Supplement to Session 2

S2.1 Logistic analysis using PROC LOGISTIC in SAS

Data from head injury study (example 2) in a SAS dataset, **head2,** with one observation for each of the 341 patients. Six variables in this dataset; patient (Patient number), age (Age in years), gcsmotor (GCS motor score), treat (Treatment), gos4 (GOS category), and bin_gos4 (Binary classification of GOS).

patient	age	gcsmotor	treat	gos4	bin_gos4
1	66	1	0	1	1
2	63	1	1	1	1
3	58	1	1	1	1
4	54	1	0	1	1
5	51	1	1	1	1
6	49	1	0	1	1
7	49	1	1	1	1
8	46	1	0	1	1
9	46	1	1	1	1
10	43	1	0	1	1
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
332	53	3	1	3	2
333	52	3	1	4	2
334	52	3	0	3	2
335	49	3	0	4	2
336	49	3	1	3	2
337	47	3	1	4	2
338	46	3	0	3	2
339	38	3	0	4	2
340	37	3	0	4	2
341	37	3	1	4	2

```
* Set up formats for GCS motor score, treatment and outcome *;

proc format;
  value motfmt 1 = 'None/Ext'
               2 = 'Flexion'
               3 = 'Localises';
  value trtfmt 0 = 'Control'
               1 = 'Treated';
  value gosfmt 1 = 'Good'
               2 = 'Moderate'
               3 = 'Severe'
               4 = 'Veg/Dead';
  value binfmt 1 = 'Favourable'
               2 = 'Unfavourable';
run;
```

```
* Fit binary logistic regression model for a favourable outcome *
* with treatment                                                  *;

proc logistic data=head2;
  class treat (ref='Control') / param=ref order=internal;
  model bin_gos4 (order=internal) = treat;
  format bin_gos4 binfmt. treat trtfmt.;
run;
```

Note: PROC LOGISTIC models the probability of the first Ordered Value in the Response Profile table when the response/outcome variable is binary. The (order=internal) option in the model statement specifies the levels of the response/outcome variable, bin_gos4, should be sorted by the unformatted values (1 and 2) and this is one way of ensuring it is the probability of a ‘Favourable’ outcome (bin_gos4 = 1) that is modelled.

The LOGISTIC Procedure

Model Information

Data Set	WORK.HEAD2
Response Variable	bin_gos4
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	341
Number of Observations Used	341

Response Variable

Response Profile

Ordered Value	bin_gos4	Total Frequency
1	**Favourable**	**170**
2	Unfavourable	171

Probability modeled is bin_gos4='Favourable'.

Explanatory Variables

Class Level Information

Class	Value	Design Variables
treat	Control	0
	Treated	1

Model Convergence Status

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

Hypothesis Testing

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	474.723	468.433
SC	478.555	476.097
-2 Log L	**472.723**	**464.433**

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	**8.2902**	1	0.0040
Score	**8.2562**	1	0.0041
Wald	**8.1885**	1	0.0042

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
treat	1	**8.1885**	0.0042

Estimation of Difference

Analysis of Maximum Likelihood Estimates

Parameter			DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept			1	-0.3302	0.1578	4.3782	0.0364
treat	Treated	$\hat{\beta}$	1	**0.6279**	**0.2194**	8.1885	0.0042

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
treat Treated vs Control	**1.874**	**1.219**	**2.880**

Association of Predicted Probabilities and Observed Responses

Percent Concordant	33.4	Somers' D	0.156
Percent Discordant	17.8	Gamma	0.304
Percent Tied	48.8	Tau-a	0.078
Pairs	29070	c	0.578

S2.2 Logistic analysis using PROC GENMOD in SAS

```
proc genmod data=head2 rorder=internal;
  class treat (ref='Control') / param=ref order=internal;
  model bin_gos4 = treat / d=binomial link=logit type1;
  format bin_gos4 binfmt. treat trtfmt.;
run;
```

Note: PROC GENMOD models the probability of the first Ordered Value in the Response Profile table when the response/outcome variable is binary. However to specify how the levels of the response/outcome variable should be sorted the rorder= option in the proc genmod statement needs to be used. So in this example with rorder=internal the two levels of bin_gos4 will be sorted by the unformatted values (1 and 2) and this results in the probability of a 'Favourable' outcome (bin_gos4 = 1) being modelled.

```
                         The GENMOD Procedure

                          Model Information

                 Data Set                WORK.HEAD2
                 Distribution              Binomial
                 Link Function                Logit
                 Dependent Variable        bin_gos4


             Number of Observations Read         341
             Number of Observations Used         341
             Number of Events                    170
             Number of Trials                    341


                       Class Level Information

                                             Design
                 Class     Value          Variables

                 treat     Control                0
                           Treated                1


                           Response Profile

               Ordered                          Total
                 Value    bin_gos4          Frequency

                     1    Favourable              170
                     2    Unfavourable            171

PROC GENMOD is modeling the probability that bin_gos4='Favourable'.


                Criteria For Assessing Goodness Of Fit

        Criterion                    DF           Value        Value/DF

        Log Likelihood                        -232.2166
        Full Log Likelihood                   -232.2166
        AIC (smaller is better)                468.4333
        AICC (smaller is better)               468.4688
        BIC (smaller is better)                476.0970
```

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates

Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square
Intercept			1	-0.3302	0.1578	-0.6396	-0.0209	4.38
treat	Treated	$\hat{\beta}$	**1**	**0.6279**	**0.2194**	**0.1978**	**1.0579**	8.19
Scale			0	1.0000	0.0000	1.0000	1.0000	

Analysis Of Maximum Likelihood Parameter Estimates

Parameter		Pr > ChiSq
Intercept		0.0364
treat	Treated	0.0042
Scale		

NOTE: The scale parameter was held fixed.

LR Statistics For Type 1 Analysis

Source	Deviance	DF	Chi-Square	Pr > ChiSq
Intercept	472.7234			
treat	464.4333	1	**8.29**	0.0040

S2.3 Extension of analysis using PROC LOGISTIC

```
proc logistic data=head2;
  model bin_gos4 (order=internal) = age;
  format bin_gos4 binfmt.;
  title1 'Model 1 fitting age';
run;

proc logistic data=head2;
  class treat (ref='Control') / param=ref order=internal;
  model bin_gos4 (order=internal) = age treat;
  format bin_gos4 binfmt. treat trtfmt.;
  title1 'Model 2 fitting age and treatment';
run;
```

```
                         Model 1 fitting age

                       The LOGISTIC Procedure

                          Model Information

              Data Set                      WORK.HEAD2
              Response Variable             bin_gos4
              Number of Response Levels     2
              Model                         binary logit
              Optimization Technique        Fisher's scoring


                 Number of Observations Read         341
                 Number of Observations Used         341


                          Response Profile

                  Ordered                            Total
                    Value     bin_gos4           Frequency

                        1     Favourable               170
                        2     Unfavourable             171

              Probability modeled is bin_gos4='Favourable'.


                      Model Convergence Status

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


                        Model Fit Statistics

                                                 Intercept
                                   Intercept           and
                  Criterion             Only    Covariates

                  AIC                474.723       468.600
                  SC                 478.555       476.264
                  -2 Log L           472.723       464.600
```

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	8.1233	1	0.0044
Score	8.0538	1	0.0045
Wald	7.9184	1	0.0049

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.7130	0.2771	6.6191	0.0101
age	1	-0.0203	0.00723	7.9184	0.0049

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
age	0.980	0.966	0.994

Association of Predicted Probabilities and Observed Responses

Percent Concordant	57.4	Somers' D	0.170
Percent Discordant	40.4	Gamma	0.174
Percent Tied	2.2	Tau-a	0.085
Pairs	29070	c	0.585

Model 2 fitting age and treatment

The LOGISTIC Procedure

Model Information

Data Set	WORK.HEAD2
Response Variable	bin_gos4
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	341
Number of Observations Used	341

Response Profile

Ordered Value	bin_gos4	Total Frequency
1	Favourable	170
2	Unfavourable	171

Probability modeled is bin_gos4='Favourable'.

Class Level Information

Class	Value	Design Variables
treat	Control	0
	Treated	1

Model Convergence Status

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

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	474.723	460.770
SC	478.555	472.265
-2 Log L	472.723	**454.770**

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	17.9539	2	0.0001
Score	17.5619	2	0.0002
Wald	16.7991	2	0.0002

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
age	1	9.3370	0.0022
treat	1	9.6396	0.0019

Analysis of Maximum Likelihood Estimates

Parameter			DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept			1	0.4337	0.2931	2.1899	0.1389
age		$\hat{\beta}_1$	1	**-0.0226**	0.00741	9.3370	0.0022
treat	Treated	$\hat{\beta}_2$	1	**0.6969**	0.2244	9.6396	0.0019

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
age	0.978	0.964	0.992
treat Treated vs Control	2.007	1.293	3.117

Association of Predicted Probabilities and Observed Responses

Percent Concordant	63.1	Somers' D	0.275
Percent Discordant	35.6	Gamma	0.278
Percent Tied	1.3	Tau-a	0.138
Pairs	29070	c	0.637

S2.4 Logistic analysis using lrm function in R

Data from head injury study (example 2) in R data frame, **head2,** which has the same data as in the SAS dataset described in section S2.1, i.e. one row for each of the 341 patients and six variables. Workspace containing just this data frame saved as head2.RData in project folder on PC.

```
# Load saved workspace which contains data frame with ungrouped data from
# head injury study (example 2)

load('head2.RData')


# Convert binary outcome variable from numeric variable to factor variable
# with 2 (Unfavourable) as first level and 1 (Favourable) as second level so
# lrm function in rms package will model probability of a favourable outcome

head2$BIN_GOS4 <- factor(head2$BIN_GOS4, levels=c('2','1'),
labels=c('Unfavourable','Favourable'))


# Convert categorical predictors from numeric variables to factor variables

head2$GCSMOTOR <- factor(head2$GCSMOTOR, levels=c('1','2','3'),
labels=c('None/Ext','Flexion','Localises'))
head2$TREAT <- factor(head2$TREAT, levels=c('0','1'),
labels=c('Control','Treated'))


# Load rms package

library(rms)


# Fit binary logistic regression model for a favourable outcome with treatment

lrm.fit1 <- lrm(BIN_GOS4 ~ TREAT, data=head2)
print(lrm.fit1)
cat('Deviance (-2 Log L)', "\n", lrm.fit1$deviance, "\n", "\n", "\n")
anova(lrm.fit1)
summary(lrm.fit1, TREAT='Control')
```

Note: The lrm function in the rms package models the probability of the second of the two values when the response/outcome variable is binary. So in this example to model the probability of a ‘Favourable‘ outcome this needs to be the second level of bin_gos4 and this ordering is the opposite of that required by PROC LOGISTIC and PROC GENMOD in SAS.

There are a small number of very minor differences in the results produced by R and SAS due to differences in the convergence criteria used by the algorithms for obtaining the maximum likelihood estimates of the model parameters. These differences are all too small to have any meaningful impact.

```
Logistic Regression Model

lrm(formula = BIN_GOS4 ~ TREAT, data = head2)

                     Model Likelihood     Discrimination    Rank Discrim.
                        Ratio Test           Indexes           Indexes

Obs           341    LR chi2      8.29    R2       0.032    C       0.578
 Unfavourable 171    d.f.            1    g        0.315    Dxy     0.156
 Favourable   170    Pr(> chi2) 0.0040    gr       1.370    gamma   0.304
max |deriv| 1e-12                         gp       0.078    tau-a   0.078
                                          Brier    0.244


                     Coef    S.E.   Wald Z Pr(>|Z|)
Intercept            -0.3302 0.1578 -2.09  0.0364
TREAT=Treated    β̂    0.6279 0.2194  2.86  0.0042


Deviance (-2 Log L)
 472.7234 464.4333


                Wald Statistics          Response: BIN_GOS4

 Factor     Chi-Square d.f. P
 TREAT      8.19       1    0.0042
 TOTAL      8.19       1    0.0042


             Effects              Response : BIN_GOS4

 Factor                      Low High Diff. Effect S.E. Lower 0.95 Upper 0.95
 TREAT - Treated:Control     1   2    NA    0.63   0.22 0.20       1.06
  Odds Ratio                 1   2    NA    1.87     NA 1.22       2.88
```

S2.5 Extension of analysis using lrm function

```
# Model fitting age

lrm.fit2 <- lrm(BIN_GOS4 ~ AGE, data=head2)
print(lrm.fit2)
cat('Deviance (-2 Log L)', "\n", lrm.fit2$deviance, "\n", "\n", "\n")
anova(lrm.fit2)
summary(lrm.fit2, AGE=c(34,34,35))

# Model fitting age and treatment

lrm.fit3 <- lrm(BIN_GOS4 ~ AGE + TREAT, data=head2)
print(lrm.fit3)
cat('Deviance (-2 Log L)', "\n", lrm.fit3$deviance, "\n", "\n", "\n")
anova(lrm.fit3)
summary(lrm.fit3, AGE=c(34,34,35), TREAT='Control')

# Likelihood ratio test for treatment adjusting for age by taking
# difference between -2 Log L for last two lrm function calls

lrtest(lrm.fit2, lrm.fit3)
```

Model fitting age

```
Logistic Regression Model

lrm(formula = BIN_GOS4 ~ AGE, data = head2)

                     Model Likelihood     Discrimination    Rank Discrim.
                        Ratio Test           Indexes           Indexes

Obs              341    LR chi2      8.12    R2       0.031    C       0.585
 Unfavourable 171    d.f.            1    g        0.355    Dxy     0.170
 Favourable   170    Pr(> chi2) 0.0044    gr       1.426    gamma   0.174
max |deriv| 1e-08                         gp       0.087    tau-a   0.085
                                          Brier    0.244


          Coef    S.E.   Wald Z Pr(>|Z|)
Intercept  0.7130 0.2771  2.57  0.0101
AGE       -0.0203 0.0072 -2.81  0.0049
```

Deviance (-2 Log L)
472.7234 464.6002

```
                Wald Statistics          Response: BIN_GOS4

 Factor     Chi-Square d.f. P
 AGE        7.92       1    0.0049
 TOTAL      7.92       1    0.0049


             Effects              Response : BIN_GOS4

 Factor       Low High Diff. Effect S.E. Lower 0.95 Upper 0.95
 AGE          34  35   1     -0.02  0.01 -0.03      -0.01
  Odds Ratio 34  35   1      0.98    NA  0.97       0.99
```

Model fitting age and treatment

```
Logistic Regression Model

lrm(formula = BIN_GOS4 ~ AGE + TREAT, data = head2)

                     Model Likelihood     Discrimination    Rank Discrim.
                        Ratio Test           Indexes           Indexes

Obs            341    LR chi2     17.95    R2       0.068    C       0.637
 Unfavourable 171    d.f.            2    g        0.540    Dxy     0.274
 Favourable   170    Pr(> chi2) 0.0001    gr       1.716    gamma   0.278
max |deriv| 5e-13                         gp       0.130    tau-a   0.138
                                          Brier    0.237
```

		Coef	S.E.	Wald Z	Pr(>|Z|)
Intercept		0.4338	0.2931	1.48	0.1388
AGE	$\hat{\beta}_1$	**-0.0226**	0.0074	-3.06	0.0022
TREAT=Treated	$\hat{\beta}_2$	**0.6970**	0.2245	3.11	0.0019

Deviance (-2 Log L)
472.7234 **454.7696**

```
                Wald Statistics          Response: BIN_GOS4

 Factor     Chi-Square d.f. P
 AGE         9.34      1    0.0022
 TREAT       9.64      1    0.0019
 TOTAL      16.81      2    0.0002


            Effects              Response : BIN_GOS4

 Factor                        Low High Diff. Effect S.E. Lower 0.95 Upper 0.95
 AGE                           34  35    1    -0.02  0.01 -0.04      -0.01
  Odds Ratio                   34  35    1     0.98    NA  0.96       0.99
 TREAT - Treated:Control        1   2   NA     0.70  0.22  0.26       1.14
  Odds Ratio                    1   2   NA     2.01    NA  1.29       3.12
```

Likelihood ratio test for treatment adjusting for age

```
Model 1: BIN_GOS4 ~ AGE
Model 2: BIN_GOS4 ~ AGE + TREAT

 L.R. Chisq        d.f.           P
9.830588747 1.000000000 0.001716334
```