

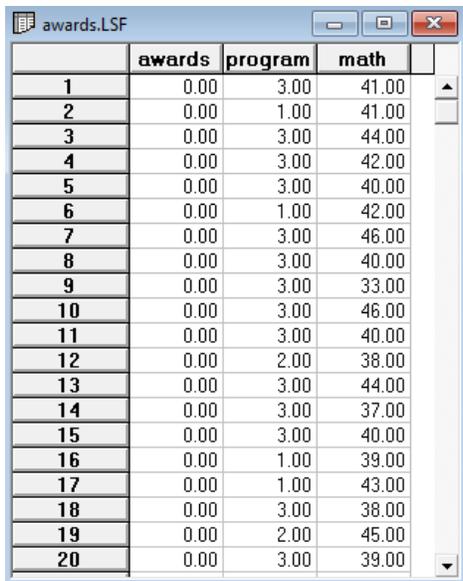
## Poisson regression with categorical covariate

### Contents

1. Introduction .....	1
2. Poisson-log model.....	2
3. Choice of appropriate reference category .....	4

### 1. Introduction

In this example we illustrate how to include a categorical variable as predictor into a Poisson model fitted using the GLIM module. The data used to illustrate is from a high school study where the number of awards earned and the math scores of 200 students were recorded. Additional information on the type of program they were enrolled in (vocational, general or academic) was also recorded. When the type of program is compared to say, for example, age categories, we note that here we are dealing with a true categorical variable and that we need to accommodate the nature of this variable in the analysis. The first 20 lines of the data file **awards.lsf** are shown below. Data and syntax files can be found in the **MVABOOK\Chapter3** folder.



	awards	program	math
1	0.00	3.00	41.00
2	0.00	1.00	41.00
3	0.00	3.00	44.00
4	0.00	3.00	42.00
5	0.00	3.00	40.00
6	0.00	1.00	42.00
7	0.00	3.00	46.00
8	0.00	3.00	40.00
9	0.00	3.00	33.00
10	0.00	3.00	46.00
11	0.00	3.00	40.00
12	0.00	2.00	38.00
13	0.00	3.00	44.00
14	0.00	3.00	37.00
15	0.00	3.00	40.00
16	0.00	1.00	39.00
17	0.00	1.00	43.00
18	0.00	3.00	38.00
19	0.00	2.00	45.00
20	0.00	3.00	39.00



On the **Titles and Options** dialog box, we enter a title and proceed to the **Distributions and Links** tab by using the **Next** button. Here we select Poisson as the **distribution type** and leave the **link function** field at its default (Log) value for a Poisson model.

On the **Dependent and Independent Variables** tab AWARDS is selected as outcome. Click next to move to the final tab and then **Finish** to generate the syntax file shown below.

```

awards.PRL
MGLimOptions Converge=0.0001 MaxIter=100 MissingCode=-999999
                Method=Quad NQUADPTS=10 ;
Title=;
SY=awards.LSF';
DEPENDENT_MISS=-999999;
Distribution=POI;
Link=LOG;
Intercept=Yes;
Scale=None;
DepVar=awards;

```

To accommodate the two covariates, we add the line

```
CoVARS=program$ math;
```

to the generated syntax file (see **awards1.prl**). The use of “\$” indicates that the variable is categorical. In this analysis, we have one continuous and one categorical covariate. LISREL will automatically a number of dummies equal to one less than the number of categories present for the categorical covariate (in this case, 2) and use the last category as the reference category. If this is not an appropriate choice, the reference category can be set using the REFCATS statement at the end of the syntax file. In this example, we use the first category as reference category (i.e., the vocational program).

The estimated regression results are given below.

Estimated Regression Weights

Parameter	Estimate	Standard Error	z Value	P Value
intcept	-4.7626	0.6126	-7.7745	0.0000
program2	0.5993	0.1722	3.4798	0.0005
program3	-0.1147	0.2274	-0.5045	0.6139
math	0.0702	0.0106	6.6187	0.0000

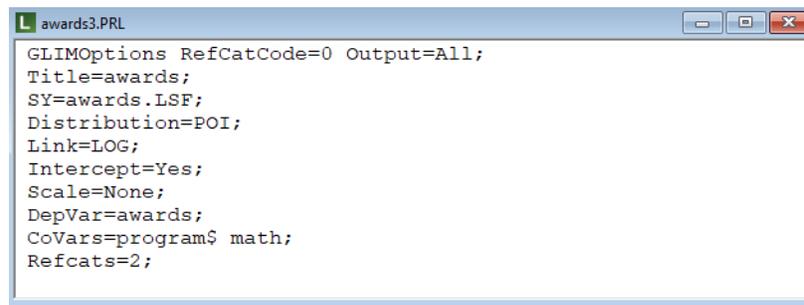
Reference Category(ies) specified for Covariate(s)

Covariate	Reference Category
program\$	1

From the results, we conclude that the estimated coefficients of math and the second category of program (general program, denoted as program2 here) are statistically significant.

### 3. Choice of appropriate reference category

In the previous analysis, we used the vocation group as the reference category. We now explore what happens when the second, and by far largest, program category is used as reference group. In **awards3.prl** we have amended the Refcat statement accordingly.



```
awards3.PRL
GLIMOptions RefCatCode=0 Output=All;
Title=awards;
SY=awards.LSF;
Distribution=POI;
Link=LOG;
Intercept=Yes;
Scale=None;
DepVar=awards;
CoVars=program$ math;
Refcats=2;
```

For this model, the estimated regressions weights are:

#### Estimated Regression Weights

Parameter	Estimate	Standard Error	z Value	P Value
intcept	-4.1633	0.6629	-6.2806	0.0000
program1	-1.0839	0.3583	-3.0254	0.0025
program3	-0.7140	0.3200	-2.2313	0.0257
math	0.0702	0.0106	6.6186	0.0000

In this case, the estimated coefficients of all predictors are statistically significant. We also see that the intcept estimate has changed, but the math estimate is the same as in the previous model. There are a number of options that can be used here; the reader is referred to Section 3.3.2 of *Multivariate Analysis with LISREL* (Joreskog, Olsson and Wallentin, 2016) for a detailed discussion of options and their implications for the estimated coefficients.