

SAS Analysis Examples Replication C9

* SAS Analysis Examples Replication for ASDA 2nd Edition
* Berglund April 2017
* Chapter 9 ;

```
libname ncsr "P:\ASDA 2\Data sets\ncsr\" ;
```

```
data c9_ncsr ;  
  set ncsr.ncsr_sub_13nov2015 ;
```

```
run ;
```

```
proc format ;
```

```
  value af 1='18-29' 2='30-44' 3='45-59' 4='60+' ;  
  value sf 1='M' 2='F' ;  
  value edf 1='0-11' 2='12' 3='13-15' 4='16+' ;  
  value mf 1='Currently Married' 2='Previously Married' 3='Never Married' ;  
  value yn 1='Yes' 0='No' ;
```

```
run ;
```

```
ods rtf style=normalprinter bodytitle ;
```

```
title " Example 9.2.6 Multinomial logistic regression using NCSR data. " ;
```

```
proc surveyl logistic ;
```

```
  strata sestrat ; cluster seclustr ; weight ncsrwtlg ;  
  class wkstat3c (ref=first) sex (ref=last) ed4cat (ref=first) ag4cat (ref=first) mar3cat (ref=first) / param=ref ;  
  model wkstat3c=sex ald mde ed4cat ag4cat mar3cat / link=glogit ;
```

```
run ;
```

* Figure 9.4, this cannot be done directly in SAS PROC SURVEYLOGISTIC, see SAS Technical Support paper here for details on approach:<https://support.sas.com/kb/22/604.html> ;

```
title " GOF test not available in SAS SURVEYLOGISTIC " ;
```

```
libname d 'P:\asda 2\data sets\ess6 russia' ;
```

```
title " 9.3.6 Example: Fitting a Cumulative Logit Regression Model to Complex Sample Survey Data " ;
```

```
data c9_russia ;
```

```
  set d.ess6 russia 2aug2016 ;  
  if stflife =. then stflife2=. ;  
  else if stflife =0 or stflife =1 then stflife2=1 ;  
  else if stflife <=4 then stflife2=2 ;  
  else if stflife <=5 then stflife2=3 ;  
  else if stflife <=8 then stflife2=4 ;  
  else if stflife <=10 then stflife2=5 ;
```

```
run ;
```

```
title "Figure 9.6, Bar Chart of Satisfaction with Life, Weighted by PSPWGHT" ;
```

```
proc freq data=c9_russia ;
```

```
  tables stflife2 / plots=freqplot ;  
  weight pspwght ;
```

```
run ;
```

```
title "Numbers for Table 9.5 and 9.6 " ;
```

* Note use of descending option for dependent variable, so this matches Stata output ;

```
proc surveyl logistic data=c9_russia ;
```

```
  strata stratify ; cluster psu ; weight pspwght ;  
  class agecat (ref=first) marcat (ref=first) / param=ref ;  
  model stflife2 (descending) = agecat marcat male ;
```

```
run ;
```

```
ods text="No Design-Adjusted GOF test for Ordinal Logistic Regression in SAS" ;
```

```
title " 9.4.7 Example: Fitting Poisson and Negative Binomial Regression Models to Complex Sample Survey Data " ;
```

```
ods text= "SURVEY PROCEDURES do not include POISSON, NEGATIVE BINOMIAL OR ZERO-INFLATED NB MODELS, see website for user written SAS macros for these models" ;
```

```
libname d2 'p:\asda 2\data sets\hrs 2012 ' ;
```

```
data c9_hrs ;
```

```
  set d2.hrs_sub_28sep2016 ;
```

```
run ;
```

```
proc univariate data=c9_hrs ;
```

```
  var numfalls24 ;  
  where age65p=1 ;
```

```
run ;
```

```
proc univariate data=c9_hrs ;
```

```
  var numfalls24 ;  
  where age65p=1 and numfalls24 >=1 ;
```

```
run ;
```

```
ods rtf close ;
```

Example 9.2.6 Multinomial logistic regression using NCSR data.

The SURVEYLOGISTIC Procedure

Model Information		
Data Set	WORK.C9_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	1	3766	3671.4725
2	2	283	289.8166
3	3	1630	1705.8959

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.	

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	9011.140	7399.903
SC	9024.425	7559.322
-2 Log L	9007.140	7351.903

Testing Global Null Hypothesis: BETA=0				
Test	F Value	Num DF	Den DF	Pr > F
Likelihood Ratio	78.21	12.6078	529.53	<.0001
Score	29.89	22	21	<.0001
Wald	73.62	22	21	<.0001
NOTE: Second-order Rao-Scott design correction 0.7450 applied to the Likelihood Ratio test.				

Type 3 Analysis of Effects				
Effect	F Value	Num DF	Den DF	Pr > F
SEX	35.61	2	41	<.0001
ald	5.03	2	41	0.0111
mde	1.13	2	41	0.3316
ED4CAT	13.63	6	37	<.0001
ag4cat	83.25	6	37	<.0001
MAR3CAT	24.72	4	39	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	WKSTAT3C	Estimate	Standard Error	t Value	Pr > t
Intercept	2	-0.6438	0.2967	-2.17	0.0357
Intercept	3	-0.3795	0.1733	-2.19	0.0341
SEX	1 2	-1.3932	0.1980	-7.04	<.0001
SEX	1 3	-0.6403	0.1103	-5.81	<.0001
ald	2	-0.1638	0.3577	-0.46	0.6494
ald	3	0.3332	0.1305	2.55	0.0144
mde	2	-0.1398	0.1576	-0.89	0.3801
mde	3	0.0985	0.0882	1.12	0.2701
ED4CAT	2 2	-0.8470	0.2359	-3.59	0.0009
ED4CAT	2 3	-0.6514	0.1413	-4.61	<.0001
ED4CAT	3 2	-1.3653	0.2580	-5.29	<.0001
ED4CAT	3 3	-0.9169	0.1468	-6.25	<.0001
ED4CAT	4 2	-1.7310	0.3111	-5.56	<.0001
ED4CAT	4 3	-1.2295	0.1599	-7.69	<.0001
ag4cat	2 2	-0.8524	0.2951	-2.89	0.0061
ag4cat	2 3	-0.3164	0.1290	-2.45	0.0184
ag4cat	3 2	-0.8377	0.2586	-3.24	0.0023
ag4cat	3 3	0.0650	0.1712	0.38	0.7062
ag4cat	4 2	1.8284	0.2953	6.19	<.0001
ag4cat	4 3	2.3806	0.1738	13.70	<.0001
MAR3CAT	2 2	-0.5899	0.2257	-2.61	0.0124
MAR3CAT	2 3	-0.0523	0.1052	-0.50	0.6220
MAR3CAT	3 2	-2.7834	0.3807	-7.31	<.0001
MAR3CAT	3 3	0.5528	0.1326	4.17	0.0002

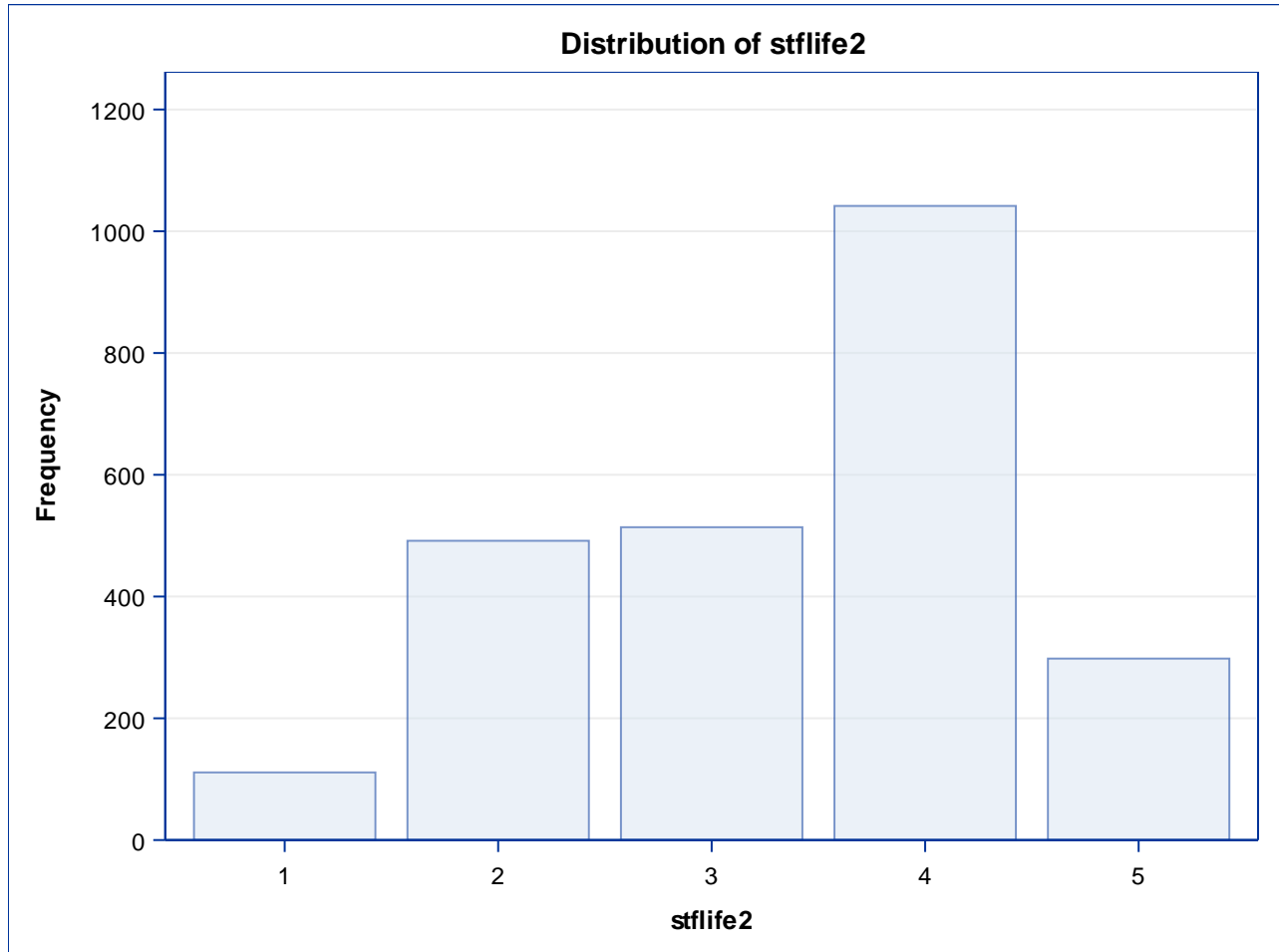
Analysis of Maximum Likelihood Estimates					
Parameter	WKSTAT3C	Estimate	Standard Error	t Value	Pr > t
NOTE: The degrees of freedom for the t tests is 42.					

Odds Ratio Estimates					
Effect	WKSTAT3C	Point Estimate	95% Confidence Limits		
SEX 1 vs 2	2	0.248	0.166	0.370	
SEX 1 vs 2	3	0.527	0.422	0.659	
ald	2	0.849	0.412	1.747	
ald	3	1.395	1.072	1.816	
mde	2	0.870	0.633	1.195	
mde	3	1.104	0.924	1.318	
ED4CAT 2 vs 1	2	0.429	0.266	0.690	
ED4CAT 2 vs 1	3	0.521	0.392	0.693	
ED4CAT 3 vs 1	2	0.255	0.152	0.430	
ED4CAT 3 vs 1	3	0.400	0.297	0.538	
ED4CAT 4 vs 1	2	0.177	0.095	0.332	
ED4CAT 4 vs 1	3	0.292	0.212	0.404	
ag4cat 2 vs 1	2	0.426	0.235	0.774	
ag4cat 2 vs 1	3	0.729	0.562	0.946	
ag4cat 3 vs 1	2	0.433	0.257	0.729	
ag4cat 3 vs 1	3	1.067	0.755	1.508	
ag4cat 4 vs 1	2	6.224	3.430	11.294	
ag4cat 4 vs 1	3	10.811	7.614	15.352	
MAR3CAT 2 vs 1	2	0.554	0.352	0.874	
MAR3CAT 2 vs 1	3	0.949	0.768	1.174	
MAR3CAT 3 vs 1	2	0.062	0.029	0.133	
MAR3CAT 3 vs 1	3	1.738	1.330	2.272	
NOTE: The degrees of freedom in computing the confidence limits is 42.					

Figure 9.6, Bar Chart of Satisfaction with Life, Weighted by PSPWGHT

The FREQ Procedure

stflife2	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	111.1011	4.52	111.1011	4.52
2	491.5778	20.01	602.6789	24.54
3	513.8991	20.92	1116.578	45.46
4	1041.598	42.41	2158.176	87.86
5	298.0937	12.14	2456.27	100.00
Frequency Missing = 27.729974041				



Numbers for Table 9.5 and 9.6
The SURVEYLOGISTIC Procedure

Model Information		
Data Set	WORK.C9_RUSSIA	
Response Variable	stflife2	
Number of Response Levels	5	
Stratum Variable	stratify	Stratification
Number of Strata	8	
Cluster Variable	psu	Primary Sampling Unit
Number of Clusters	184	
Weight Variable	PSPWGHT	Post-stratification weight including design weight
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	2484
Number of Observations Used	2415
Sum of Weights Read	2484
Sum of Weights Used	2422.138

Response Profile			
Ordered Value	stflife2	Total Frequency	Total Weight
1	5	283	294.8272
2	4	1036	1035.3251
3	3	492	502.8743
4	2	497	478.0106
5	1	107	111.1011

Probabilities modeled are cumulated over the lower Ordered Values.

Note: 69 observations were deleted due to missing values for the response or explanatory variables.
--

Class Level Information				
Class	Value	Design Variables		
agecat	1	0	0	0
	2	1	0	0
	3	0	1	0
	4	0	0	1
marcat	1	0	0	
	2	1	0	
	3	0	1	

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

Score Test for the Proportional Odds Assumption		
Chi-Square	DF	Pr > ChiSq

Score Test for the Proportional Odds Assumption		
Chi-Square	DF	Pr > ChiSq
22.4126	18	0.2142

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	6827.069	6766.264
SC	6850.238	6824.188
-2 Log L	6819.069	6746.264

Testing Global Null Hypothesis: BETA=0				
Test	F Value	Num DF	Den DF	Pr > F
Likelihood Ratio	11.25	5.6214	989.37	<.0001
Score	8.33	6	171	<.0001
Wald	7.60	6	171	<.0001

NOTE: Second-order Rao-Scott design correction 0.0673 applied to the Likelihood Ratio test.

Type 3 Analysis of Effects				
Effect	F Value	Num DF	Den DF	Pr > F
agecat	9.94	3	174	<.0001
marcat	2.18	2	175	0.1166
male	1.33	1	176	0.2509

Analysis of Maximum Likelihood Estimates					
Parameter		Estimate	Standard Error	t Value	Pr > t
Intercept	5	-1.3841	0.1575	-8.79	<.0001
Intercept	4	0.8348	0.1643	5.08	<.0001
Intercept	3	1.7928	0.1713	10.47	<.0001
Intercept	2	3.7111	0.2176	17.05	<.0001
agecat	2	-0.5293	0.1392	-3.80	0.0002
agecat	3	-0.7455	0.1463	-5.10	<.0001
agecat	4	-0.8081	0.1661	-4.86	<.0001
marcat	2	-0.2089	0.1049	-1.99	0.0480
marcat	3	-0.1372	0.1349	-1.02	0.3105
male		-0.1096	0.0952	-1.15	0.2509

NOTE: The degrees of freedom for the t tests is 176.

Odds Ratio Estimates			
Effect	Point Estimate	95% Confidence Limits	
agecat 2 vs 1	0.589	0.448	0.775
agecat 3 vs 1	0.474	0.356	0.633
agecat 4 vs 1	0.446	0.321	0.619
marcat 2 vs 1	0.811	0.660	0.998
marcat 3 vs 1	0.872	0.668	1.138
male	0.896	0.743	1.081

NOTE: The degrees of freedom in computing the confidence limits is 176.

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	54.1	Somers' D	0.149
Percent Discordant	39.2	Gamma	0.159
Percent Tied	6.7	Tau-a	0.106
Pairs	2089159	c	0.574

No Design-Adjusted GOF test for Ordinal Logistic Regression in SAS

SURVEY PROCEDURES do not include POISSON, NEGATIVE BINOMIAL OR ZERO-INFLATED NB MODELS, see website for user written SAS macros for these models

9.4.7 Example: Fitting Poisson and Negative Binomial Regression Models to Complex Sample Survey Data, NOT AVAILABLE IN SAS SURVEY PROCEDURES

The UNIVARIATE Procedure

Variable: numfalls24 (Number of Falls Past 2 Years)

Moments			
N	10594	Sum Weights	10594
Mean	1.19124032	Sum Observations	12620
Std Deviation	3.3183257	Variance	11.0112855
Skewness	7.98181453	Kurtosis	90.9991099
Uncorrected SS	131676	Corrected SS	116642.547
Coeff Variation	278.560559	Std Error Mean	0.03223955

Basic Statistical Measures			
Location		Variability	
Mean	1.191240	Std Deviation	3.31833
Median	0.000000	Variance	11.01129
Mode	0.000000	Range	50.00000
		Interquartile Range	1.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	36.94965	Pr > t 	<.0001
Sign	M	1917.5	Pr >= M 	<.0001
Signed Rank	S	3677765	Pr >= S 	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	50
99%	13
95%	5
90%	3
75% Q3	1
50% Median	0
25% Q1	0
10%	0
5%	0
1%	0
0% Min	0

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0	20540	50	9173
0	20536	50	9327
0	20515	50	9982
0	20498	50	11024
0	20494	50	11545

Missing Values			
Missing Value	Count	Percent Of	
		All Obs	Missing Obs
.	141	1.31	100.00

9.4.7 Example: Fitting Poisson and Negative Binomial Regression Models to Complex Sample Survey Data, NOT AVAILABLE IN SAS SURVEY PROCEDURES

The UNIVARIATE Procedure

Variable: numfalls24 (Number of Falls Past 2 Years)

Moments			
N	3835	Sum Weights	3835
Mean	3.29074316	Sum Observations	12620
Std Deviation	4.84896618	Variance	23.512473
Skewness	5.7153534	Kurtosis	42.9379926
Uncorrected SS	131676	Corrected SS	90146.8214
Coeff Variation	147.351706	Std Error Mean	0.07830085

Basic Statistical Measures			
Location		Variability	
Mean	3.290743	Std Deviation	4.84897
Median	2.000000	Variance	23.51247
Mode	1.000000	Range	49.00000
		Interquartile Range	2.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	42.02692	Pr > t 	<.0001
Sign	M	1917.5	Pr >= M 	<.0001
Signed Rank	S	3677765	Pr >= S 	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	50
99%	25
95%	10
90%	6
75% Q3	3
50% Median	2
25% Q1	1
10%	1
5%	1
1%	1
0% Min	1

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
1	20341	50	9173
1	19401	50	9327
1	19367	50	9982
1	19323	50	11024
1	19260	50	11545