.219	-0.797	3.213
RACE_REV.2	3.49	0.772	4.522	0.000	1.846	5.138
RACE_REV.3	1.80	0.632	2.843	0.012	0.449	3.143
RACE_REV.4	1.20	1.085	1.107	0.286	-1.112	3.514
MARCAT_R.1	-0.35	0.606	-0.572	0.576	-1.637	0.945
MARCAT_R.2	0.91	0.658	1.378	0.188	-0.496	2.310
GENDER_R.1	-3.24	0.713	-4.542	0.000	-4.756	-1.718
AGECENT	0.12	0.020	5.999	0.000	0.076	0.160
AGESQ	-0.01	0.001	-10.177	0.000	-0.016	-0.011
agecent*GENDER_R.1	0.01	0.028	0.505	0.621	-0.045	0.073
AGESQ*GENDER_R.1	0.00	0.002	1.067	0.303	-0.002	0.005

TEST	F VALUE	NUM. DF	DENOM. DF	PROB>F	NOTE
OVERALL FIT	94.222	11	5	0.000	
RACE_REV[5]	5.224	4	12	0.011	
MARCAT_R[3]	1.099	2	14	0.360	
GENDER_R[2]	20.633	1	15	0.000	
AGECENT	70.482	1	15	0.000	
AGESQ	251.119	1	15	0.000	
agecent*GENDER_R[2]	0.255	1	15	0.621	
AGESQ*GENDER_R[2]	1.138	1	15	0.303	
TEST1	1.716	2	14	0.215	

ANALYSIS EXAMPLE 7.5: RE-FIT FINAL MODEL WITHOUT INTERACTIONS AS THEY WERE NON-SIGNIFICANT

OPTIONS : Intercept,
 No Standardized Coefficient,
 Degrees of Freedom = 15
 t VALUE : 2.131
 BY : AGE18P = 0
 MISSING : 3837 (UNWEIGHTED)
 58380235.173233 (WEIGHTED)
 NONMISSING : 948 (UNWEIGHTED)
 15536185.383222 (WEIGHTED)
 BY : AGE18P = 1
 MISSING : 985 (UNWEIGHTED)
 27852349.583949 (WEIGHTED)
 NONMISSING : 4578 (UNWEIGHTED)
 189848121.699553 (WEIGHTED)

PARAMETER	PARAMETER ESTIMATE	STANDARD ERROR OF ESTIMATE	TEST FOR H0: PARAMETER=0	PROB> T	LOWER 95%	UPPER 95%
INTERCEPT	73.86	0.454	162.766	0.000	72.892	74.826
RACE_REV.1	1.19	0.940	1.264	0.226	-0.816	3.193
RACE_REV.2	3.47	0.774	4.475	0.000	1.815	5.116
RACE_REV.3	1.78	0.631	2.823	0.013	0.436	3.125
RACE_REV.4	1.19	1.072	1.110	0.285	-1.095	3.474
MARCAT_R.1	-0.34	0.599	-0.573	0.575	-1.620	0.933
MARCAT_R.2	1.04	0.629	1.654	0.119	-0.301	2.382
GENDER_R.1	-2.72	0.336	-8.099	0.000	-3.437	-2.005
AGECENT	0.13	0.015	8.358	0.000	0.093	0.157
AGECENT**2	-0.01	0.001	-15.835	0.000	-0.014	-0.011

TEST	F VALUE	NUM. DF	DENOM. DF	PROB>F	NOTE
OVERALL FIT	89.573	9	7	0.000	
RACE_REV[5]	5.169	4	12	0.012	
MARCAT_R[3]	1.483	2	14	0.260	
GENDER_R[2]	65.590	1	15	0.000	
AGECENT	69.848	1	15	0.000	
AGECENT**2	250.741	1	15	0.000	