

Fitting a Poisson Regression Model for Rate Data

Let's see how to fit a Poisson regression model for rate data.

The syntax for PROC GENMOD is very similar to what we've used previously. This time we specify the skin cancer data set **mydata.skin**. In the CLASS statement we specify the two categorical predictor variables **City** and **Age**. In the MODEL statement we specify the outcome as **Cases**, and the predictors as **City** and **Age**. Then we use the OFFSET option to specify a variable in the input data set to be used as the offset variable. This variable cannot be a classification variable, and it cannot be the response variable or one of the other explanatory variables. In this case we use the **Log_Pop** variable. This variable contains the log of the population of each age group at risk within each city.

```
proc genmod data=mydata.skin;
  class city age;
  model cases= city age / offset=log_pop dist=poi link=log type3;
  title 'Poisson Regression Model for Skin Cancer Rates';
run;
```

Let's run the program and review the results. The Model Information table shows that the offset variable is the log of the population.

If you used reference cell coding in the CLASS statement (PARAM=REF), you would see a different presentation of the Class Level Information table. The default coding is GLM coding. Note that you can use whichever parameterization works best for the problem being solved.

The Criteria For Assessing Goodness Of Fit table shows that the model fits your data reasonably well because the Value/DF value is close to 1 for the scaled deviance and scaled Pearson chi-square statistics. So the model fits quite well.

The LR Statistics For Type 3 Analysis table confirms that the variables **City** and **Age** are significantly related to the outcome. In other words, there is a significant difference in skin cancer rates between Minneapolis-St. Paul and Dallas-Fort Worth, and there is at least one significant difference in skin cancer rates between the age groups.

The Analysis Of Maximum Likelihood Parameter Estimates table shows that there is a significant difference in skin cancer rates between Dallas-Fort Worth and Minneapolis-St. Paul. The positive coefficient for Dallas-Fort Worth indicates that Dallas-Fort Worth has a higher skin cancer rate. There are also significant differences between the age groups 15-24 and 85+, 25-34 and 85+, 35-44 and 85+, 45-54 and 85+, 55-64 and 85+, and 65-74 and 85+. The only group that shows no significant difference in skin cancer rates when compared to the 85+ age group is the age group 75-84.

You can write CONTRAST or ESTIMATE statements (or both) with the EXP option to obtain meaningful contrasts of interests. More information can be found in the Fitting Poisson Regression Models Using the GENMOD Procedure Live Web class.