

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 GENERALIZED LINEAR MODELS IN SPSS/PASW V18 COMPLEX SAMPLES MODULE

SPSS/PASW ORDINAL commands can perform some of the analyses presented in Chapter 9 of ASDA. CSORDINAL performs multinomial logit and cumulative logit regression but many of the other analyses such as Poisson, negative binomial and the zero-inflated versions of Poisson and negative binomial regression are not available in the SPSS/PASW Complex Samples module. Note that SPSS CSORDINAL includes a test of the parallel lines assumption and this is demonstrated in the ordinal logistic regression example in this chapter.

Some of the fine points of these procedures are the use of a SUBPOP statement for subpopulation analyses, various output statistics specified on the STATISTICS subcommand, and use of an analysis Plan file for all Complex Samples commands. The plan file should be prepared prior to working with any Complex Samples commands and offers the ability to declare weights and design variables to the program. For matching the reference group to Stata v10.1, we use a reverse coding strategy as this is one way to match the omitted categories of Stata (lowest category is omitted by default). Other approaches might be to use individual indicator variables for each level of the categorical variables. Finally, use of the /CUSTOM command for hypothesis testing is demonstrated in the multinomial logit model. This sub-command is required to define some of the hypothesis tests not already included in the default output.

*MULTINOMIAL LOGIT REGRESSION: ANALYSIS EXAMPLE TABLE 9.2 AND 9.3 NCS-R DATA

*NOTE THAT MODEL IS RUN TWICE: FIRST IS WITH FACTOR VARIABLES FOR EDUCATION, AGE AND MARITAL STATUS AND 2ND MODEL USES EDUCATION AS A SERIES OF INDICATOR VARIABLES FOR USE IN /CUSTOM HYPOTHESIS TESTING.

FIRST RUN OF MODEL:

* Complex Samples Logistic Regression.

* EXAMPLE 9.3 MULTINOMIAL LOGISTIC WITH CUSTOM HYPOTHESIS TESTS: RUN WITH EDUCATION AS FACTOR VARIABLE.

CSLOGISTIC WKSTAT3C(LOW) BY revedcat revag4cat revmar3cat WITH sexm ald mde

```

/PLAN FILE='F:\applied_analysis_book\SPSS Analysis Examples Replication\Analysis Examples Replication Winter 2010 SPSSv18\ncsr_p2wt.csaplan'
/MODEL revedcat revag4cat revmar3cat sexm ald mde
/INTERCEPT INCLUDE=YES SHOW=YES
/STATISTICS PARAMETER EXP SE CINTERVAL TTEST
/TEST TYPE=F PADJUST=LSD
/MISSING CLASSMISSING=EXCLUDE
/CRITERIA MXITER=100 MXSTEP=5 PCONVERGE=[1E-006 RELATIVE] LCONVERGE=[0] CHKSEP=20 CILEVEL=95
/PRINT SUMMARY VARIABLEINFO SAMPLEINFO.

```

Sample Design Information

		N
Unweighted Cases	Valid	5679
	Invalid	3603
	Total	9282
Population Size		5.667E3
Stage 1	Strata	42
	Units	84
Sampling Design Degrees of Freedom		42

Categorical Variable Information

		Weighted Count	Weighted Percent
Work Status 3 categories	1 ^b	3671.472	64.8%
1=Employed 2=Unemployed	2	289.817	5.1%
3=NLF ^a	3	1705.896	30.1%
1=16+ 2=13-15 3=12 4=0-11	1.0000	1315.579	23.2%
	2.0000	1567.870	27.7%
	3.0000	1848.467	32.6%
	4.0000	935.269	16.5%
1=60+ 2=45-59 3=30-44	1.0000	1202.804	21.2%
	2.0000	1502.135	26.5%
	3.0000	1633.099	28.8%
4=18-29	4.0000	1329.147	23.5%
	1.0000	1312.250	23.2%
	2.0000	1177.332	20.8%
3=married	3.0000	3177.603	56.1%
Population Size		5667.185	100.0%

a. Dependent Variable

b. Reference Category

Covariate Information

	Mean
sexm	.47
ald	.05
mde	.19

Pseudo R Squares

Cox and Snell	.253
Nagelkerke	.318
McFadden	.184

Dependent Variable: Work

Status 3 categories

1=Employed 2=Unemployed

3=NLF (reference category

= 1)

Model: (Intercept),

revedcat, revag4cat,

revmar3cat, sexm, ald, mde

Tests of Model Effects

Source	df1	df2	Wald F	Sig.
(Corrected Model)	22.000	21.000	73.913	.000
(Intercept)	2.000	41.000	127.647	.000
revedcat	6.000	37.000	13.681	.000
revag4cat	6.000	37.000	83.591	.000
revmar3cat	4.000	39.000	24.813	.000
sexm	2.000	41.000	35.755	.000
ald	2.000	41.000	5.048	.011
mde	2.000	41.000	1.139	.330

Dependent Variable: Work Status 3 categories 1=Employed

2=Unemployed 3=NLF (reference category = 1)

Model: (Intercept), revedcat, revag4cat, revmar3cat, sexm, ald,

mde

Parameter Estimates

Work Status 3 categories 1=Employed 2=Unemployed 3=NLF	Parameter	B	Std. Error	95% Confidence Interval		Hypothesis Test			Exp(B)	95% Confidence Interval for Exp(B)		
				Lower	Upper	t	df	Sig.		Lower	Upper	
				2	(Intercept)	-.644	.296	-1.241		-.046	-2.174	42.000
	[revedcat=1.0000]	-1.731	.310	-2.358	-1.104	-5.575	42.000	.000	.177	.095	.331	
	[revedcat=2.0000]	-1.365	.258	-1.885	-.846	-5.302	42.000	.000	.255	.152	.429	
	[revedcat=3.0000]	-.847	.235	-1.322	-.372	-3.598	42.000	.001	.429	.267	.689	
	[revedcat=4.0000]	.000 ^a	1.000	.	.	
	[revag4cat=1.0000]	1.828	.295	1.234	2.423	6.204	42.000	.000	6.224	3.434	11.281	
	[revag4cat=2.0000]	-.838	.258	-1.359	-.317	-3.246	42.000	.002	.433	.257	.728	
	[revag4cat=3.0000]	-.852	.295	-1.447	-.258	-2.894	42.000	.006	.426	.235	.773	
	[revag4cat=4.0000]	.000 ^a	1.000	.	.	
	[revmar3cat=1.0000]	-2.785	.380	-3.552	-2.017	-7.323	42.000	.000	.062	.029	.133	
	[revmar3cat=2.0000]	-.590	.225	-1.044	-.135	-2.619	42.000	.012	.554	.352	.873	
	[revmar3cat=3.0000]	.000 ^a	1.000	.	.	
	sexm	-1.393	.198	-1.792	-.994	-7.049	42.000	.000	.248	.167	.370	
	ald	-.164	.357	-.884	.557	-.459	42.000	.649	.849	.413	1.745	
	mde	-.140	.157	-.457	.178	-.889	42.000	.379	.870	.633	1.194	
dimension0	3	(Intercept)	-.379	.173	-.728	-.031	-2.194	42.000	.034	.684	.483	.970
		[revedcat=1.0000]	-1.230	.160	-1.552	-.907	-7.704	42.000	.000	.292	.212	.404
		[revedcat=2.0000]	-.917	.146	-1.213	-.621	-6.259	42.000	.000	.400	.297	.537
		[revedcat=3.0000]	-.651	.141	-.936	-.367	-4.619	42.000	.000	.521	.392	.693
		[revedcat=4.0000]	.000 ^a	1.000	.	.	
		[revag4cat=1.0000]	2.381	.173	2.031	2.731	13.729	42.000	.000	10.811	7.619	15.341
		[revag4cat=2.0000]	.065	.171	-.280	.410	.380	42.000	.706	1.067	.756	1.507
		[revag4cat=3.0000]	-.316	.129	-.576	-.057	-2.457	42.000	.018	.729	.562	.945
		[revag4cat=4.0000]	.000 ^a	1.000	.	.	
		[revmar3cat=1.0000]	.553	.132	.286	.820	4.176	42.000	.000	1.738	1.331	2.270
		[revmar3cat=2.0000]	-.052	.105	-.264	.160	-.498	42.000	.621	.949	.768	1.173
		[revmar3cat=3.0000]	.000 ^a	1.000	.	.	
		sexm	-.640	.110	-.862	-.418	-5.818	42.000	.000	.527	.422	.658
		ald	.333	.130	.070	.596	2.559	42.000	.014	1.395	1.073	1.815
		mde	.099	.088	-.079	.276	1.120	42.000	.269	1.104	.924	1.318

Dependent Variable: Work Status 3 categories 1=Employed 2=Unemployed 3=NLF (reference category = 1)

Model: (Intercept), revedcat, revag4cat, revmar3cat, sexm, ald, mde

a. Set to zero because this parameter is redundant.

McFadden .184

Dependent Variable: Work

Status 3 categories

1=Employed 2=Unemployed

3=NLF (reference category

= 1)

Model: (Intercept), ed12,

ed1315, ed16, revag4cat,

revmar3cat, sexm, ald, mde

Tests of Model Effects

Source	df1	df2	Wald F	Sig.
(Corrected Model)	22.000	21.000	73.913	.000
(Intercept)	2.000	41.000	26.179	.000
ed12	2.000	41.000	11.778	.000
ed1315	2.000	41.000	24.303	.000
ed16	2.000	41.000	35.191	.000
revag4cat	6.000	37.000	83.591	.000
revmar3cat	4.000	39.000	24.813	.000
sexm	2.000	41.000	35.755	.000
ald	2.000	41.000	5.048	.011
mde	2.000	41.000	1.139	.330

Dependent Variable: Work Status 3 categories 1=Employed

2=Unemployed 3=NLF (reference category = 1)

Model: (Intercept), ed12, ed1315, ed16, revag4cat, revmar3cat,

sexm, ald, mde

Parameter Estimates

Work Status 3 categories 1=Employed 2=Unemployed 3=NLF	Parameter	B	Std. Error	95% Confidence Interval		Hypothesis Test			Exp(B)	95% Confidence Interval for Exp(B)		
				Lower	Upper	t	df	Sig.		Lower	Upper	
				2	(Intercept)	-.644	.296	-1.241		-.046	-2.174	42.000
	ed12	-.847	.235	-1.322	-.372	-3.598	42.000	.001	.429	.267	.689	
	ed1315	-1.365	.258	-1.885	-.846	-5.302	42.000	.000	.255	.152	.429	
	ed16	-1.731	.310	-2.358	-1.104	-5.575	42.000	.000	.177	.095	.331	
	[revag4cat=1.0000]	1.828	.295	1.234	2.423	6.204	42.000	.000	6.224	3.434	11.281	
	[revag4cat=2.0000]	-.838	.258	-1.359	-.317	-3.246	42.000	.002	.433	.257	.728	
	[revag4cat=3.0000]	-.852	.295	-1.447	-.258	-2.894	42.000	.006	.426	.235	.773	
	[revag4cat=4.0000]	.000 ^a	1.000	.	.	
	[revmar3cat=1.0000]	-2.785	.380	-3.552	-2.017	-7.323	42.000	.000	.062	.029	.133	
	[revmar3cat=2.0000]	-.590	.225	-1.044	-.135	-2.619	42.000	.012	.554	.352	.873	
	[revmar3cat=3.0000]	.000 ^a	1.000	.	.	
	sexm	-1.393	.198	-1.792	-.994	-7.049	42.000	.000	.248	.167	.370	
	ald	-.164	.357	-.884	.557	-.459	42.000	.649	.849	.413	1.745	
	mde	-.140	.157	-.457	.178	-.889	42.000	.379	.870	.633	1.194	
dimension0	3	(Intercept)	-.379	.173	-.728	-.031	-2.194	42.000	.034	.684	.483	.970
	ed12	-.651	.141	-.936	-.367	-4.619	42.000	.000	.521	.392	.693	
	ed1315	-.917	.146	-1.213	-.621	-6.259	42.000	.000	.400	.297	.537	
	ed16	-1.230	.160	-1.552	-.907	-7.704	42.000	.000	.292	.212	.404	
	[revag4cat=1.0000]	2.381	.173	2.031	2.731	13.729	42.000	.000	10.811	7.619	15.341	
	[revag4cat=2.0000]	.065	.171	-.280	.410	.380	42.000	.706	1.067	.756	1.507	
	[revag4cat=3.0000]	-.316	.129	-.576	-.057	-2.457	42.000	.018	.729	.562	.945	
	[revag4cat=4.0000]	.000 ^a	1.000	.	.	
	[revmar3cat=1.0000]	.553	.132	.286	.820	4.176	42.000	.000	1.738	1.331	2.270	
	[revmar3cat=2.0000]	-.052	.105	-.264	.160	-.498	42.000	.621	.949	.768	1.173	
	[revmar3cat=3.0000]	.000 ^a	1.000	.	.	
	sexm	-.640	.110	-.862	-.418	-5.818	42.000	.000	.527	.422	.658	
	ald	.333	.130	.070	.596	2.559	42.000	.014	1.395	1.073	1.815	
	mde	.099	.088	-.079	.276	1.120	42.000	.269	1.104	.924	1.318	

Dependent Variable: Work Status 3 categories 1=Employed 2=Unemployed 3=NLF (reference category = 1)

Model: (Intercept), ed12, ed1315, ed16, revag4cat, revmar3cat, sexm, ald, mde

a. Set to zero because this parameter is redundant.

Contrast Coefficients^a

Work Status 3 categories 1=Employed 2=Unemployed 3=NLF		Parameter	L1	L2	L3
dimension0	2	(Intercept)	.000	.000	.000
		ed12	1.000	.000	.000
		ed1315	.000	1.000	.000
		ed16	.000	.000	1.000
		[revag4cat=1.0000]	.000	.000	.000
		[revag4cat=2.0000]	.000	.000	.000
		[revag4cat=3.0000]	.000	.000	.000
		[revag4cat=4.0000]	.000	.000	.000
		[revmar3cat=1.0000]	.000	.000	.000
		[revmar3cat=2.0000]	.000	.000	.000
		[revmar3cat=3.0000]	.000	.000	.000
		sexm	.000	.000	.000
		ald	.000	.000	.000
		mde	.000	.000	.000
	3	(Intercept)	.000	.000	.000
	ed12	-1.000	.000	.000	
	ed1315	.000	-1.000	.000	
	ed16	.000	.000	-1.000	
	[revag4cat=1.0000]	.000	.000	.000	
	[revag4cat=2.0000]	.000	.000	.000	
	[revag4cat=3.0000]	.000	.000	.000	
	[revag4cat=4.0000]	.000	.000	.000	
	[revmar3cat=1.0000]	.000	.000	.000	
	[revmar3cat=2.0000]	.000	.000	.000	
	[revmar3cat=3.0000]	.000	.000	.000	
	sexm	.000	.000	.000	
	ald	.000	.000	.000	
	mde	.000	.000	.000	

a. The default display of this matrix is the transpose of the corresponding L matrix.

Individual Test Results

Contrast	Contrast Estimate	Hypothesized Value	Difference (Estimate - Hypothesized)	Std. Error	df1	df2	Wald F	Sig.
d L1	-.196	.000	-.196	.212	1.000	42.000	.852	.361
i L2	-.448	.000	-.448	.246	1.000	42.000	3.324	.075
m L3	-.501	.000	-.501	.307	1.000	42.000	2.668	.110

Overall Test Results

df1	df2	Wald F	Sig.
3.000	40.000	1.254	.303

*ORDINAL LOGISTIC REGRESSION: ANALYSIS EXAMPLE TABLE 9.5 HRS DATA

Warning # 3211

On at least one case, the value of the weight variable was zero, negative, or missing. Such cases are invisible to statistical procedures and graphs which need positively weighted cases, but remain on the file and are processed by non-statistical facilities such as LIST and SAVE.

* Define Variable Properties.

*selfrhealth.

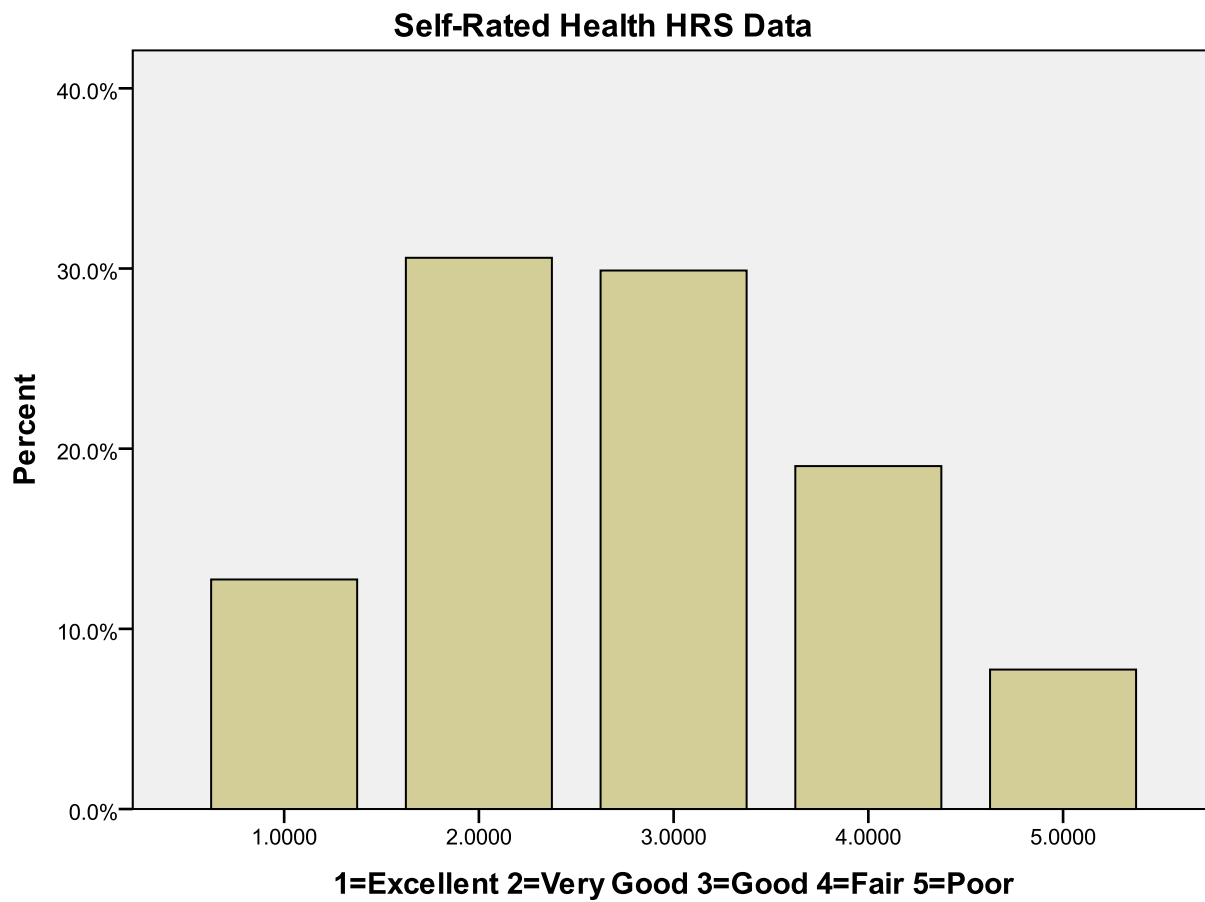
VARIABLE LABELS selfrhealth '1=Excellent 2=Very Good 3=Good 4=Fair 5=Poor'.

EXECUTE.

GRAPH

/BAR(SIMPLE)=PCT BY selfrhealth

/TITLE='Self-Rated Health HRS Data'.



Cases weighted by 2006 weight: respondent level

* Complex Samples Ordinal Regression.

CSORDINAL selfrhealth (ASCENDING) BY GENDER WITH KAGE

```

/PLAN FILE='F:\applied_analysis_book\SPSS Analysis Examples Replication\Analysis Examples Replication Winter 2010 SPSSv18\hrs.csaplan'
/LINK FUNCTION=LOGIT
/MODEL GENDER KAGE
/STATISTICS PARAMETER EXP SE CINTERVAL TTEST
/NONPARALLEL TEST
/TEST TYPE=F PADJUST=LSD
/MISSING CLASSMISSING=EXCLUDE
/CRITERIA MXITER=100 MXSTEP=5 PCONVERGE=[1e-006 RELATIVE] LCONVERGE=[0] METHOD=NEWTON CHKSEP=20 CILEVEL=95
/PRINT SUMMARY SAMPLEINFO.

```

Sample Design Information

		N
Unweighted Cases	Valid	18442
	Invalid	25
	Total	18467
Population Size		7.644E7
Stage 1	Strata	56
	Units	112
Sampling Design Degrees of Freedom		56

Pseudo R Squares

Cox and Snell	.028
Nagelkerke	.030
McFadden	.010

Dependent Variable:

1=Excellent 2=Very Good

3=Good 4=Fair 5=Poor

(Ascending)

Model: (Threshold),

GENDER, KAGE

Link function: Logit

Tests of Model Effects

Source	df1	df2	Wald F	Sig.
GENDER	1.000	56.000	4.780	.033
KAGE	1.000	56.000	174.992	.000

Dependent Variable: 1=Excellent 2=Very Good 3=Good 4=Fair

5=Poor (Ascending)

Model: (Threshold), GENDER, KAGE

Link function: Logit

Parameter Estimates

Parameter		B	Std. Error	95% Confidence Interval		Hypothesis Test			Exp(B)	95% Confidence Interval for Exp(B)	
				Lower	Upper	t	df	Sig.		Lower	Upper
Threshold	[selfrhealth=1.000 0]	-.071	.153	-.377	.236	-.463	56.000	.645	.932	.686	1.266
	[selfrhealth=2.000 0]	1.614	.153	1.308	1.920	10.560	56.000	.000	5.024	3.699	6.824
	[selfrhealth=3.000 0]	2.917	.159	2.599	3.235	18.367	56.000	.000	18.481	13.445	25.403
	[selfrhealth=4.000 0]	4.405	.165	4.074	4.736	26.650	56.000	.000	81.882	58.800	114.024
Regression	[GENDER=1]	-.071	.032	-.135	-.006	-2.186	56.000	.033	.932	.873	.994
	[GENDER=2]	.000 ^a	1.000	.	.
	KAGE	.029	.002	.024	.033	13.228	56.000	.000	1.029	1.025	1.034

Dependent Variable: 1=Excellent 2=Very Good 3=Good 4=Fair 5=Poor (Ascending)

Model: (Threshold), GENDER, KAGE

Link function: Logit

Parameter Estimates

Parameter	B	Std. Error	95% Confidence Interval		Hypothesis Test			Exp(B)	95% Confidence Interval for Exp(B)	
			Lower	Upper	t	df	Sig.		Lower	Upper
Threshold [selfrhealth=1.000 0]	-.071	.153	-.377	.236	-.463	56.000	.645	.932	.686	1.266
[selfrhealth=2.000 0]	1.614	.153	1.308	1.920	10.560	56.000	.000	5.024	3.699	6.824
[selfrhealth=3.000 0]	2.917	.159	2.599	3.235	18.367	56.000	.000	18.481	13.445	25.403
[selfrhealth=4.000 0]	4.405	.165	4.074	4.736	26.650	56.000	.000	81.882	58.800	114.024
Regression [GENDER=1]	-.071	.032	-.135	-.006	-2.186	56.000	.033	.932	.873	.994
[GENDER=2]	.000 ^a	1.000	.	.
KAGE	.029	.002	.024	.033	13.228	56.000	.000	1.029	1.025	1.034

Dependent Variable: 1=Excellent 2=Very Good 3=Good 4=Fair 5=Poor (Ascending)

Model: (Threshold), GENDER, KAGE

Link function: Logit

a. Set to zero because this parameter is redundant.

Generalized Cumulative Model

Test of Parallel Lines

df1	df2	Wald F	Sig.
6.000	51.000	3.939	.003

Dependent Variable: 1=Excellent 2=Very Good

3=Good 4=Fair 5=Poor (Ascending)

Model: (Threshold), GENDER, KAGE

Link function: Logit