

Poisson Data

The other response variable of interest records the number of blue spots observed after incubation. This type of data, i.e. count data, is often skewed showing numerous small values with occasional large ones. Note that these counts differ from the binomial case above. There, the count of successes was bounded by the number of experimental units: 3 reps x 8 explants = 24. In this example, the number of spots observed is unbounded or open ended. The maximum number possible is not known. A probability distribution which can describe this type of data is the Poisson.

The SAS code to analyze Poisson data is similar to that of the Binomial:

```
proc genmod;  
  class treat;  
  model spots = treat/dist=poisson link = log type3;  
  lsmeans treat/diff;  
  contrast 'Osmotic Effect' treat -4 1 1 1 1;  
  by cultivar;
```

with the obvious exceptions of the DIST and LINK options. In the

case of the Poisson distribution, the log LINK is most appropriate (log here designates natural log, ln). However, as was seen with the binomial, a count of 0 can cause errors. Therefore, a data step can be used prior to the analysis to adjust these values before PROC GENMOD is called:

```
data osmotic;
  set osmotic;
  if spots = 0 then spots = .5;
```

Again, a small value is added to those observations having a count of 0 spots. The output for cultivar MUM follows.

----- cult=Mum -----

Class Level Information			
Class	Levels	Values	
treat	5	0	0.2 0.4 0.6 1

Source	DF	Chi-Square	Pr > ChiSq
treat	5	6895.92	<.0001

Least Squares Means

Effect	treat	Estimate	Standard Error	DF	Chi-Square	Pr > ChiSq
treat	0	1.3652	0.1031	1	175.20	<.0001
treat	0.2	3.2974	0.0393	1	7056.4	<.0001
treat	0.4	2.7267	0.0533	1	2613.4	<.0001
treat	0.6	3.1481	0.0423	1	5540.0	<.0001
treat	1	0.9328	0.1280	1	53.08	<.0001

Differences of Least Squares Means

Effect	treat	_treat	Estimate	Standard Error	DF	Chi-Square	Pr > ChiSq
treat	0	0.2	-1.9321	0.1104	1	306.52	<.0001
treat	0	0.4	-1.3615	0.1161	1	137.48	<.0001
treat	0	0.6	-1.7829	0.1115	1	255.78	<.0001
treat	0	1	0.4324	0.1644	1	6.92	0.0085
treat	0.2	0.4	0.5707	0.0662	1	74.25	<.0001
treat	0.2	0.6	0.1493	0.0577	1	6.69	0.0097
treat	0.2	1	2.3646	0.1339	1	311.76	<.0001
treat	0.4	0.6	-0.4214	0.0681	1	38.32	<.0001
treat	0.4	1	1.7939	0.1387	1	167.27	<.0001
treat	0.6	1	2.2153	0.1348	1	269.90	<.0001

Contrast Results				
Contrast	DF	Chi-Square	Pr > ChiSq	Type
Osmotic Effect	1	144.89	<.0001	LR

The treatment level information is given followed by the overall test of model effects. Here, TREAT is highly significant with a p-value less than 0.0001.

The means are given in the next table. As was the case above, these reflect the transformed (log) values. All the treatment means are significantly different from zero indicating that all treatments had more than one spot. Pair-wise comparison of the means is given next. In this example, all the pair-wise comparisons are significant, i.e. all treatments are apparently different from one another.

Finally, the contrast for any osmotic effect is also highly significant with a chi-square value of 144.89.

Given this significance, it might be of interest to actually estimate the mean number of spots for each component of the contrast, i.e. a mean for no osmotic pressure (equilibrium) and a mean where osmotic pressure was applied. This can be accomplished with the ESTIMATE statement. However, because SAS implicitly includes an overall intercept in the model, the actual ESTIMATE statement can be

rather tricky to formulate. One method of avoiding this is to rerun the model without the intercept term:

```
proc genmod;
  class treat;
  model spots = treat/dist=poisson link = log type3 noint;
  lsmeans treat/diff;
  contrast 'Osmotic Effect' treat -4 1 1 1 1;
  estimate 'Equilibrium' treat 1 0 0 0 0;
  estimate 'Osmotic Applied' treat 0 .25 .25 .25 .25;
  by cult;
```

The intercept is omitted with the NOINT option in the MODEL statement. Two ESTIMATE statements have been added; one for the osmotic treatment of 0 and another for the average of the remaining treatments. Note that, like CONTRAST, each statement has a unique label given in quotes. The resulting output is:

Contrast Estimate Results							
Label	Estimate	Standard Error	Alpha	Confidence Limits		Chi-Square	Pr > ChiSq
Equilibrium	1.3652	0.1031	0.05	1.1631	1.5674	175.20	<.0001

Osmotic Applied 2.5263 0.0376 0.05 2.4526 2.5999 4524.6 <.0001

Each ESTIMATE statement produces the value of its estimate, the associated standard error, and confidence limits. Also given are test results for the hypothesis that the estimate is equal to 0. (Notice that the 'Equilibrium' estimate is the same as that for TREAT = 0 above).

The results show that the application of osmotic pressure increased the number of spots over the equilibrium condition. Since these are still the log transformed values, the actual count estimates can be determined by utilizing the exponential (anti-log) function:

$$\text{Equilibrium number of spots} = e^{1.3652} = 3.92$$

and

$$\text{Avg. number of spots with treatment} = e^{2.5263} = 12.51.$$

This shows the osmotic treatments on average gave more than a 3 fold increase in the number of spots. The associated standard errors and confidence limits may also be back-transformed in a similar manner.