An Animated Guide: An Introduction To Poisson Regression  
Russ Lavery, K&L Consulting Services, King of Prussia, PA, U.S.A.

ABSTRACT: 
This paper will be a brief introduction to Poisson regression (theory, steps to be followed, complications and interpretation) via a worked example. It is hoped that this will increase motivation towards learning this useful statistical technique.

INTRODUCTION: 
Poisson regression is available in SAS through the GENMOD procedure (generalized modeling). It is appropriate when: 1) the process that generates the conditional Y distributions would, theoretically, be expected to be a Poisson random process and 2) when there is no evidence of overdispersion and 3) when the mean of the marginal distribution is less than ten (preferably less than five and ideally close to one).

THE POISSON DISTRIBUTION: 
The Poisson distribution is a discrete distribution and is appropriate for modeling counts of observations. Counts are observed cases, like the count of measles cases in cities. You can simply model counts if all data were collected in the same measuring unit (e.g. the same number of days or same number of square feet).

You can use the Poisson Distribution for modeling rates (rates are counts per unit) if the units of collection were different.

If the cities being investigated have different populations (or the data is collected over different amounts of time) then it would be appropriate to model a measles rate. For counts collected over different time periods, or for cities with different numbers of inhabitants, we might model a rate (count / ( population*months)).

Unlike the familiar normal distribution, which is described by two parameters (mean and variance), the Poisson distribution is completely described by just one parameter, lambda (λ). Lambda is the average of a Poisson distribution as well as the variance and λ can take on non-integer values. While it is impossible to have 1.5 cases of measles in a city, it is possible to have the average number of cases per 1000 person-months be non-integer (like 3.12 cases/1000 person-months).

Figure 1 shows the formula for the Poisson distribution and plots of the Poisson distribution for several different values of λ. Note how the distribution "spreads out" as lambda increases. This is to be expected since, for a Poisson distribution, λ is both the mean and the variance.

The smooth curves in Figure 1 are a bit deceiving. Poisson distributions are discreet distributions and can only take on integer values. The values that the Poisson distributions can assume are show by the different symbols used in the illustration above – not by the curved lines connecting the symbols.

The curved, smooth lines, connecting the symbols, make it seem like a value of X=5.5 could be observed and that is not true. To be accurately represented, a Poisson distribution should be represented by a histogram, not by using smooth curves as was done above. However; the graphic above lets us, in one graph, easily compare several distributions with several different values of lambda. The graphic has uses as well as flaws.
THE HISTORY OF THE POISSON AND COMPUTING POISSON PROBABILITIES

The Poisson distribution was not made famous by Poisson. It was made famous by Ladislaus von Bortkiewicz who used it to model the number of Prussian calvarymen who were killed by getting kicked by a horse. The world (or the folks who read statistical journals back then) was exited by how close the Poisson distribution modeled the observed data.

Bortkiewicz was able to get fatality-by-kicking data for twenty one-year periods from each of ten army corps. This gave him 200 data points and a unit of measure of “deaths/corps-year” (and he assumed all corps were of approximately equal size and had horses of comparable training). There were 122 fatalities in the 200 observations and a mean deaths per corps-year of .61. \( \lambda \) is simply 122/200 or .61.

Obviously no corps had .61 deaths in a year. Many corps had zero deaths. Some corps had one death. Fewer corps had two deaths and so on.

Bortkiewicz’s data, and some calculations, are seen below.

<table>
<thead>
<tr>
<th>x = deaths / corps-year</th>
<th>Actual number of observations with that number of deaths in a corps-year</th>
<th>Percent of observations expected to have that number of deaths in a corps-year</th>
<th>Expected observations with x deaths, given 200 corps-years</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 deaths per corps-year</td>
<td>109</td>
<td>( P(X=0, \lambda=.61) = \left( e^{-.61} \cdot .61^0 \right) / 0! = .543 )</td>
<td>108.67</td>
</tr>
<tr>
<td>1 deaths per corps-year</td>
<td>65</td>
<td>( P(X=1, \lambda=.61) = \left( e^{-.61} \cdot .61^1 \right) / 1! = .334 )</td>
<td>66.28</td>
</tr>
<tr>
<td>2 deaths per corps-year</td>
<td>22</td>
<td>( P(X=2, \lambda=.61) = \left( e^{-.61} \cdot .61^2 \right) / 2! = .101 )</td>
<td>20.21</td>
</tr>
<tr>
<td>3 deaths per corps-year</td>
<td>3</td>
<td>( P(X=3, \lambda=.61) = \left( e^{-.61} \cdot .61^3 \right) / 3! = .020 )</td>
<td>4.11</td>
</tr>
<tr>
<td>4 deaths per corps-year</td>
<td>1</td>
<td>( P(X=4, \lambda=.61) = \left( e^{-.61} \cdot .61^4 \right) / 4! = .003 )</td>
<td>0.62</td>
</tr>
<tr>
<td>5 deaths per corps-year</td>
<td>0</td>
<td>( P(X=5, \lambda=.61) = \left( e^{-.61} \cdot .61^5 \right) / 5! = .000 )</td>
<td>0.07</td>
</tr>
<tr>
<td>6 deaths per corps-year</td>
<td>0</td>
<td>( P(X=6, \lambda=.61) = \left( e^{-.61} \cdot .61^6 \right) / 6! = .000 )</td>
<td>0.00</td>
</tr>
<tr>
<td>Total = 200 obs.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FIGURE 2

COMPARISON OF O.L.S. AND POISSON

We all have had the assumption of ordinary least squares (O.L.S.) drilled into our heads. Those assumptions are: 1) uncorrelated errors and 2) equal and constant variance of the error terms. Poisson regression is appropriate in a different set of circumstances. An easy way to see the differences between O.L.S. and Poisson regression is to compare the pictures in FIGURES 3 and 4. The little distributions, that are shown on the pictures, are the representation of many observations taken at the same value of X and then plotted.

You can see on the ordinary least squares regression graphic (FIGURE 3) little pictures of normal distributions at each observed level of X. These pictures represent the conditional frequency of an observed value of Y – conditional that X has that particular value. For the ordinary least squares regression, the histograms are all normal, smooth and have the same dispersion in Y (except when X=1 and the distribution is clipped because Y can not be negative). For ordinary least squares, no matter what the value is of X (almost), the conditional distribution of Y is normal and has the same spread.

Poisson regression is appropriate when the conditional distributions of Y are expected to be Poisson distributions (FIGURE 4). This often happens when you are trying to regress on count data (measles cases per 1000 elementary school students or paint imperfections in a square foot of automobile surface).

Count data will be, by its very nature, discrete as opposed to continuous (you cannot observe 1.5 cases of measles or 3.2 defects in a square foot of paint). So instead of a smooth looking curve, when we look at a Poisson distribution (FIGURE 4), we see a spiked and stepped histogram at each value of X. The second thing to notice about the Poisson regression picture is that the distribution is often skewed. The distribution of Y, for a small value of Y, is not symmetric. Another thing to notice about the Poisson regression is that the distribution of conditional Y values changes shape and spread as Y changes. A Poisson distribution becomes normal shaped, and wider, as the mean of the distribution gets larger.
In the appendix of this paper, there is a figure that shows how different distributions approach other distributions, under certain conditions. This sort of handout was very useful in the days before computers. In those “olden times” knowing that a theoretical distribution had approximated another distribution, with a computationally simpler formula, was very useful.

![Figure 3](image-url)

**FIGURE 3**

![Figure 4](image-url)

**FIGURE 4**

**WHEN IS POISSON REGRESSION APPROPRIATE**

Poisson regression is appropriate when the conditional value of Y is likely to have a Poisson distribution. The Poisson distribution is discrete, so Y should be discrete. The SAS training course instructs that the marginal mean of Y, must be less than 10. It suggests that the marginal mean of Y be less than 5, and mentions, for best results, it should be near 1.

**DIFFICULTIES**

There are two common difficulties in Poisson regression and they are both caused by heterogeneity in the data. By heterogeneity in the data, we mean that the data was collected, unknowingly, on more than one group of people. It means that there is more than one process that is generating the data, though it doesn’t appear to be the case on a simple overview of the data.

One difficulty is called overdispersion. This is when the variance of the fitted model is larger than what is expected by the assumptions (the mean and the variance are equal) of the Poisson model. Overdispersion is typically caused by a Poisson regression that is missing an important independent variable or by data being collected in clusters (like collecting data inside family units).
In the example, that follows, we will originally model the number of phone calls into software help desk (imagine the help desk is similar to SAS tech support). In the first model we will create, we will only use one independent variable, the size of the client company and this model will be missing an important variable. This will be an under-specified model. The client companies are not homogeneous – some client companies have in-house user groups and some do not.

In our example, companies that have an internal user group, and therefore conduct training at lunch time and have a sharing environment for knowledge, have a lower rate of calls. When this fact is recognized, it can be modeled by adding a variable for in-house user group and the heterogeneity within groups is reduced. When this variable is not in the model, and every client with the same number of employees is considered to be one group, we have an (unrecognized) heterogeneous population and overdispersion.

The second problem, also caused by heterogeneity, is excess zeros. In this situation, the distribution has more zeros than would be expected from a Poisson distribution. Often this is caused by two processes creating the data set, one of the processes having an expected count of zero. An example of this might be a distribution of holes in fabric being woven by a loom. Assume, if the loom is “in adjustment” it produces almost zero defects. Assume as the loom goes out of adjustment, it produces more defects, and produces defects with a Poisson count of defects per square yard.

Now imagine looking at data from one morning (7Am to 12 noon) for a loom. Assume that the loom was in adjustment from seven o’clock until nine o’clock (producing zero defects during that time period) and then out of adjustment from nine o’clock until noon. When trying to model that morning of data as Poisson, one would see excess zeros. Again, the problem is heterogeneity. Data coming from two different data sources, or processes, was combined into one data set.

Both of these problems can be fixed and we will try and do so in the example included in this paper.

**BUSINESS SITUATION:**

Our business situation will be an attempt to model calls to a software support help desk. In our example the client is a company that provides software support, similar to what SAS technical support does for us, for a proprietary software product.

The company has many clients and clients have different numbers of employees. A survey was used to collect information about the clients. Employee count was not recorded exactly, but was recorded via the mechanism of checking a box in a series of pre-printed boxes. As a result employee count is grouped and not continuous (note values of EmpCnt as the macro is called on the next page).

The code that generates our data is shown to the right.

The number of calls increases as the number of employees increases and is expected to be discrete and to follow a Poisson distribution.
Unrecorded in the original survey was the fact that some companies have in-house user groups (like an in-house SUG). The meetings held and the environment created by the in-house user groups will reduce the calls to the help desk.

This passes parameters to the macro.

Some of the output from a PROC UNIVARIATE is shown below. Note that the marginal distribution has a mean MUCH GREATER than 10. Modeling this distribution with a Poisson regression is not good practice, but is a learning exercise. Readers can explore different situations by changing macro parameters.

The shape of the marginal, or overall, distribution is determined by the “processes” (in this case, the mix of the company types) that generated it.

In the code above, we can see that the process that generated the distribution to the right was comprised of companies with 1,000, 3,000, 5,000 and 10,000 employees. The marginal distribution is skewed to the left.

If the company mix had been 1,000, 7,000, 8,000, 9,000 and 10,000 employees the marginal distribution would look different. It would be skewed to the right.

Additionally, the distribution suggests that it might be multimodal (see arrows). This is caused by the fact that a number of different processes, each with a different mean, were involved in the creation of the data.

This histogram has a “Poisson look” but the variance is likely be greater than the mean. This is because of “heterogeneity in the subjects”.

Some of the output from a PROC UNIVARIATE is shown below. Note that the marginal distribution has a mean MUCH GREATER than 10. Modeling this distribution with a Poisson regression is not good practice, but is a learning exercise. Readers can explore different situations by changing macro parameters.

The shape of the marginal, or overall, distribution is determined by the “processes” (in this case, the mix of the company types) that generated it.

In the code above, we can see that the process that generated the distribution to the right was comprised of companies with 1,000, 3,000, 5,000 and 10,000 employees. The marginal distribution is skewed to the left.

If the company mix had been 1,000, 7,000, 8,000, 9,000 and 10,000 employees the marginal distribution would look different. It would be skewed to the right.

Additionally, the distribution suggests that it might be multimodal (see arrows). This is caused by the fact that a number of different processes, each with a different mean, were involved in the creation of the data.

This histogram has a “Poisson look” but the variance is likely be greater than the mean. This is because of “heterogeneity in the subjects”.

Some of the output from a PROC UNIVARIATE is shown below. Note that the marginal distribution has a mean MUCH GREATER than 10. Modeling this distribution with a Poisson regression is not good practice, but is a learning exercise. Readers can explore different situations by changing macro parameters.
To illustrate how the Poisson distribution approaches normality as the mean gets large, FIGURE 7 shows
the marginal distribution, and some statistics, for the CoType= Avg emps=10000.  Also note that the mean is
49.9 and the variance is 68.12.  This data shows overdispersion and is the mix of the two types of clients.

<table>
<thead>
<tr>
<th>CoType=AvgWUserGroup emps=10000</th>
<th>Distribution for CoType=Avg emps=10000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moments</td>
<td></td>
</tr>
<tr>
<td>N 50</td>
<td>Sum Weights 50</td>
</tr>
<tr>
<td>Mean 49.96</td>
<td>Sum Obs. 2498</td>
</tr>
<tr>
<td>Std Dev. 8.25</td>
<td>Variance 68.12</td>
</tr>
<tr>
<td>Skew. 0.65</td>
<td>Kurtosis 0.415</td>
</tr>
<tr>
<td>UCSS1 28138</td>
<td>Corrected SS 3337.92</td>
</tr>
<tr>
<td>Coeff 16.52</td>
<td>Std Error Mean 1.167</td>
</tr>
</tbody>
</table>

Tests for Normality
Test --Statistic---       -----p Value-----
Shapiro-Wilk       W 0.96       Pr < W   0.1788
Kolmogorov-Smirnov D 0.10       Pr > D  >0.1500
Cramer-von Mises W-Sq 0.09  Pr > W-Sq 0.13
Anderson-Darling A-Sq 0.52  Pr > A-Sq 0.17

GENERALIZED LINEAR MODELS AND PROC GENMOD
The generalized linear model is an extension of the general linear model.  The \textit{generalized} model adds, to
the general linear model, the idea that Y is related to the x values through a \textit{link function}.  We will define the
link function as a transform, done to the Y variable, to correct “problems” caused by the conditional
distributions of Y not meeting model assumptions.  There are many link functions and data with many types
of conditional distributions can be modeled using the different link functions. PROC GENMOD can be used
to model several different conditional distributions, but generally the conditional distributions must be in the
exponential family.  This allows PROC GENMOD to use just one algorithm.

The link function for Poison data is the log transform. Accordingly, the model for a Poisson regression is:

For count data
\[
\text{Log(count)} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3
\]

For Rate Data
\[
\text{Log(}\mu/ T\text{)} = \text{Log(}\text{Expected # of events / Index of exposure e.g. hours, sq. miles, people at risk )}
\]
\[
= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3
\]
so
\[
\text{Log(}\mu\text{-Log(T)} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3;
\]
\[
\text{Log(}\mu\text{)} = \text{Log(T)} + \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3;
\]

The Log(T) is called the offset variable and is the log of the measure of exposure.  So if all the data is
collected with the same measure of exposure, you might as well model “counts observed” as model the rate
(“count per square mile” etc, or “count per 1000 exposed subjects”)
SUGGESTED HIGH LEVEL FLOWCHART:

The process of building a Poisson regression model is not straightforward. Building a Poisson Regression model is a multi-step, multi-decision process. At each step of the process, results must be examined and a decision made as to how to proceed. Sometimes the decision might be to abandon the use of a Poisson regression and to use another method.

Below is a suggested flowchart for the process of building a Poisson regression.

---

FIGURE 8

---
Every project should start with Exploratory Data Analysis (E.D.A.). Above, please see that the marginal distribution is skewed and the plot of calls versus the number of employees seems as if it could be linear. The scatter plot uses the new SAS SGPLOT (statistical graphics plot). In this example, while employees at a company is close to being a continuous variable, it is recorded by checking one of only a few pre-printed boxes on a form and appears discrete. If employees had been recorded as actual numbers of employees, employees could be “bucketed” to make for easier to read E.D.A. plots. The code for the plot is:

```sas
PROC SQL;
CREATE TABLE LogMeanByEmps AS
  SELECT emps AS Mean_Emps,
         log(mean(calls)) AS Log_Mean_Calls
  FROM CallsPerHour
  GROUP BY emps;
RUN;

PROC SGLOT DATA=LogMeanByEmps NOAUTOLEGEND;
SCATTER Y=Log_Mean_Calls X=Mean_Emps;
PBSPLINE Y=Log_Mean_Calls X=Mean_Emps / NOMARKERS SMOOTH=100 NKNOTS=10;
XAXIS LABEL= "Mean (bucketed) of emps";
YAXIS LABEL= "Log of Mean of calls";
RUN;
```

Some conditional distributions (plots of Y for each level of X were also examined. Some hinted at skewness and others did not. This is not surprising. We know that, as the average gets larger, skewness disappears and the data in our example had some very large conditional means.

Poisson regression should not be used on data with a marginal mean as large as we see in this data. The example proceeded in this manner to show results of violating assumptions.

The flowchart (FIGURE 8) indicates that we should try a Gamma or Lognormal Regression on this data. These techniques are not discussed in this paper. A planned paper, on these techniques, will compare results from those techniques to the results shown in this paper. It is, sometimes, useful to see what happens when assumptions are violated. Additionally, the reader can cut-and-paste the code from this paper into his/her copy of SAS and create a data set more in line with the assumptions of Poisson regression.
The SAS code to run the first model is shown below. This model does not have a variable to model “in-house user group”. This mistake will cause overdispersion. The values of dist and link make this a Poisson regression.

```
proc genmod data=CallsPerHour;
    model calls=emps /dist=poi link=log type3;
    title "Poisson Regression calls=emps ";
    run;
```

Important output is:

This tells us that we asked for a Poisson model. All observations that were read were used in the model. We have no missing data problems.

The first thing to check is the red number. It is a measure of overdispersion and we would like this number to be close to 1. If the number is much greater than one, we have overdispersion. As said before, overdispersion is where the fitted variance is larger than the mean.

Overdispersion is typically caused by an under-specified model (here we know we are missing the variable company-type) or by “clustering” in the data. Clustering in the data is collecting data within groups.

The first response of the modeler, to overdispersion, is to look for more variables that can be used to predict Y (either as variables added to the model or as a way of understanding clustering). In our example, we know that there is a variable company-type that can be used to predict the number of calls.

In our example, the existence of in-house user groups was discovered and added to the data. The model maker then added this variable to the model. The SAS code to run the augmented model is:
Proc Genmod data=CallsPerHour;
   class cotype (param=ref ref="Avg");
   model calls=emps cotype cotype*emps
      /dist=poi link=log type3;
   title "Poisson Regression calls=emps cotype cotype*emps ";
   estimate "average com Vs. Company With User Group" CoType 1;
   estimate "Effect of 1000 employees" emps 1000;
   estimate "Effect of 1 employee" emps 1;
run;

Important output is:

This tells us that we asked for a Poisson model.

All observations read were used. We have no missing data problems.

The base case is a company of Average type.

The overdispersion has dropped considerably but still exists. This will cause the S.E. to be underestimated and inflate the type I error.

The note is oddly placed, but this tells up that the algorithm converged.

The output below, from the model above, shows several things. Emps and CoType are significant but the interaction is not. Below, I will re-run the model without the interaction.

Poisson Regression calls=emps cotype cotype*emps

The GENMOD Procedure

Analysis Of Maximum Likelihood Parameter Estimates

Parameter  DF  Estimate  Error  Limits  Chi-Square  Pr > ChiSq
Intercept  1   2.5037   0.0265  2.4519  2.5556  8956.58  <.0001
emps      1   0.0002   0.0000  0.0002  0.0002  3239.01  <.0001
CoType     1  -0.4759   0.0427 -0.5596 -0.3922  124.17  <.0001
emps*CoType 1  0.0000   0.0000  0.0000  0.0000   0.02  0.8916
Scale      0   1.0000   0.0000  1.0000  1.0000

FIGURE 13
The new code is:

```sql
proc genmod data=CallsPerHour;
  class cotype (param=ref ref="Avg");
  model calls=emps cotype /dist=poi link=log type3;
  title "Poisson Regression calls=emps cotype  cotype*emps  ";
  estimate "average com Vs. Company With User Group" CoType 1;
  estimate "Effect of 1000 employees" emps 1000;
  estimate "Effect of 1 employee"     emps 1;
run;
```

The box to the right contains the results of the run. As you can see, the indicator of dispersion did not change much. There was a slight improvement in the AIC, AICC and BIC.

Overdispersion exists and we will, in the next section, use a Negative Binomial model to analyze the data.

Before doing that, we will do some more interpretation of the results of the Poisson model.

<table>
<thead>
<tr>
<th>Parameter Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
</tr>
<tr>
<td>Prm1</td>
</tr>
<tr>
<td>Prm2</td>
</tr>
<tr>
<td>Prm3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Criteria For Assessing Goodness Of Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criterion</td>
</tr>
<tr>
<td>Deviance</td>
</tr>
<tr>
<td>Scaled Deviance</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
</tr>
<tr>
<td>Log Likelihood</td>
</tr>
<tr>
<td>Full Log Likelihood</td>
</tr>
<tr>
<td>AIC (smaller is better)</td>
</tr>
<tr>
<td>AICC (smaller is better)</td>
</tr>
<tr>
<td>BIC (smaller is better)</td>
</tr>
</tbody>
</table>

Algorithm converged.

FIGURE 14
Below we can see that emps and CoType are significant. The -0.4706 means that there is a .47 decrease in the log as we go from an "average company" to a "company with an in-house user group". The 0.6246 means that the companies with in-house user groups make 62% of the calls that a company without an in-house user group would make. The 1.2119 means that if a company adds a thousand employees, the number of calls made is predicted to increase by 21.19%.

Poisson Regression calls=emps cotype cotype*emps

The GENMOD Procedure

Analysis Of Maximum Likelihood Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>95% Confidence Limits</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>2.5017</td>
<td>0.0220</td>
<td>2.4587 - 2.5447</td>
<td>12986.1</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>emps</td>
<td>1</td>
<td>0.0002</td>
<td>0.0000</td>
<td>0.0000 - 0.0002</td>
<td>5275.22</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>CoType AvgWUserGroup</td>
<td>1</td>
<td>-0.4706</td>
<td>0.0185</td>
<td>-0.5069 - -0.4344</td>
<td>646.52</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000 - 1.0000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis

Chi-Square Source DF  Square  Pr > ChiSq
Emps 1 5535.73 <.0001
CoType 1 864.53 <.0001

Contrast Estimate Results

Label Estimate Confidence Limits L'Beta Estimate Error Alpha
average com Vs. Company With User Group 0.6246 0.6024 0.6477 -0.4706 0.0185 0.05
Effect of 1000 employees 1.2119 1.2057 1.2182 0.1922 0.0026 0.05
Effect of 1 employee 1.0002 1.0002 1.0002 0.0002 0.0000 0.05

Contrast Estimate Results

Label L'Beta Chi-Square Pr > ChiSq
Average com Vs. Company With User Group -0.5069 -0.4344 646.52 <.0001
Effect of 1000 employees 0.1870 0.1974 245.64 <.0001
Effect of 1 employee 0.0002 0.0002 245.64 <.0001

FIGURE 15

Estimates of the betas are shown below.

Poisson Regression calls=emps cotype cotype*emps

The GENMOD Procedure

Analysis Of Maximum Likelihood Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>95% Confidence Limits</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>2.4189</td>
<td>0.0330</td>
<td>2.3543 - 2.4835</td>
<td>5388.80</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>emps</td>
<td>1</td>
<td>0.0002</td>
<td>0.0000</td>
<td>0.0000 - 0.0002</td>
<td>1956.37</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>CoType AvgWUserGroup</td>
<td>1</td>
<td>-0.4715</td>
<td>0.0301</td>
<td>-0.5305 - -0.4126</td>
<td>245.64</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Dispersion</td>
<td>1</td>
<td>0.0462</td>
<td>0.0075</td>
<td>0.0336 - 0.0636</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: The negative binomial dispersion parameter was estimated by maximum likelihood.

LR Statistics For Type 3 Analysis

Chi-Square Source DF  Square  Pr > ChiSq
Emps 1 702.43 <.0001
CoType 1 175.99 <.0001

Contrast Estimate Results

Label Estimate Confidence Limits L'Beta Estimate Error Alpha
average com Vs. Company With User Group 0.6246 0.5883 0.6620 -0.4715 0.0301 0.05
Effect of 1000 employees 1.2294 1.2182 1.2407 0.2065 0.0047 0.05
Effect of 1 employee 1.0002 1.0002 1.0002 0.0002 0.0000 0.05

Contrast Estimate Results

Label L'Beta Chi-Square Pr > ChiSq
average com Vs. Company With User Group -0.5069 -0.4344 646.52 <.0001
Effect of 1000 employees 0.1974 0.2157 1956.4 <.0001
Effect of 1 employee 0.0002 0.0002 1956.4 <.0001

FIGURE 16

- 12 -
Clients will ask how well we did and will want the answers in real units and lot logs or ratios. Let’s do a few calculations to see. Immediately below is the result of a PROC MEANS on the raw data. Below we can see the raw averages for each crossing of company type and number of employees. We will compare this to predicted values.

<table>
<thead>
<tr>
<th>Analysis Variable : calls</th>
<th>N</th>
<th>CoType</th>
<th>emps</th>
<th>Obs</th>
<th>N</th>
<th>Minimum</th>
<th>Mean</th>
<th>Maximum</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg</td>
<td></td>
<td>Avg</td>
<td>1000</td>
<td>50</td>
<td>50</td>
<td>2.000000</td>
<td>8.220000</td>
<td>16.000000</td>
<td>411.00</td>
</tr>
<tr>
<td>3000</td>
<td></td>
<td>3000</td>
<td>50</td>
<td>50</td>
<td>16.000000</td>
<td>24.860000</td>
<td>39.000000</td>
<td>1243.00</td>
<td></td>
</tr>
<tr>
<td>5000</td>
<td></td>
<td>5000</td>
<td>50</td>
<td>50</td>
<td>28.000000</td>
<td>39.440000</td>
<td>55.000000</td>
<td>1972.00</td>
<td></td>
</tr>
<tr>
<td>10000</td>
<td></td>
<td>10000</td>
<td>50</td>
<td>50</td>
<td>63.000000</td>
<td>79.320000</td>
<td>98.000000</td>
<td>3966.00</td>
<td></td>
</tr>
<tr>
<td>AvgWUserGroup</td>
<td></td>
<td>AvgWUserGroup</td>
<td>1000</td>
<td>50</td>
<td>50</td>
<td>0</td>
<td>4.800000</td>
<td>10.000000</td>
<td>240.00</td>
</tr>
<tr>
<td>3000</td>
<td></td>
<td>AvgWUserGroup</td>
<td>50</td>
<td>50</td>
<td>9.000000</td>
<td>15.580000</td>
<td>32.000000</td>
<td>779.00</td>
<td></td>
</tr>
<tr>
<td>5000</td>
<td></td>
<td>AvgWUserGroup</td>
<td>50</td>
<td>50</td>
<td>14.000000</td>
<td>25.000000</td>
<td>34.000000</td>
<td>1250.00</td>
<td></td>
</tr>
<tr>
<td>10000</td>
<td></td>
<td>AvgWUserGroup</td>
<td>50</td>
<td>50</td>
<td>38.000000</td>
<td>49.460000</td>
<td>73.000000</td>
<td>2473.00</td>
<td></td>
</tr>
</tbody>
</table>

**FIGURE 17**

For a company with 5000 employees and an in-house user group, we expect to see

\[ \mu = e^{(b_0 + b_{emps} \times 5000 + b_{cotype} \times 1)} = e^{(2.5017 + 1 - 0.4706)} = 20.72 \text{ vs } 25 \]

For an average company with 10000 employees, we expect to see

\[ \mu = e^{(b_0 + b_{emps} \times 10000 + b_{cotype} \times 0)} = e^{(2.5017 + 2 + 0)} = 90.17 \text{ vs } 79 \]
There was evidence of overdispersion in the results above. We did not have any more new variables to bring into the model and did not have an indication that clustering might have been the cause of the problem. The next step is to run a negative binomial model. If a modeler cannot improve a Poisson model, and overdispersion exists, it is suggested that a Negative Binomial Model be attempted. The code is shown immediately below.

```plaintext
proc genmod data=CallsPerHour;
  class cotype (param=ref ref="Avg");
  model calls=emps cotype dist=nb link=log type3;
  title "Poisson Regression calls=emps cotype cotype*emps ";
  estimate "average com Vs. Company With User Group" CoType 1;
  estimate "Effect of 1000 employees" emps 1000;
  estimate "Effect of 1 employee" emps 1;
run;
```

Important output is:

This tells us that we used a Negative Binomial to model the data.

We had no missing value issues. All the data that was read, was used in the model.

The base case is the average company,

The overdispersion is gone. The measure of overdispersion is 1.1476.

The model converged.

The GenMod Procedure

Model Information

Data Set               WORK.CALLSPERHOUR
Distribution          Negative Binomial
Link Function                       Log
Dependent Variable                calls

Number of Observations Read         400
Number of Observations Used         400

Class Level Information

<table>
<thead>
<tr>
<th>Class</th>
<th>Value</th>
<th>Design Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>CoType</td>
<td>Avg</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>AvgWUserGroup</td>
<td>1</td>
</tr>
</tbody>
</table>

Parameter Information

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Effect</th>
<th>CoType</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prm1</td>
<td>Intercept</td>
<td></td>
</tr>
<tr>
<td>Prm2</td>
<td>emps</td>
<td></td>
</tr>
<tr>
<td>Prm3</td>
<td>CoType</td>
<td>AvgWUserGroup</td>
</tr>
</tbody>
</table>

Criteria For Assessing Goodness Of Fit

<table>
<thead>
<tr>
<th>Criterion</th>
<th>DF</th>
<th>Value</th>
<th>Value/DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviance</td>
<td>397</td>
<td>494.2319</td>
<td>1.2449</td>
</tr>
<tr>
<td>Scaled Deviance</td>
<td>397</td>
<td>494.2319</td>
<td>1.2449</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>397</td>
<td>455.6018</td>
<td>1.1476</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
<td>397</td>
<td>455.6018</td>
<td>1.1476</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>33110.3107</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full Log Likelihood</td>
<td>-1387.5418</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC (smaller is better)</td>
<td>2783.0836</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AICC (smaller is better)</td>
<td>2783.1849</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIC (smaller is better)</td>
<td>2799.0495</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Algorithm converged.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FIGURE 18
More output is shown below.

Poisson Regression calls=emps cotype cotype*emps

The GENMOD Procedure

Analysis Of Maximum Likelihood Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>Limits</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>2.4189</td>
<td>0.0330</td>
<td>2.3543</td>
<td>2.4835</td>
<td>5388.80</td>
</tr>
<tr>
<td>emps</td>
<td>1</td>
<td>0.0002</td>
<td>0.0000</td>
<td>0.0002</td>
<td>0.0002</td>
<td>1956.37</td>
</tr>
<tr>
<td>CoType AvgUserGroup</td>
<td>1</td>
<td>-0.4715</td>
<td>0.0301</td>
<td>-0.5305</td>
<td>-0.4126</td>
<td>245.64</td>
</tr>
<tr>
<td>Dispersion</td>
<td>1</td>
<td>0.0462</td>
<td>0.0075</td>
<td>0.0336</td>
<td>0.0636</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: The negative binomial dispersion parameter was estimated by maximum likelihood.

LR Statistics For Type 3 Analysis

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>emps</td>
<td>1</td>
<td>702.43</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>CoType</td>
<td>1</td>
<td>175.99</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Contrast Estimate Results

<table>
<thead>
<tr>
<th>Label</th>
<th>Mean Estimate</th>
<th>Mean Confidence Limits</th>
<th>L'Beta Estimate</th>
<th>Standard Error</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average com Vs. Company With User Group</td>
<td>0.6241</td>
<td>0.5883 - 0.6620</td>
<td>-0.4715</td>
<td>0.0301</td>
<td>0.05</td>
</tr>
<tr>
<td>Effect of 1000 employees</td>
<td>1.2294</td>
<td>1.2182 - 1.2407</td>
<td>0.2065</td>
<td>0.0047</td>
<td>0.05</td>
</tr>
<tr>
<td>Effect of 1 employee</td>
<td>1.0002</td>
<td>1.0002 - 1.0002</td>
<td>0.0002</td>
<td>0.0000</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Contrast Estimate Results

<table>
<thead>
<tr>
<th>Label</th>
<th>L'Beta</th>
<th>Confidence Limits</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average com Vs. Company With User Group</td>
<td>-0.5305</td>
<td>-0.4126 - 0.6426</td>
<td>245.64</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Effect of 1000 employees</td>
<td>0.1974</td>
<td>0.1857 - 0.2087</td>
<td>1956.4</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Effect of 1 employee</td>
<td>0.0002</td>
<td>0.0000 - 0.0002</td>
<td>1956.4</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

FIGURE 19

The N.B.D. model made the predictions below.

\[ \mu = e^{(b_0 + b_{\text{emps}} \times 5000 + b_{\text{cotype}} \times 1)} = e^{(2.4189 + 1 + -0.4715)} = 19.06 \text{ vs } 25 \]

For an average company with 10000 employees, we expect to see

\[ \mu = e^{(b_0 + b_{\text{emps}} \times 10000 + b_{\text{cotype}} \times 0)} = e^{(2.4189 + 2 + 0)} = 83.00 \text{ vs } 79 \]

CONCLUSION

These techniques are appropriate when the conditional distribution can be assumed to be Poisson and the marginal mean is small (ideally close to one, hopefully less than five definitely less than ten).

While the results of the modeling did not predict well, the rule of thumb given above was grossly violated. This was done intentionally to show that it is not wise to violate the rule of thumb.

Poisson regression is a useful tool in and is implemented in PROC GENMOD.

REFERENCES

“SAS training Introduction to Poisson regression with Proc GenMod”

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Russell Lavery Contractor Russ.Lavery@verizon.net

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.
Statistics and Analysis

Poisson Approx to Binomial

If \( N \geq 100 \) trials and \( NP \leq 20 \) and \( Np \leq 0.05 \) you can use the Poisson instead of the binomial. Set \( \lambda = np \). This allows you to not calculate all the combinations in the binomial formula. The Poisson is a special case of the binomial. Start with binomial \( p.m.f \) \( P(x; n, p) \) let \( n \rightarrow \infty \) and \( p \rightarrow 0 \) keeping \( NP \) fixed at the value \( \lambda \) (must be GT 0) then \( b(n; p) \rightarrow p(x; \lambda) \)

\[
P(X=x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}
\]

Normal approx. to Binomial

When using this you will be predicting successes. \( P \) = prob. of success. \( N \) = number of trials.

Rule of thumb is: \( Np \geq 5 \) and \( Nq \geq 5 \).

Don’t forget the continuity correction! add \(-0.5\) to the \( x \) values

Mean = \( \mu = \lambda \) \( \sigma = \sqrt{\lambda} \)

Normal Distribution

We like this distribution because it has just ONE table and the table is easy to use.

Binomial Approx To Hypergeometric

If probability of success does not change much (and this is left to your judgment) as you select items you can use the binomial.

Rule of thumb for use is: \( n/N \leq 0.05 \) and \( p \) not too near 0 or 1.

Hypergeometric

From a small \( N \), with a certain number of (successes) good parts and (failures) bad parts we can say \( N=S+F \). Since the population is finite, the chance of a success changes as parts are picked from the “shipping box”

\( p=S/N \), \( q=F/N \)

Exponential

\( X \) is a continuous random variable and often is the time to service a customer at a bank. If we expect to service one customer in two minutes we expect to service 1/2 customers in one minute - and \( \lambda = .5 \)

\( E(x) = 1/\lambda \)

\( V = 1/(\lambda^2) \)

\( \lambda = \) Average \# of events per unit measure (time or area)

Multinomial

There are more than two “kinds” of outcomes. Here we will assume 3. (Imagine computer chips being made and having the following probabilities: Chance of being scrap=10%. Chance of being commercial quality=70%. Chance of being Military grade=20%.)

In a batch of 100, what is the chance of 30 scrap, 60 commercial and 10 government?