The GENMOD Toolkit: What do you do when things aren't Normal?

The GENMOD Toolkit: What do you do when things aren’t Normal?

Danny Modlin – SAS Cary, NC
Senior Analytical Training Consultant

General Linear Models

$$y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + \varepsilon_i$$

$$\varepsilon_i \sim i.i.d.\ N(0, \sigma^2)$$

$$\Rightarrow \quad E(y_i) = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki}$$

$$\mathrm{var}(y_i) = \sigma^2$$

Generalized Linear Models

$$g(E(y_i)) = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} = X\beta$$

- The distribution of the observations can come from the exponential family of distributions.
- The variance of the response variable can be expressed as a function of its mean.
- $X\beta$ is fit to a function of $E(y)$ (called a link function), which is suggested by the distribution of the observations: $g(E(y)) = g(\mu) = X\beta$.

Link function

Examples of Generalized Linear Models

Model	Response	Distribution	Mean	Variance	Canonical Link
Linear Regression	Continuous	Normal	μ	σ^2	identity μ
Logistic regression	Dichotomous	Binomial	π	$\pi(1-\pi)/n$	logit $\log[\pi/(1-\pi)]$
Poisson Regression	Count	Poisson	λ	λ	log $\log(\lambda)$
Gamma Regression	Continuous	Gamma	μ	μ^2/ν	*inverse $1/\mu$

* Models often use the LOG link in practice.

The GENMOD Procedure

General form of the GENMOD procedure:

```
PROC GENMOD options PLOTS=requests;
    CLASS variables;
    MODEL response=effects / options;
    ESTIMATE 'label' effect values / options;
RUN;
```


Available Distributions

Click to edit subtitle

Poisson

Geometric

Binomial

Gamma

Multinomial

Neg Binomial

Normal

Inv Gaussian

Zero-Inflated Neg Bin*

Zero-Inflated Poisson*

Available Link Functions

Inverse

Logit

Identity

Power

Probit

Complementary Log-Log

Cumulative Logit

Log

*Can define your own link functions using programming statements within proc

In [4]:

```
/*Patients in each of two centers are randomly assigned to groups receiving the active
  During treatment, respiratory status, represented by the variable outcome (coded her
  is determined for each of four visits. The variables center, treatment, sex, and bas
  are classification variables with two levels. The variable age (age at time of entry
*/

data resp;
   input center id treatment $ sex $ age baseline visit1-visit4;
   datalines;
1  1 P M 46 0 0 0 0 0
1  2 P M 28 0 0 0 0 0
1  3 A M 23 1 1 1 1 1
1  4 P M 44 1 1 1 1 0
1  5 P F 13 1 1 1 1 1
1  6 A M 34 0 0 0 0 0
1  7 P M 43 0 1 0 1 1
1  8 A M 28 0 0 0 0 0
1  9 A M 31 1 1 1 1 1
1 10 P M 37 1 0 1 1 0
1 11 A M 30 1 1 1 1 1
1 12 A M 14 0 1 1 1 0
1 13 P M 23 1 1 0 0 0
1 14 P M 30 0 0 0 0 0
1 15 P M 20 1 1 1 1 1
1 16 A M 22 0 0 0 0 1
1 17 P M 25 0 0 0 0 0
1 18 A F 47 0 0 1 1 1
1 19 P F 31 0 0 0 0 0
1 20 A M 20 1 1 0 1 0
1 21 A M 26 0 1 0 1 0
1 22 A M 46 1 1 1 1 1
1 23 A M 32 1 1 1 1 1
1 24 A M 48 0 1 0 0 0
1 25 P F 35 0 0 0 0 0
1 26 A M 26 0 0 0 0 0
```

```
1 27 P M 23 1 1 0 1 1
1 28 P F 36 0 1 1 0 0
1 29 P M 19 0 1 1 0 0
1 30 A M 28 0 0 0 0 0
1 31 P M 37 0 0 0 0 0
1 32 A M 23 0 1 1 1 1
1 33 A M 30 1 1 1 1 0
1 34 P M 15 0 0 1 1 0
1 35 A M 26 0 0 0 1 0
1 36 P F 45 0 0 0 0 0
1 37 A M 31 0 0 1 0 0
1 38 A M 50 0 0 0 0 0
1 39 P M 28 0 0 0 0 0
1 40 P M 26 0 0 0 0 0
1 41 P M 14 0 0 0 0 1
1 42 A M 31 0 0 1 0 0
1 43 P M 13 1 1 1 1 1
1 44 P M 27 0 0 0 0 0
1 45 P M 26 0 1 0 1 1
1 46 P M 49 0 0 0 0 0
1 47 P M 63 0 0 0 0 0
1 48 A M 57 1 1 1 1 1
1 49 P M 27 1 1 1 1 1
1 50 A M 22 0 0 1 1 1
1 51 A M 15 0 0 1 1 1
1 52 P M 43 0 0 0 1 0
1 53 A F 32 0 0 0 1 0
1 54 A M 11 1 1 1 1 0
1 55 P M 24 1 1 1 1 1
1 56 A M 25 0 1 1 0 1
2  1 P F 39 0 0 0 0 0
2  2 A M 25 0 0 1 1 1
2  3 A M 58 1 1 1 1 1
2  4 P F 51 1 1 0 1 1
2  5 P F 32 1 0 0 1 1
2  6 P M 45 1 1 0 0 0
2  7 P F 44 1 1 1 1 1
2  8 P F 48 0 0 0 0 0
2  9 A M 26 0 1 1 1 1
2 10 A M 14 0 1 1 1 1
2 11 P F 48 0 0 0 0 0
2 12 A M 13 1 1 1 1 1
2 13 P M 20 0 1 1 1 1
2 14 A M 37 1 1 0 0 1
2 15 A M 25 1 1 1 1 1
2 16 A M 20 0 0 0 0 0
2 17 P F 58 0 1 0 0 0
2 18 P M 38 1 1 0 0 0
2 19 A M 55 1 1 1 1 1
2 20 A M 24 1 1 1 1 1
2 21 P F 36 1 1 0 0 1
2 22 P M 36 0 1 1 1 1
2 23 A F 60 1 1 1 1 1
2 24 P M 15 1 0 0 1 1
2 25 A M 25 1 1 1 1 0
2 26 A M 35 1 1 1 1 1
2 27 A M 19 1 1 0 1 1
```

```
2 28 P F 31 1 1 1 1 1
2 29 A M 21 1 1 1 1 1
2 30 A F 37 0 1 1 1 1
2 31 P M 52 0 1 1 1 1
2 32 A M 55 0 0 1 1 0
2 33 P M 19 1 0 0 1 1
2 34 P M 20 1 0 1 1 1
2 35 P M 42 1 0 0 0 0
2 36 A M 41 1 1 1 1 1
2 37 A M 52 0 0 0 0 0
2 38 P F 47 0 1 1 0 1
2 39 P M 11 1 1 1 1 1
2 40 P M 14 0 0 0 1 0
2 41 P M 15 1 1 1 1 1
2 42 P M 66 1 1 1 1 1
2 43 A M 34 0 1 1 0 1
2 44 P M 43 0 0 0 0 0
2 45 P M 33 1 1 1 0 1
2 46 P M 48 1 1 0 0 0
2 47 A M 20 0 1 1 1 1
2 48 P F 39 1 0 1 0 0
2 49 A M 28 0 1 0 0 0
2 50 P F 38 0 0 0 0 0
2 51 A M 43 1 1 1 1 0
2 52 A F 39 0 1 1 1 1
2 53 A M 68 0 1 1 1 1
2 54 A F 63 1 1 1 1 1
2 55 A M 31 1 1 1 1 1
;

data resp;
   set resp;
   visit=1;  outcome=visit1;  output;
   visit=2;  outcome=visit2;  output;
   visit=3;  outcome=visit3;  output;
   visit=4;  outcome=visit4;  output;
run;

proc print data=resp (obs=20);
run;

proc genmod data=resp;
   class id treatment(ref="P") center(ref="1") sex(ref="M")
      baseline(ref="0");
   model outcome(event='1')=treatment center sex age baseline / dist=bin;
   repeated subject=id(center) / corr=unstr corrw;
run;
```

The SAS System

Obs	center	id	treatment	sex	age	baseline	visit1	visit2	visit3	visit4	visit	outcome
1	1	1	P	M	46	0	0	0	0	0	1	0
2	1	1	P	M	46	0	0	0	0	0	2	0
3	1	1	P	M	46	0	0	0	0	0	3	0
4	1	1	P	M	46	0	0	0	0	0	4	0
5	1	2	P	M	28	0	0	0	0	0	1	0
6	1	2	P	M	28	0	0	0	0	0	2	0
7	1	2	P	M	28	0	0	0	0	0	3	0
8	1	2	P	M	28	0	0	0	0	0	4	0
9	1	3	A	M	23	1	1	1	1	1	1	1
10	1	3	A	M	23	1	1	1	1	1	2	1
11	1	3	A	M	23	1	1	1	1	1	3	1
12	1	3	A	M	23	1	1	1	1	1	4	1
13	1	4	P	M	44	1	1	1	1	0	1	1
14	1	4	P	M	44	1	1	1	1	0	2	1
15	1	4	P	M	44	1	1	1	1	0	3	1
16	1	4	P	M	44	1	1	1	1	0	4	0
17	1	5	P	F	13	1	1	1	1	1	1	1
18	1	5	P	F	13	1	1	1	1	1	2	1
19	1	5	P	F	13	1	1	1	1	1	3	1
20	1	5	P	F	13	1	1	1	1	1	4	1

The SAS System

The GENMOD Procedure

Model Information	
Data Set	WORK.RESP
Distribution	Binomial
Link Function	Logit
Dependent Variable	outcome

Number of Observations Read	444
Number of Observations Used	444
Number of Events	248

Number of Trials	444

Class Level Information		
Class	**Levels**	**Values**
id	56	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56
treatment	2	A P
center	2	2 1
sex	2	F M
baseline	2	1 0

Response Profile		
Ordered Value	**outcome**	**Total Frequency**
1	1	248
2	0	196

PROC GENMOD is modeling the probability that outcome='1'. One way to change this to model the probability that outcome='0' is to specify the DESCENDING option in the PROC statement.

Parameter Information					
Parameter	**Effect**	**treatment**	**center**	**sex**	**baseline**
Prm1	Intercept				
Prm2	treatment	A			
Prm3	treatment	P			
Prm4	center		2		
Prm5	center		1		
Prm6	sex			F	
Prm7	sex			M	
Prm8	age				
Prm9	baseline				1
Prm10	baseline				0

Algorithm converged.

GEE Model Information	
Correlation Structure	Unstructured
Subject Effect	id(center) (111 levels)

GEE Model Information	
Number of Clusters	111
Correlation Matrix Dimension	4
Maximum Cluster Size	4
Minimum Cluster Size	4

Algorithm converged.

Working Correlation Matrix				
	Col1	Col2	Col3	Col4
Row1	1.0000	0.3351	0.2140	0.2953
Row2	0.3351	1.0000	0.4429	0.3581
Row3	0.2140	0.4429	1.0000	0.3964
Row4	0.2953	0.3581	0.3964	1.0000

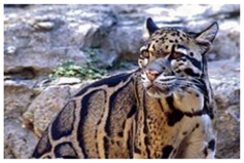
GEE Fit Criteria	
QIC	512.3416
QICu	499.6081

Analysis Of GEE Parameter Estimates Empirical Standard Error Estimates							
Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept		-0.8882	0.4568	-1.7835	0.0071	-1.94	0.0519
treatment	A	1.2442	0.3455	0.5669	1.9214	3.60	0.0003
treatment	P	0.0000	0.0000	0.0000	0.0000	.	.
center	2	0.6558	0.3512	-0.0326	1.3442	1.87	0.0619
center	1	0.0000	0.0000	0.0000	0.0000	.	.
sex	F	0.1128	0.4408	-0.7512	0.9768	0.26	0.7981
sex	M	0.0000	0.0000	0.0000	0.0000	.	.
age		-0.0175	0.0129	-0.0427	0.0077	-1.36	0.1728
baseline	1	1.8981	0.3441	1.2237	2.5725	5.52	<.0001
baseline	0	0.0000	0.0000	0.0000	0.0000	.	.

Poisson Regression

Poisson regression has the following characteristics:

- is one type of generalized linear model
- assumes that the response variable follows a Poisson distribution that is conditional on the values of the predictor variables
- can be used to model the number of occurrences of an event of interest or the rate of occurrence of an event of interest as a function of some predictor variables
- is most appropriate for counts of rare events



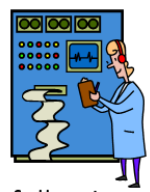
Overdispersion

- Poisson regression models assume that the variance is equal to the mean.
- Count data often exhibit variability that exceeds the mean.
- Overdispersion leads to underestimates of the standard errors of parameter estimates.
- Overdispersion results in overestimates of the test statistic and liberal *p*-values.

Causes of Overdispersion

- subject heterogeneity due to an under-specified model
- outliers in the data
- positive correlation between the responses in clustered data

Correcting for Overdispersion



- Make sure that you do not have erroneous data.
- Recheck your model to include all important variables. Re-specify your model if necessary.

After these assessments are completed, you can do one of the following:

- Use the negative binomial distribution to model the overdispersion (DIST=NEGBIN option in the MODEL statement in PROC GENMOD).

  or

- Apply a multiplicative adjustment factor to adjust the standard errors accordingly (PSCALE or DSCALE option in the MODEL statement in PROC GENMOD).

Negative Binomial Distribution

The negative binomial distribution has the following attributes:

- It is the distribution for count data that permits the variance to exceed the mean.
- This distribution enables the model to have greater flexibility in modeling the relationship between the mean and the variance of the response variable than the Poisson model has.

Negative Binomial Model

Response Variable	Distribution	Link Function	Variance Function
Count	Negative Binomial	Natural Log	$\mu + k\mu^2$

Dispersion Parameter *k*

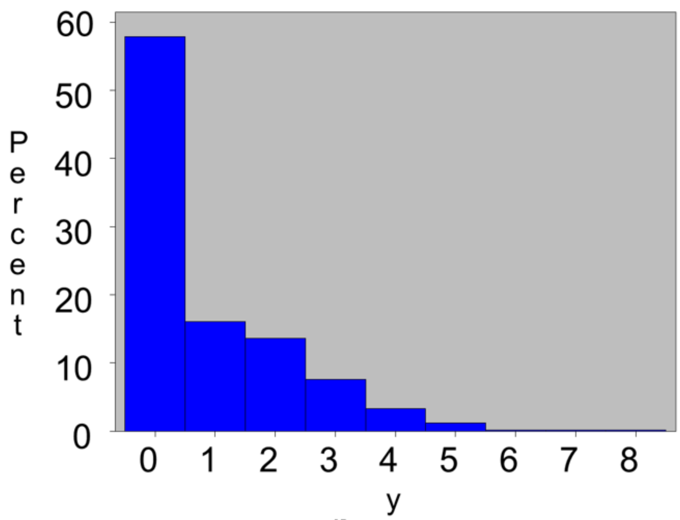
- The dispersion parameter *k* is not allowed to vary over observations.
- The limiting case when the parameter *k* is equal to 0 corresponds to a Poisson regression model.
- When the parameter is greater than 0, overdispersion is evident and the standard errors increase. The fitted values are similar, but the larger standard errors reflect the overdispersion that is uncaptured with the Poisson model.

Zero Inflation

Zero-Inflated Poisson Models

- In some settings, the incidence of zero counts will be much greater than expected for the Poisson distribution.
- Poisson regression models will exhibit overdispersion when they are fit to data with an excess number of zeros.
- Zero-inflated Poisson (ZIP) models might be a better fit to the data.

Count Data with Many Zeros



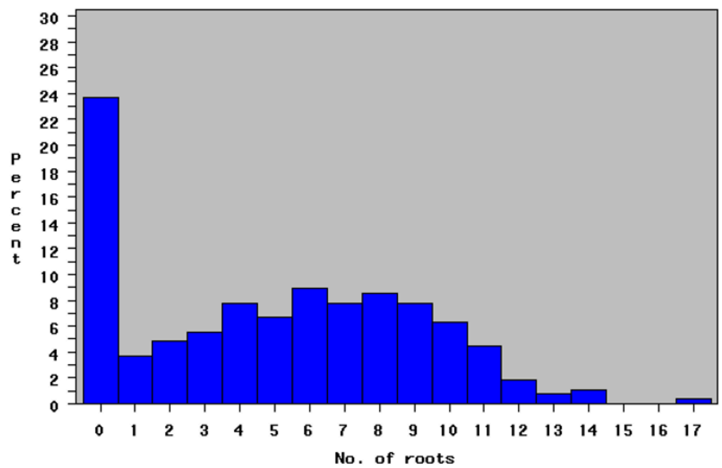
ZIP Models

- The population that can be modeled with the zero-inflated Poisson distribution is considered to consist of two types of responses.
- The first type gives Poisson distributed counts, which can produce the zero outcome or some other positive outcome.
- The second type always gives a zero count.
- Therefore, the relevant distribution is a mixture of a Poisson distribution and a distribution that is constant at zero.

A Biological Example

photoperiod (hour)	concentration (μM)			
	2.2	4.4	8.8	17.6
8	Number of roots	Number of roots	Number of roots	Number of roots
16	Number of roots	Number of roots	Number of roots	Number of roots

Distribution



The GENMOD Procedure

Inclusion of the ZEROMODEL statement:

PROC GENMOD *options* PLOTS=*requests*;
    CLASS *variables*;
    MODEL *response=effects* / *options*;
    ZEROMODEL *response=effects* / *options*;
RUN;

*Default LINK in ZEROMODEL is LOGIT
*Same effects can appear in both MODEL and ZEROMODEL

In [2]:

```
/*We have data on 250 groups that went to a park.
  Each group was questioned about how many fish they caught (count),
  how many children were in the group (child),
  how many people were in the group (persons),
  and whether or not they brought a camper to the park (camper).
*/

data WORK.FISH(label='Written by SAS  ');
  infile datalines dsd truncover;
  input nofish:32. livebait:32. camper:32. persons:32. child:32. xb:32. zg:32. count:3
datalines4;
1,0,0,1,0,-0.896314561,3.0504047871,0
```

```
0,1,1,1,0,-0.55834496,1.7461489439,0
0,1,0,1,0,-0.401731014,0.2799388766,0
0,1,1,2,1,-0.956298113,-0.601525664,0
0,1,0,1,0,0.4368909597,0.5277091265,1
0,1,1,4,2,1.3944854736,-0.70753479,0
0,1,0,3,1,0.1847167462,-3.398022175,0
0,1,0,4,3,2.3291065693,-5.450901508,0
1,0,1,3,2,0.188386485,-1.527417779,0
0,1,1,1,0,0.2876899242,1.3938905001,1
0,1,0,4,1,1.9909527302,-1.933189988,0
0,1,1,3,2,1.3178931475,-2.471574545,0
1,0,0,3,0,0.2980416715,1.5912652016,1
0,1,0,3,0,1.2908734083,0.8295348883,2
0,1,1,1,0,-0.06088984,2.8205792904,0
1,1,1,1,0,0.3700492084,2.1583449841,1
0,1,0,4,1,1.9790934324,-3.069952726,0
1,1,1,3,2,0.7153370976,-1.952804923,0
0,1,1,2,1,1.5160530806,-0.186567351,1
0,1,0,3,1,-0.034895968,-0.118922494,0
0,1,0,4,1,1.1782302856,0.0018565661,1
0,1,1,4,0,1.6421117783,1.892821312,5
0,1,1,2,1,0.5977273583,-0.294278145,0
0,1,1,2,0,1.1397230625,1.9317910671,3
0,1,1,3,0,3.5002975464,1.4512708187,30
0,1,1,2,0,-0.789978385,2.817448616,0
0,1,1,4,0,2.6623561382,1.6565625668,13
0,1,0,2,1,1.6061724424,-1.064542532,0
0,1,0,1,0,0.0185501799,0.0807982683,0
0,1,0,4,3,3.0345590115,-4.824044704,0
0,1,0,1,0,0.0510007553,0.9218238592,0
0,1,1,3,1,0.7452588081,-0.663867116,0
0,1,1,4,0,2.4531366825,3.5083677769,11
0,1,1,4,1,2.3527066708,0.1766097248,5
0,1,1,1,0,-1.108257532,0.7720884085,0
1,0,0,2,0,0.5154194236,1.6566575766,1
1,1,1,2,1,1.9827685356,-0.642237127,1
1,1,1,4,1,2.0668559074,1.244507432,7
1,1,1,3,1,0.0950117707,-2.268660784,0
0,1,1,4,1,5.3526740074,-1.472992539,14
0,1,1,1,0,-0.711813927,3.0204780102,0
0,1,1,4,0,3.4840462208,2.3556528091,32
1,0,0,3,2,2.4009063244,-3.086473703,0
1,0,1,4,0,0.3760284781,2.6760778427,1
1,1,0,4,2,1.0850564241,-2.654790401,0
0,1,1,1,0,-1.067198753,2.1330449581,0
0,1,1,2,1,0.3229970634,0.3034364879,0
0,1,0,1,0,0.5010663271,1.5531616211,1
0,1,1,2,1,2.0114884377,0.6270364523,5
0,1,0,2,1,0.9571473598,-2.05817461,0
0,1,1,2,1,1.1144340038,0.2847320735,1
0,1,1,2,1,-0.673830032,-0.708101988,0
0,1,1,4,0,3.1967058182,1.5154833794,22
1,0,0,2,0,-0.3863208,2.0795879364,0
```

```
0,1,1,3,0,2.7185115814,2.6293129921,15
1,0,1,1,0,-1.269290566,4.179599762,0
0,1,1,1,0,-1.087805986,2.1226129532,0
0,1,1,1,0,-0.984578311,1.351520896,0
1,1,1,4,1,1.8724677563,1.2611768246,5
1,1,1,1,0,1.547955513,1.6289883852,4
0,1,0,2,0,1.004094243,1.0836246014,2
0,1,1,2,1,-0.165202618,2.0952503681,0
0,1,0,2,1,1.471645236,-0.073470898,2
0,1,1,4,0,3.4893360138,2.5479974747,32
0,1,0,4,3,1.8857758045,-4.232519627,0
0,1,1,1,0,-2.272530079,1.6007531881,0
0,1,0,1,0,0.6138409972,1.1117559671,1
1,1,0,3,2,2.8789975643,-2.766067028,0
1,1,1,1,0,-0.944848835,2.0116078854,0
0,1,1,2,1,0.8200352192,-1.285437942,0
0,1,1,3,0,2.0881836414,2.2687482834,7
0,1,0,4,3,2.1653063297,-5.035178185,0
1,0,0,4,2,0.55752635,-2.696651697,0
0,1,1,3,2,-0.627695501,-3.22431159,0
1,0,1,1,0,-3.275050163,0.9133918285,0
0,1,0,2,0,0.3073975742,-0.431482762,0
0,1,0,3,2,0.4593037367,-3.140106678,0
1,1,1,1,0,-0.188099161,3.2674453259,0
0,1,0,2,1,-0.19742322,-0.238331914,0
0,1,0,4,0,0.9011332393,1.3939222097,2
0,1,1,4,1,2.2297375202,-0.410112113,3
1,1,0,2,1,1.1500788927,-0.320476443,1
0,1,1,3,0,1.7165309191,2.6540598869,5
0,1,0,1,0,-0.465738386,0.2460825294,0
0,1,1,1,0,1.0195733309,1.7188441753,2
1,0,0,3,1,1.8674137592,-0.54847914,1
1,0,0,4,1,0.7143785357,-2.550681114,0
1,1,1,1,0,0.0440081209,2.2629549503,1
0,1,1,4,0,5.0050396919,3.5721342564,149
0,1,1,3,2,2.4615564346,-2.769872427,0
0,1,1,3,1,1.5704168081,-0.390616268,1
0,1,0,2,0,-1.490852475,0.0889994577,0
0,1,1,3,0,-0.865133345,0.9720797539,0
0,1,0,2,1,0.8338031769,0.0233576857,1
0,1,0,4,2,2.2084999084,-1.734373331,0
0,1,1,3,1,1.6332079172,-1.501252651,0
0,1,0,4,2,1.5240097046,-4.324279785,0
1,0,0,4,0,2.4913747311,-0.722057521,2
0,1,1,2,0,1.0847690105,2.9630026817,2
0,1,0,4,0,3.5946772099,0.860208869,29
0,1,1,1,0,1.1288346052,2.0607004166,3
1,0,1,2,0,-0.385939926,2.2662763596,0
0,1,0,4,2,2.2135095596,-2.06809473,0
1,0,0,3,0,2.5187780857,-0.110921405,5
0,1,1,2,0,0.0115894843,-0.33921802,0
0,1,0,4,1,3.2988061905,-2.671430111,0
0,1,1,1,0,-0.258302957,0.7147980928,0
```

```
0,1,0,3,1,0.931161046,-1.925231814,0
0,1,1,4,1,3.2343406677,-1.934120536,0
0,1,1,4,1,0.5665621161,1.3611514568,1
0,1,1,1,0,1.9740951061,3.1780841351,7
0,1,1,2,0,0.1099227741,2.1244986057,1
0,1,0,1,0,-1.052196026,-1.837709427,0
1,0,0,4,1,1.1932103634,0.7328076363,2
0,1,0,3,2,0.3515802324,-2.184268236,0
1,1,0,2,0,1.1800631285,1.2939589024,2
0,1,0,2,1,0.4423512518,-0.252750695,0
0,1,0,4,1,0.3395077288,-0.768189728,0
1,0,1,3,1,0.368057549,-0.848371208,0
1,1,1,3,2,2.4312150478,-0.945337653,1
0,1,0,2,1,-0.267399877,-1.620429158,0
1,1,0,1,0,-0.36659652,2.514541626,0
1,0,1,4,3,1.4816583395,-2.82096076,0
0,1,0,1,0,-0.620765746,0.5831140876,0
1,0,1,4,3,1.5936685801,-2.947399616,0
0,1,0,4,1,2.8242943287,-0.742032945,3
0,1,1,1,0,1.4875385761,2.2652254105,4
0,1,1,3,0,1.1390730143,3.4486987591,3
0,1,0,2,0,1.4158217907,1.2972024679,3
0,1,1,4,0,2.1720778942,4.2631850243,8
1,0,1,3,0,0.8180824518,3.0585968494,2
1,1,1,1,0,0.3325906694,2.0458590984,1
0,1,0,4,0,1.8568427563,2.5167958736,6
1,1,1,4,2,0.5936886668,-1.977207899,0
0,1,1,2,1,0.0617640056,1.2698074579,0
1,1,0,4,0,1.9847823381,0.8246141672,5
0,1,1,4,1,1.9638563395,0.0315203704,3
0,1,0,3,1,3.7260725498,0.7016750574,31
0,1,1,2,1,0.3493249118,-0.161479443,0
0,1,0,2,0,0.7040555477,2.3449170589,2
1,1,1,4,3,1.5692437887,-4.469278812,0
0,1,1,2,1,0.781064868,-1.890136719,0
1,0,0,3,0,0.139555037,0.3391566873,0
0,1,1,3,1,1.5416204929,-1.531070232,0
1,1,0,1,0,0.0718735531,1.3057957888,0
0,1,0,2,0,0.7465148568,-0.650594831,0
1,1,1,4,0,1.9250824451,3.5670778751,6
0,1,0,3,0,2.2142121792,2.6915719509,9
1,0,1,3,2,-0.599701524,-2.059751987,0
1,0,1,2,1,-2.107331038,0.1413461715,0
0,1,1,1,0,-2.490455389,2.0737595558,0
0,1,1,2,1,-0.070476264,-0.478490353,0
0,1,0,1,0,-0.233503744,0.1525063664,0
0,1,0,2,0,2.1488587856,-0.488574415,2
0,1,0,3,0,3.1737546921,0.378390342,15
0,1,1,1,0,0.6767058372,1.6237850189,1
1,1,1,2,0,0.9033687115,2.0153822899,2
0,1,1,3,1,2.1214718819,-0.091606267,3
1,1,1,1,0,-1.680656552,2.2772350311,0
0,1,1,4,0,4.2684879303,1.3622959852,65
```

```
0,1,1,3,0,1.7596712112,1.4781696796,5
0,1,0,1,0,0.0071598762,0.1845640987,0
1,1,1,3,2,3.5338106155,-3.252695322,0
0,1,0,4,1,1.902828455,-1.590401888,0
0,1,0,4,2,2.0543143749,-3.139189243,0
1,1,0,1,0,0.1320486069,2.2440671921,1
0,1,1,4,1,2.8445887566,-0.078207836,8
0,1,1,3,0,-0.150227517,2.4962151051,0
0,1,1,3,2,0.3775939643,-2.29437232,0
0,1,1,4,3,1.461001873,-3.943372011,0
0,1,1,3,0,0.9519529343,2.6063089371,2
0,1,1,2,1,1.531072855,1.537260294,4
1,1,0,4,0,1.7160952091,1.8269017935,5
0,1,0,2,0,2.2420418262,2.3986170292,9
0,1,0,4,2,3.3890376091,-2.893302679,0
0,1,0,2,0,-0.457464665,3.17933321,0
0,1,1,2,0,-0.263669461,1.7608680725,0
0,1,0,1,0,-0.71243757,1.1771656275,0
1,1,1,2,0,3.0539169312,2.4691078663,21
0,1,0,1,0,-2.708276749,0.3473309875,0
0,1,1,3,0,1.8720514774,2.3079695702,6
0,1,0,2,1,0.1697490066,-0.835122049,0
1,0,1,2,0,-0.348856091,2.94145751,0
0,1,1,2,1,0.2402192205,-0.403173774,0
1,1,0,3,1,2.4038770199,-1.568528771,0
0,1,1,4,1,2.9885745049,0.9501032233,16
1,0,0,3,2,0.005257762,-3.463099957,0
0,1,1,3,2,0.4007853568,-1.719052434,0
1,1,1,4,0,1.5923120975,3.4721469879,4
0,1,1,2,1,1.022531271,0.5986690521,2
0,1,1,3,0,2.3773417473,2.5786168575,10
1,0,1,1,0,-2.112927198,1.6301124096,0
0,1,1,1,0,-0.557507157,1.9381258488,0
1,0,0,1,0,-1.276221037,0.145860225,0
1,0,1,2,0,1.0607860088,2.9731755257,2
1,1,1,2,0,0.4099714458,1.287543416,1
0,1,0,1,0,1.4727476835,1.0017284155,3
0,1,1,3,1,1.1222449541,-1.014991045,0
0,1,1,4,2,2.2423763275,-2.386286736,0
0,1,1,2,0,3.1638433933,1.2968349457,21
0,1,0,2,1,-0.585306764,-2.07876277,0
0,1,0,2,0,-0.015607119,2.2990167141,0
1,0,1,1,0,0.8344883323,2.2055110931,2
1,0,1,2,1,-1.268876076,1.5709195137,0
0,1,1,4,1,2.7570819855,-0.866619527,3
0,1,1,3,1,1.703949213,-1.154581547,0
0,1,1,4,0,3.7650024891,1.1892702579,38
1,1,1,4,3,2.7190511227,-4.614192963,0
0,1,1,1,0,-0.154021844,2.4653201103,0
0,1,0,1,0,-0.411525637,0.257135123,0
0,1,1,1,0,0.5311220288,2.9289655685,1
0,1,0,1,0,1.3403061628,3.5663721561,3
0,1,1,1,0,-1.562050104,1.4067739248,0
```

```
0,1,0,2,0,0.2253839523,2.2209911346,1
1,1,0,4,2,4.5026636124,-4.730084896,0
1,0,0,2,1,0.1621997654,-1.125854254,0
0,1,0,4,2,3.0470135212,-3.184262991,0
0,1,0,3,1,1.6307533979,-2.707653284,0
0,1,1,4,2,1.9832550287,0.5738066435,5
0,1,0,4,2,3.8771114349,-2.179812431,0
1,0,0,4,1,1.5847555399,-1.289599061,0
0,1,0,3,0,1.7011432648,0.0051179253,2
1,1,1,2,1,-0.170604989,1.469863534,0
0,1,1,1,0,-0.49698019,0.957906127,0
0,1,0,2,1,1.2182699442,-0.664241612,0
0,1,1,2,0,0.0375366136,2.8617730141,1
1,1,1,3,1,1.5232441425,3.0395727158,4
1,1,0,1,0,-0.907698154,0.6114888191,0
0,1,1,3,2,2.3313765526,-2.355525732,0
0,1,1,2,1,1.0680860281,1.375056386,2
0,1,1,1,0,1.4012657404,2.1410496235,3
1,1,0,4,2,3.9304218292,-2.74390316,0
0,1,0,3,1,0.1962469816,-1.781073928,0
0,1,0,2,0,0.0295273811,1.002204299,0
1,0,0,2,0,-0.766901433,-0.44208771,0
0,1,1,1,0,0.6611886621,3.2413334846,1
0,1,1,2,0,1.0150740147,0.9489593506,2
0,1,1,1,0,-1.046668053,0.7609766126,0
0,1,1,3,0,1.9380548,2.1590342522,6
0,1,1,2,0,1.7247505188,0.927846849,4
0,1,1,2,0,0.602655232,3.5716090202,1
0,1,1,4,2,2.4137816429,-1.316399097,1
1,0,0,2,0,-1.20076859,1.0577998161,0
0,1,1,3,1,1.9638493061,-0.733001232,1
0,1,0,3,0,-0.291065693,1.3155082464,0
1,1,1,2,0,-0.755235732,2.3242087364,0
0,1,1,4,3,1.7948591709,-5.625943661,0
0,1,1,2,1,-0.392648846,0.6772754192,0
1,1,1,3,2,1.374640584,-2.595630169,0
1,1,1,2,1,0.8288341165,-1.457115412,0
;;;;
run;quit;

proc means data = fish mean std min max var;
  var count child persons;
run;

proc univariate data = fish noprint;
  histogram count / midpoints = 0 to 50 by 1 vscale = count ;
run;

proc freq data = fish;
  tables camper;
run;

proc genmod data = fish;
  class camper;
```

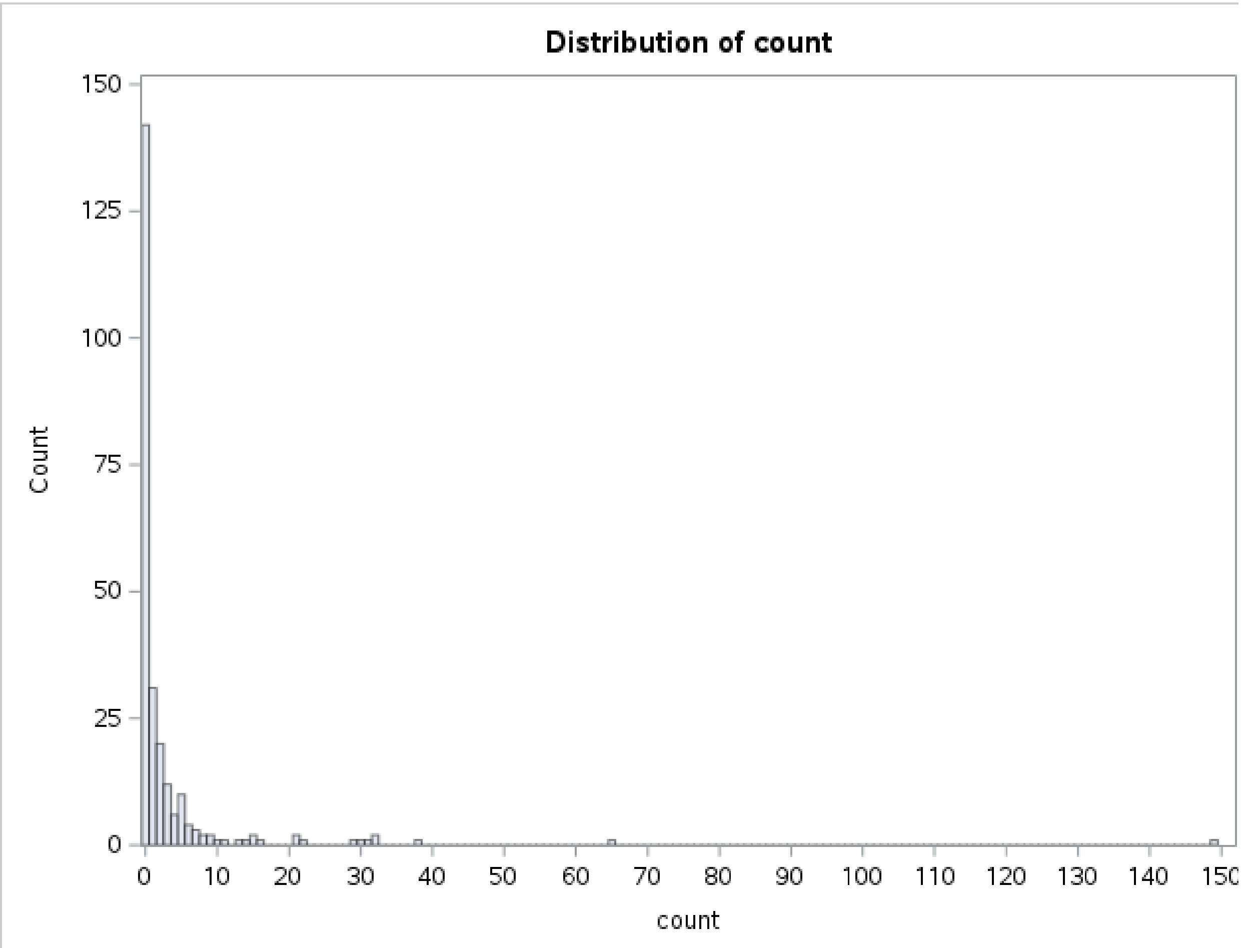
```
  model count = child camper /dist=zip;
  zeromodel persons /link = logit;
run;

proc genmod data = fish;
  class camper;
  model count = child camper /dist=zip;
  zeromodel persons /link = logit ;
  estimate "camper = 0" intercept 1 child .684 camper 1 0 @ZERO intercept 1 persons 2.
  estimate "camper = 1" intercept 1 child .684 camper 0 1 @ZERO intercept 1 persons 2.
run;
```

The MEANS Procedure

Variable	Mean	Std Dev	Minimum	Maximum	Variance
count	3.2960000	11.6350281	0	149.0000000	135.3738795
child	0.6840000	0.8503153	0	3.0000000	0.7230361
persons	2.5280000	1.1127303	1.0000000	4.0000000	1.2381687

The UNIVARIATE Procedure



The FREQ Procedure

camper	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	103	41.20	103	41.20

camper	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	147	58.80	250	100.00

The SAS System

The GENMOD Procedure

Model Information		
Data Set	WORK.FISH	Written by SAS
Distribution	Zero Inflated Poisson	
Link Function	Log	
Dependent Variable	count	
Zero Model Link Function	Logit	

Number of Observations Read	250
Number of Observations Used	250

Class Level Information		
Class	Levels	Values
camper	2	0 1

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance		2063.2168	
Scaled Deviance		2063.2168	
Pearson Chi-Square	245	1543.4597	6.2998
Scaled Pearson X2	245	1543.4597	6.2998
Log Likelihood		774.8999	
Full Log Likelihood		-1031.6084	
AIC (smaller is better)		2073.2168	
AICC (smaller is better)		2073.4627	
BIC (smaller is better)		2090.8241	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	2.4319	0.0413	2.3510	2.5128	3472.23	<.0001
child		1	-1.0428	0.1000	-1.2388	-0.8469	108.78	<.0001
camper	0	1	-0.8340	0.0936	-1.0175	-0.6505	79.35	<.0001
camper	1	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

Analysis Of Maximum Likelihood Zero Inflation Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	1.2974	0.3739	0.5647	2.0302	12.04	0.0005
persons	1	-0.5643	0.1630	-0.8838	-0.2449	11.99	0.0005

The SAS System

The GENMOD Procedure

Model Information		
Data Set	WORK.FISH	Written by SAS
Distribution	Zero Inflated Poisson	
Link Function	Log	
Dependent Variable	count	
Zero Model Link Function	Logit	

Number of Observations Read	250
Number of Observations Used	250

Class Level Information		
Class	Levels	Values
camper	2	0 1

Parameter Information		
Parameter	Effect	camper
Prm1	Intercept	

Parameter Information		
Parameter	Effect	camper
Prm2	child	
Prm3	camper	0
Prm4	camper	1

Zero Inflation Parameter Information	
Parameter	Effect
Prm5	Intercept
Prm6	persons

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance		2063.2168	
Scaled Deviance		2063.2168	
Pearson Chi-Square	245	1543.4597	6.2998
Scaled Pearson X2	245	1543.4597	6.2998
Log Likelihood		774.8999	
Full Log Likelihood		-1031.6084	
AIC (smaller is better)		2073.2168	
AICC (smaller is better)		2073.4627	
BIC (smaller is better)		2090.8241	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	2.4319	0.0413	2.3510	2.5128	3472.23	<.0001
child		1	-1.0428	0.1000	-1.2388	-0.8469	108.78	<.0001
camper	0	1	-0.8340	0.0936	-1.0175	-0.6505	79.35	<.0001
camper	1	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

Analysis Of Maximum Likelihood Zero Inflation Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	1.2974	0.3739	0.5647	2.0302	12.04	0.0005
persons	1	-0.5643	0.1630	-0.8838	-0.2449	11.99	0.0005

Contrast Estimate Result										
Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiS
camper = 0	2.4220	1.9724	2.9741	0.8846	0.1048	0.05	0.6792	1.0899	71.28	<.000
camper = 0 (Zero Inflation)	0.4677	0.3838	0.5536	-0.1292	0.1756	0.05	-0.4735	0.2150	0.54	0.461
camper = 1	5.5768	4.8823	6.3701	1.7186	0.0679	0.05	1.5856	1.8516	641.42	<.000
camper = 1 (Zero Inflation)	0.4677	0.3838	0.5536	-0.1292	0.1756	0.05	-0.4735	0.2150	0.54	0.461

In [3]:
```
proc genmod data = fish;
  class camper;
  model count = child camper /dist=nb;
run;
```

The GENMOD Procedure

Model Information		
Data Set	WORK.FISH	Written by SAS
Distribution	Negative Binomial	
Link Function	Log	
Dependent Variable	count	

Number of Observations Read	250
Number of Observations Used	250

Class Level Information		
Class	Levels	Values
camper	2	0 1

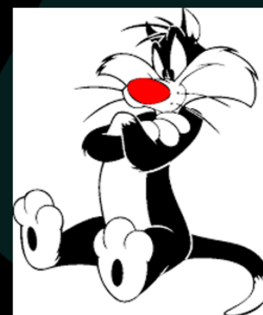
Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	247	201.8949	0.8174
Scaled Deviance	247	201.8949	0.8174
Pearson Chi-Square	247	551.5932	2.2332
Scaled Pearson X2	247	551.5932	2.2332
Log Likelihood		1366.7980	
Full Log Likelihood		-439.7103	
AIC (smaller is better)		887.4206	
AICC (smaller is better)		887.5839	
BIC (smaller is better)		901.5064	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	1.9821	0.2040	1.5823	2.3818	94.45	<.0001
child		1	-1.3753	0.2082	-1.7834	-0.9672	43.62	<.0001
camper	0	1	-0.9094	0.2831	-1.4642	-0.3545	10.32	0.0013
camper	1	0	0.0000	0.0000	0.0000	0.0000	.	.
Dispersion		1	3.9171	0.5060	3.0410	5.0455		

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

Tweedie Distribution



What is Tweedie?

- Nonnegative support
- Can have a discrete mass at zero

- Is a member of the Exponential family (is found in GENMOD)

- Includes several important distributions as special cases:
  - Normal (p=0)
  - Poisson (p=1)
  - Gamma (p=2)
  - Inv Gaussian (p=3)

Distribution	Range Requirements
Beta	Values must be between 0 and 1, exclusive.
Binary	Exactly two distinct values
Exponential	Nonnegative real values
Gamma	Nonnegative real values
Geometric	Positive integers
Inverse Gaussian	Positive real values
Negative Binomial	Nonnegative integers
Normal	Real values
Poisson	Nonnegative integers
Tweedie	Nonnegative real values

In [1]:

```
ods listing;

%let nObs = 250;
%let nClass = 5;
%let nLevs = 4;
%let seed = 100;

data tmp1;
   array c{&nClass};

   keep c1-c&nClass yTweedie d1 d2;

   /* Tweedie parms */
   phi=0.5;
   p=1.5;

   do i=1 to &nObs;

      do j=1 to &nClass;
         c{j} = int(ranuni(1)*&nLevs);
      end;

      d1 = ranuni(&seed);
      d2 = ranuni(&seed);

      xBeta   =  0.5*((c2<2) - 2*(c1=1) + 0.5*c&nClass + 0.05*d1);
      mu      =  exp(xBeta);

      /* Poisson distributions parms */
      lambda = mu**(2-p)/(phi*(2-p));
      /* Gamma distribution parms */
      alpha = (2-p)/(p-1);
      gamma = phi*(p-1)*(mu**(p-1));
```

```
      rpoi = ranpoi(&seed,lambda);
      if rpoi=0 then yTweedie=0;
      else do;
         yTweedie=0;
         do j=1 to rpoi;
         yTweedie = yTweedie + rangam(&seed,alpha);
         end;
         yTweedie = yTweedie * gamma;
      end;
      output;
   end;
run;

proc genmod data=tmp1;
   class C1-C5;
   model yTweedie = C1-C5 D1 D2 / dist=Tweedie type3;
run;

proc genmod data=tmp1;
   class C1 C2;
   model yTweedie = C1 C2 D1 / dist=Tweedie(p=1.5) type3;
run;
```

```
SAS server started using Context SAS Studio compute context with SESSION_ID=548e576e-ac08-475e-a48a-f14ce7225f2e-ses0000
```

The GENMOD Procedure

Model Information	
Data Set	WORK.TMP1
Distribution	Tweedie
Link Function	Log
Dependent Variable	yTweedie
Number of Threads	32

Number of Observations Read	250
Number of Observations Used	250

Class Level Information		
Class	Levels	Values
c1	4	0 1 2 3
c2	4	0 1 2 3
c3	4	0 1 2 3
c4	4	0 1 2 3
c5	4	0 1 2 3

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Pearson Chi-Square	232	101.9124	0.4393
Scaled Pearson X2	232	251.5826	1.0844
Log Likelihood		-297.2106	
Full Log Likelihood		-297.2106	
AIC (smaller is better)		634.4212	
AICC (smaller is better)		638.0893	
BIC (smaller is better)		704.8504	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	0.7570	0.1684	0.4268	1.0871	20.20	<.0001
c1	0	1	-0.1840	0.0990	-0.3781	0.0101	3.45	0.0632

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
c1	1	1	-0.9397	0.1107	-1.1567	-0.7227	72.03	<.0001
c1	2	1	0.0267	0.0962	-0.1618	0.2153	0.08	0.7810
c1	3	0	0.0000	0.0000	0.0000	0.0000	.	.
c2	0	1	0.6219	0.1035	0.4190	0.8249	36.08	<.0001
c2	1	1	0.4987	0.1043	0.2943	0.7032	22.86	<.0001
c2	2	1	0.1062	0.1102	-0.1098	0.3222	0.93	0.3353
c2	3	0	0.0000	0.0000	0.0000	0.0000	.	.
c3	0	1	0.0103	0.1014	-0.1885	0.2090	0.01	0.9195
c3	1	1	-0.0503	0.0992	-0.2447	0.1441	0.26	0.6122
c3	2	1	-0.0437	0.1084	-0.2562	0.1687	0.16	0.6865
c3	3	0	0.0000	0.0000	0.0000	0.0000	.	.
c4	0	1	-0.1151	0.1072	-0.3252	0.0949	1.15	0.2826
c4	1	1	0.1244	0.1044	-0.0801	0.3290	1.42	0.2332
c4	2	1	0.1714	0.0953	-0.0154	0.3582	3.23	0.0721
c4	3	0	0.0000	0.0000	0.0000	0.0000	.	.
c5	0	1	-0.6632	0.1008	-0.8608	-0.4656	43.26	<.0001
c5	1	1	-0.5171	0.1036	-0.7202	-0.3141	24.93	<.0001
c5	2	1	-0.2547	0.0921	-0.4351	-0.0742	7.65	0.0057
c5	3	0	0.0000	0.0000	0.0000	0.0000	.	.
d1		1	-0.0067	0.1326	-0.2666	0.2531	0.00	0.9595
d2		1	-0.1440	0.1255	-0.3899	0.1019	1.32	0.2511
Dispersion		1	0.4051	0.0326	0.3412	0.4690		
Power		1	1.3607	0.0660	1.2313	1.4901		

Note: The Tweedie dispersion parameter was estimated by maximum likelihood.

Note: The Tweedie power parameter was estimated by maximum likelihood.

LR Statistics For Type 3 Analysis			
Source	DF	Chi-Square	Pr > ChiSq
c1	3	85.46	<.0001
c2	3	48.18	<.0001
c3	3	0.56	0.9050
c4	3	9.38	0.0247

LR Statistics For Type 3 Analysis			
Source	DF	Chi-Square	Pr > ChiSq
c5	3	47.76	<.0001
d1	1	0.00	0.9595
d2	1	1.31	0.2518

The SAS System

The GENMOD Procedure

Model Information	
Data Set	WORK.TMP1
Distribution	Tweedie
Link Function	Log
Dependent Variable	yTweedie
Number of Threads	32

Number of Observations Read	250
Number of Observations Used	250

Class Level Information		
Class	Levels	Values
c1	4	0 1 2 3
c2	4	0 1 2 3

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Pearson Chi-Square	242	125.9834	0.5206
Scaled Pearson X2	242	254.4442	1.0514
Log Likelihood		-328.1245	
Full Log Likelihood		-328.1245	
AIC (smaller is better)		674.2490	
AICC (smaller is better)		674.9990	
BIC (smaller is better)		705.9421	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	0.3440	0.1347	0.0801	0.6080	6.53	0.0106
c1	0	1	-0.0722	0.1101	-0.2880	0.1436	0.43	0.5120
c1	1	1	-0.8952	0.1196	-1.1296	-0.6607	56.01	<.0001
c1	2	1	0.0770	0.1073	-0.1334	0.2873	0.51	0.4733
c1	3	0	0.0000	0.0000	0.0000	0.0000	.	.
c2	0	1	0.6138	0.1161	0.3862	0.8414	27.93	<.0001
c2	1	1	0.5103	0.1150	0.2849	0.7356	19.70	<.0001
c2	2	1	0.1001	0.1215	-0.1380	0.3381	0.68	0.4099
c2	3	0	0.0000	0.0000	0.0000	0.0000	.	.
d1		1	-0.0211	0.1493	-0.3136	0.2714	0.02	0.8876
Dispersion		1	0.4951	0.0398	0.4172	0.5731		
Power		0	1.5000	0.0000	1.5000	1.5000		

Note: The Tweedie dispersion parameter was estimated by maximum likelihood.

Note: The Tweedie power parameter was held fixed.

LR Statistics For Type 3 Analysis			
Source	DF	Chi-Square	Pr > ChiSq
c1	3	72.11	<.0001
c2	3	39.24	<.0001
d1	1	0.02	0.8876

Resources

https://go.documentation.sas.com/doc/en/pgmsascdc/v_054/statug/statug_genmod_overview07.htı

https://support.sas.com/resources/papers/proceedings/proceedings/sugi26/p264-26.pdf

https://www.lexjansen.com/wuss/2006/tutorials/TUT-Smith.pdf

https://www.lexjansen.com/nesug/nesug99/po/po140.pdf

https://support.sas.com/resources/papers/sgf2008/countreg.pdf

https://stats.oarc.ucla.edu/sas/dae/zero-inflatedpoisson-regression/


sas.com
sas
Copyright © SAS Institute Inc. All rights reserved.

In [ ]: