Data from the SAS support on a study on cancer remission (Lee; [1974](http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/statug_logistic_sect067.htm#lee_e_74)). The data consist of patient characteristics and whether or not cancer remission occurred.

**data** Remission;

input remiss cell smear infil li blast temp;

label remiss='Complete Remission';

datalines;

1 .8 .83 .66 1.9 1.1 .996

1 .9 .36 .32 1.4 .74 .992

0 .8 .88 .7 .8 .176 .982

0 1 .87 .87 .7 1.053 .986

1 .9 .75 .68 1.3 .519 .98

0 1 .65 .65 .6 .519 .982

1 .95 .97 .92 1 1.23 .992

0 .95 .87 .83 1.9 1.354 1.02

0 1 .45 .45 .8 .322 .999

0 .95 .36 .34 .5 0 1.038

0 .85 .39 .33 .7 .279 .988

0 .7 .76 .53 1.2 .146 .982

0 .8 .46 .37 .4 .38 1.006

0 .2 .39 .08 .8 .114 .99

0 1 .9 .9 1.1 1.037 .99

1 1 .84 .84 1.9 2.064 1.02

0 .65 .42 .27 .5 .114 1.014

0 1 .75 .75 1 1.322 1.004

0 .5 .44 .22 .6 .114 .99

1 1 .63 .63 1.1 1.072 .986

0 1 .33 .33 .4 .176 1.01

0 .9 .93 .84 .6 1.591 1.02

1 1 .58 .58 1 .531 1.002

0 .95 .32 .3 1.6 .886 .988

1 1 .6 .6 1.7 .964 .99

1 1 .69 .69 .9 .398 .986

0 1 .73 .73 .7 .398 .986

;

**proc** **logistic** data=Remission;

model remiss(event='1')=cell smear infil li blast temp/lackfit;

**run**;

|  |
| --- |
| The SAS System |

The LOGISTIC Procedure

|  |  |  |
| --- | --- | --- |
| **Model Information** | | |
| **Data Set** | WORK.REMISSION |  |
| **Response Variable** | remiss | Complete Remission |
| **Number of Response Levels** | 2 |  |
| **Model** | binary logit |  |
| **Optimization Technique** | Fisher's scoring |  |

|  |  |
| --- | --- |
| **Number of Observations Read** | 27 |
| **Number of Observations Used** | 27 |

|  |  |  |
| --- | --- | --- |
| **Response Profile** | | |
| **Ordered Value** | **remiss** | **Total Frequency** |
| **1** | 0 | 18 |
| **2** | 1 | 9 |

Probability modeled is remiss=1.

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

|  |  |  |
| --- | --- | --- |
| **Model Fit Statistics** | | |
| **Criterion** | **Intercept Only** | **Intercept and Covariates** |
| **AIC** | 36.372 | 35.751 |
| **SC** | 37.668 | 44.822 |
| **-2 Log L** | 34.372 | 21.751 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Testing Global Null Hypothesis: BETA=0** | | | |
| **Test** | **Chi-Square** | **DF** | **Pr > ChiSq** |
| **Likelihood Ratio** | 12.6211 | 6 | 0.0495 |
| **Score** | 9.4609 | 6 | 0.1493 |
| **Wald** | 4.5302 | 6 | 0.6053 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Analysis of Maximum Likelihood Estimates** | | | | | |
| **Parameter** | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| **Intercept** | 1 | 58.0385 | 71.2364 | 0.6638 | 0.4152 |
| **cell** | 1 | 24.6613 | 47.8376 | 0.2658 | 0.6062 |
| **smear** | 1 | 19.2933 | 57.9499 | 0.1108 | 0.7392 |
| **infil** | 1 | -19.6009 | 61.6814 | 0.1010 | 0.7507 |
| **li** | 1 | 3.8960 | 2.3371 | 2.7789 | 0.0955 |
| **blast** | 1 | 0.1511 | 2.2786 | 0.0044 | 0.9471 |
| **temp** | 1 | -87.4337 | 67.5735 | 1.6742 | 0.1957 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Odds Ratio Estimates** | | | |
| **Effect** | **Point Estimate** | **95% Wald Confidence Limits** | |
| **cell** | >999.999 | <0.001 | >999.999 |
| **smear** | >999.999 | <0.001 | >999.999 |
| **infil** | <0.001 | <0.001 | >999.999 |
| **li** | 49.203 | 0.504 | >999.999 |
| **blast** | 1.163 | 0.013 | 101.191 |
| **temp** | <0.001 | <0.001 | >999.999 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Association of Predicted Probabilities and Observed Responses** | | | |
| **Percent Concordant** | 88.3 | **Somers' D** | 0.765 |
| **Percent Discordant** | 11.7 | **Gamma** | 0.765 |
| **Percent Tied** | 0.0 | **Tau-a** | 0.353 |
| **Pairs** | 162 | **c** | 0.883 |

Here some contraceptive, showing the distribution of 1607 currently married and fecund women interviewed in the Fiji Fertility Survey, according to age, education, desire for more children and current use of contraception.

**data** conception;

input age $ education $ wantsMore $ notUsing using;

total = notUsing + using;

cards;

<25 low yes 53 6

<25 low no 10 4

<25 high yes 212 52

<25 high no 50 10

25-29 low yes 60 14

25-29 low no 19 10

25-29 high yes 155 54

25-29 high no 65 27

30-39 low yes 112 33

30-39 low no 77 80

30-39 high yes 118 46

30-39 high no 68 78

40-49 low yes 35 6

40-49 low no 46 48

40-49 high yes 8 8

40-49 high no 12 31

;

**proc** **logistic** data=conception;

class age education wantsMore;

model notUsing/total = age education wantsMore;

**run**;

**quit**;

|  |
| --- |
| The SAS System |
| Predicted Probabilities and 95% Confidence Limits of best model |

The LOGISTIC Procedure

|  |  |
| --- | --- |
| **Model Information** | |
| **Data Set** | WORK.CONCEPTION |
| **Response Variable (Events)** | notUsing |
| **Response Variable (Trials)** | total |
| **Model** | binary logit |
| **Optimization Technique** | Fisher's scoring |

|  |  |
| --- | --- |
| **Number of Observations Read** | 16 |
| **Number of Observations Used** | 16 |
| **Sum of Frequencies Read** | 1607 |
| **Sum of Frequencies Used** | 1607 |

|  |  |  |
| --- | --- | --- |
| **Response Profile** | | |
| **Ordered Value** | **Binary Outcome** | **Total Frequency** |
| **1** | Event | 1100 |
| **2** | Nonevent | 507 |

Note the default settings in SAS for the design matrix

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class Level Information** | | | | |
| **Class** | **Value** | **Design Variables** | | |
| **age** | **25-29** | 1 | 0 | 0 |
|  | **30-39** | 0 | 1 | 0 |
|  | **40-49** | 0 | 0 | 1 |
|  | **<25** | -1 | -1 | -1 |
| **education** | **high** | 1 |  |  |
|  | **low** | -1 |  |  |

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

|  |  |  |  |
| --- | --- | --- | --- |
| **Model Fit Statistics** | | | |
| **Criterion** | **Intercept Only** | **Intercept and Covariates** | **With Constant** |
| **AIC** | 2005.694 | 1928.339 | 161.926 |
| **SC** | 2011.076 | 1955.250 | 188.837 |
| **-2 Log L** | 2003.694 | 1918.339 | 151.926 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Testing Global Null Hypothesis: BETA=0** | | | |
| **Test** | **Chi-Square** | **DF** | **Pr > ChiSq** |
| **Likelihood Ratio** | 85.3541 | 4 | <.0001 |
| **Score** | 83.5708 | 4 | <.0001 |
| **Wald** | 79.8091 | 4 | <.0001 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Type 3 Analysis of Effects** | | | |
| **Effect** | **DF** | **Wald Chi-Square** | **Pr > ChiSq** |
| **age** | 3 | 79.1370 | <.0001 |
| **education** | 1 | 6.0783 | 0.0137 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Analysis of Maximum Likelihood Estimates** | | | | | | |
| **Parameter** |  | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| **Intercept** |  | 1 | 0.8044 | 0.0615 | 171.0876 | <.0001 |
| **age** | **25-29** | 1 | 0.3197 | 0.1022 | 9.7920 | 0.0018 |
| **age** | **30-39** | 1 | -0.3411 | 0.0847 | 16.2131 | <.0001 |
| **age** | **40-49** | 1 | -0.7804 | 0.1240 | 39.6413 | <.0001 |
| **education** | **high** | 1 | -0.1500 | 0.0608 | 6.0783 | 0.0137 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Odds Ratio Estimates** | | | | | | |
| **Effect** | | **Point Estimate** | | **95% Wald Confidence Limits** | | |
| **age 25-29 vs <25** | | 0.618 | | 0.440 | | 0.867 |
| **age 30-39 vs <25** | | 0.319 | | 0.233 | | 0.436 |
| **age 40-49 vs <25** | | 0.206 | | 0.137 | | 0.307 |
| **education high vs low** | | 0.741 | | 0.584 | | 0.940 |
| **Association of Predicted Probabilities and Observed Responses** | | | | | | |
| **Percent Concordant** | 55.9 | | **Somers' D** | | 0.273 | |
| **Percent Discordant** | 28.6 | | **Gamma** | | 0.323 | |
| **Percent Tied** | 15.5 | | **Tau-a** | | 0.118 | |
| **Pairs** | 557700 | | **c** | | 0.636 | |

We can also fit this model with proc genmod using the following (note the referent groups have changed).

**proc** **genmod** data = conception;

class age (ref="25-29" param=ref) education (ref="low" param=ref) wantsMore (ref="no" param=ref);

model notUsing = age education wantsMore / type3 dist=poisson link=log offset=ltot;

**run**;

|  |
| --- |
| The SAS System |

The GENMOD Procedure

|  |  |
| --- | --- |
| **Model Information** | |
| **Data Set** | WORK.CONCEPTION |
| **Distribution** | Poisson |
| **Link Function** | Log |
| **Dependent Variable** | notUsing |
| **Offset Variable** | ltot |

|  |  |
| --- | --- |
| **Number of Observations Read** | 16 |
| **Number of Observations Used** | 16 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class Level Information** | | | | |
| **Class** | **Value** | **Design Variables** | | |
| **age** | **25-29** | 0 | 0 | 0 |
|  | **30-39** | 1 | 0 | 0 |
|  | **40-49** | 0 | 1 | 0 |
|  | **<25** | 0 | 0 | 1 |
| **education** | **high** | 1 |  |  |
|  | **low** | 0 |  |  |
| **wantsMore** | **no** | 0 |  |  |
|  | **yes** | 1 |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Criteria For Assessing Goodness Of Fit** | | | |
| **Criterion** | **DF** | **Value** | **Value/DF** |
| **Deviance** | 10 | 14.7987 | 1.4799 |
| **Scaled Deviance** | 10 | 14.7987 | 1.4799 |
| **Pearson Chi-Square** | 10 | 15.0292 | 1.5029 |
| **Scaled Pearson X2** | 10 | 15.0292 | 1.5029 |
| **Log Likelihood** |  | 3871.7418 |  |
| **Full Log Likelihood** |  | -53.0467 |  |
| **AIC (smaller is better)** |  | 118.0934 |  |
| **AICC (smaller is better)** |  | 127.4267 |  |
| **BIC (smaller is better)** |  | 122.7289 |  |

|  |
| --- |
| Algorithm converged. |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Analysis Of Maximum Likelihood Parameter Estimates** | | | | | | | | |
| **Parameter** |  | **DF** | **Estimate** | **Standard Error** | **Wald 95% Confidence Limits** | | **Wald Chi-Square** | **Pr > ChiSq** |
| **Intercept** |  | 1 | -0.4309 | 0.0912 | -0.6096 | -0.2521 | 22.33 | <.0001 |
| **age** | **30-39** | 1 | -0.1595 | 0.0801 | -0.3166 | -0.0025 | 3.96 | 0.0465 |
| **age** | **40-49** | 1 | -0.2828 | 0.1219 | -0.5217 | -0.0439 | 5.38 | 0.0203 |
| **age** | **<25** | 1 | 0.0791 | 0.0806 | -0.0788 | 0.2370 | 0.96 | 0.3262 |
| **education** | **high** | 1 | -0.0962 | 0.0666 | -0.2267 | 0.0344 | 2.08 | 0.1488 |
| **wantsMore** | **yes** | 1 | 0.2754 | 0.0688 | 0.1405 | 0.4102 | 16.01 | <.0001 |
| **Scale** |  | 0 | 1.0000 | 0.0000 | 1.0000 | 1.0000 |  |  |

|  |  |
| --- | --- |
| Note: | The scale parameter was held fixed. |

|  |  |  |  |
| --- | --- | --- | --- |
| **LR Statistics For Type 3 Analysis** | | | |
| **Source** | **DF** | **Chi-Square** | **Pr > ChiSq** |
| **age** | 3 | 12.87 | 0.0049 |
| **education** | 1 | 2.07 | 0.1502 |
| **wantsMore** | 1 | 16.43 | <.0001 |

The number of awards earned by students at one high school. Predictors of the number of awards earned include the type of program in which the student was enrolled (e.g., vocational, general or academic) and the score on their final exam in math. 1 = "General", 2 = "Academic" and 3 = "Vocational";

**data** pois;

input id num\_awards prog math;

prog\_cat = 'Gen';

if prog = **2** then prog\_cat = 'Acad';

if prog = **3** then prog\_cat = 'Voca';

cards;

45 0 3 41

108 0 1 41

15 0 3 44

67 0 3 42

153 0 3 40

…………………Data edited………

100 2 2 71

143 2 3 75

68 1 2 71

57 0 2 72

132 3 2 73

**run**;

**proc** **freq** data = pois;

tables prog\_cat;

**run**;

|  |
| --- |
| The SAS System |

The FREQ Procedure

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **prog\_cat** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **Aca** | 105 | 52.50 | 105 | 52.50 |
| **Gen** | 45 | 22.50 | 150 | 75.00 |
| **Voc** | 50 | 25.00 | 200 | 100.00 |

**proc** **genmod** data = pois;

class prog\_cat /param=glm;

model num\_awards = prog\_cat math / type3 dist=poisson;

**run**;

|  |
| --- |
| The SAS System |

The GENMOD Procedure

|  |  |
| --- | --- |
| **Model Information** | |
| **Data Set** | WORK.POIS |
| **Distribution** | Poisson |
| **Link Function** | Log |
| **Dependent Variable** | num\_awards |

|  |  |
| --- | --- |
| **Number of Observations Read** | 200 |
| **Number of Observations Used** | 200 |

|  |  |  |
| --- | --- | --- |
| **Class Level Information** | | |
| **Class** | **Levels** | **Values** |
| **prog\_cat** | 3 | Aca Gen Voc |

|  |  |  |  |
| --- | --- | --- | --- |
| **Criteria For Assessing Goodness Of Fit** | | | |
| **Criterion** | **DF** | **Value** | **Value/DF** |
| **Deviance** | 196 | 189.4496 | 0.9666 |
| **Scaled Deviance** | 196 | 189.4496 | 0.9666 |
| **Pearson Chi-Square** | 196 | 212.1437 | 1.0824 |
| **Scaled Pearson X2** | 196 | 212.1437 | 1.0824 |
| **Log Likelihood** |  | -135.1052 |  |
| **Full Log Likelihood** |  | -182.7523 |  |
| **AIC (smaller is better)** |  | 373.5045 |  |
| **AICC (smaller is better)** |  | 373.7096 |  |
| **BIC (smaller is better)** |  | 386.6978 |  |

|  |
| --- |
| Algorithm converged. |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Analysis Of Maximum Likelihood Parameter Estimates** | | | | | | | | | | | | | | | |
| **Parameter** | |  | | **DF** | | **Estimate** | **Standard Error** | | **Wald 95% Confidence Limits** | | | **Wald Chi-Square** | | | **Pr > ChiSq** |
| **Intercept** | |  | | 1 | | -4.8773 | 0.6282 | | -6.1085 | | -3.6461 | 60.28 | | | <.0001 |
| **prog\_cat** | | **Aca** | | 1 | | 0.7140 | 0.3200 | | 0.0868 | | 1.3413 | 4.98 | | | 0.0257 |
| **prog\_cat** | | **Gen** | | 1 | | -0.3698 | 0.4411 | | -1.2343 | | 0.4947 | 0.70 | | | 0.4018 |
| **prog\_cat** | | **Voc** | | 0 | | 0.0000 | 0.0000 | | 0.0000 | | 0.0000 | . | | | . |
| **math** | |  | | 1 | | 0.0702 | 0.0106 | | 0.0494 | | 0.0909 | 43.81 | | | <.0001 |
| **Scale** | |  | | 0 | | 1.0000 | 0.0000 | | 1.0000 | | 1.0000 |  | | |  |
| Note: | | The scale parameter was held fixed. | | | | | | | | | | |
| **LR Statistics For Type 3 Analysis** | | | | | | | | | | | | |
| **Source** | | | | | **DF** | | | **Chi-Square** | | **Pr > ChiSq** | | |
| **prog\_cat** | | | | | 2 | | | 14.57 | | 0.0007 | | |
| **math** | | | | | 1 | | | 45.01 | | <.0001 | | |

Now we'll go back to the conception data and analyze it with a Poisson model with an offset term. First, we need to create the offset term.

**data** conception2;

set conception;

ltot = log(total);

**run**;

**proc** **genmod** data = conception2;

class age (ref="25-29" param=ref) education (ref="low" param=ref) wantsMore (ref="no" param=ref);

model notUsing = age education wantsMore age\*education / type3 dist=poisson link=log offset=ltot;

**run**;

|  |
| --- |
| The SAS System |

The GENMOD Procedure

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model Information** | | | | | | | | |
| **Data Set** | | | WORK.CONCEPTION | | | | | |
| **Distribution** | | | Poisson | | | | | |
| **Link Function** | | | Log | | | | | |
| **Dependent Variable** | | | notUsing | | | | | |
| **Offset Variable** | | | ltot | | | | | |
| **Number of Observations Read** | | | | | | 16 |
| **Number of Observations Used** | | | | | | 16 |
| **Class Level Information** | | | | | | | |
| **Class** | **Value** | **Design Variables** | | | | | |
| **age** | **25-29** | 0 | | 0 | 0 | | |
|  | **30-39** | 1 | | 0 | 0 | | |
|  | **40-49** | 0 | | 1 | 0 | | |
|  | **<25** | 0 | | 0 | 1 | | |
| **education** | **high** | 1 | |  |  | | |
|  | **low** | 0 | |  |  | | |
| **wantsMore** | **no** | 0 | |  |  | | |
|  | **yes** | 1 | |  |  | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **Criteria For Assessing Goodness Of Fit** | | | |
| **Criterion** | **DF** | **Value** | **Value/DF** |
| **Deviance** | 7 | 10.6279 | 1.5183 |
| **Scaled Deviance** | 7 | 10.6279 | 1.5183 |
| **Pearson Chi-Square** | 7 | 11.0160 | 1.5737 |
| **Scaled Pearson X2** | 7 | 11.0160 | 1.5737 |
| **Log Likelihood** |  | 3873.8272 |  |
| **Full Log Likelihood** |  | -50.9613 |  |
| **AIC (smaller is better)** |  | 119.9226 |  |
| **AICC (smaller is better)** |  | 149.9226 |  |
| **BIC (smaller is better)** |  | 126.8759 |  |

|  |
| --- |
| Algorithm converged. |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Analysis Of Maximum Likelihood Parameter Estimates** | | | | | | | | | |
| **Parameter** |  | | **DF** | **Estimate** | **Standard Error** | **Wald 95% Confidence Limits** | | **Wald Chi-Square** | **Pr > ChiSq** |
| **Intercept** |  |  | 1 | -0.4689 | 0.1244 | -0.7127 | -0.2251 | 14.21 | 0.0002 |
| **age** | **30-39** |  | 1 | -0.1403 | 0.1348 | -0.4046 | 0.1239 | 1.08 | 0.2980 |
| **age** | **40-49** |  | 1 | -0.1331 | 0.1606 | -0.4478 | 0.1816 | 0.69 | 0.4071 |
| **age** | **<25** |  | 1 | 0.0952 | 0.1690 | -0.2360 | 0.4264 | 0.32 | 0.5732 |
| **education** | **high** |  | 1 | -0.0420 | 0.1312 | -0.2991 | 0.2151 | 0.10 | 0.7488 |
| **wantsMore** | **yes** |  | 1 | 0.2733 | 0.0688 | 0.1384 | 0.4082 | 15.78 | <.0001 |
| **age\*education** | **30-39** | **high** | 1 | -0.0134 | 0.1670 | -0.3407 | 0.3139 | 0.01 | 0.9360 |
| **age\*education** | **40-49** | **high** | 1 | -0.5196 | 0.2820 | -1.0724 | 0.0332 | 3.39 | 0.0654 |
| **age\*education** | **<25** | **high** | 1 | -0.0247 | 0.1921 | -0.4012 | 0.3518 | 0.02 | 0.8976 |
| **Scale** |  |  | 0 | 1.0000 | 0.0000 | 1.0000 | 1.0000 |  |  |

|  |  |
| --- | --- |
| Note: | The scale parameter was held fixed. |

|  |  |  |  |
| --- | --- | --- | --- |
| **LR Statistics For Type 3 Analysis** | | | |
| **Source** | **DF** | **Chi-Square** | **Pr > ChiSq** |
| **age** | 3 | 3.13 | 0.3717 |
| **education** | 1 | 0.10 | 0.7496 |
| **wantsMore** | 1 | 16.18 | <.0001 |
| **age\*education** | 3 | 4.17 | 0.2436 |