SAS Work Shop
PROC GENMOD
Handout #3

Statistical Programs
College of Agriculture
HTTP://WWW.UIDAHO.EDU/AG/STATPROG

Discrete Data Examples

Binomial Data

In the example data set, GUS is a binary variable recording 1 if blue spots did appear after incubation or 0 if they did not. The appropriate probability distribution for this data would be the binomial. Since the experiment was conducted as a completely random design, only the treatment effect is necessary for the model. In this example, the response, GUS, is either {1, 0}. The PROC GENMOD code used directly for this binary data form is:

```
proc sort data=osmotic;
     by cult;

proc genmod;
     class treat;
     model gus = treat/dist=binomial link = logit type3;
     lsmeans treat/diff;
     by cult;
run;
```

The CLASS statement tells SAS to use the variable TREAT as discrete levels. The binomial distribution is specified with the DIST= option and the LINK is logit. The logit transformation is often used with binomial data for theoretical reasons and is defined as:

$$\text{Logit} = \ln(p / (1-p)) = \ln(p) - \ln(1-p)$$

where p is the proportion of successes. In this case, SAS uses the GUS value of 1 to define a success as the proportion of explants showing blue spots.

The LSMEANS statement requests mean values for each level of TREAT and the DIFF option will test for equality among all possible pairs of treatments. The TYPE3 option produces a table similar to that of an ANOVA, and the 'BY' statement will result in separate analyses for each cultivar.

While this code is technically correct, a problem will occur for the example data. For some of the replication - treatment combinations, the proportion of blue spots may be 0 or 1. Either of these values are problematic for the logit calculation and will result in an error. One method of handling this condition is to redefine those values so that p is not exactly 1 or 0, e.g. .9999 and .0001, respectively. With the raw binary data, this can be done only by adding pseudo observations to the data and results in very coarse adjustments.

It would be much more advantageous to work with the proportions directly. To do this, a summary data set with the proportion of 1's is initially generated. This can be obtained using PROC MEANS as follows:

```
proc sort data=osmotic;
     by cultivar treatment;

proc means data=osmotic noprint;
     var gus;
     output out=test mean = prop n=total;
     by cultivar treatment;
```

Here, SAS will calculate the proportion of successes (1's) for each cultivar/treatment combination. These are put into variable PROP in the data set TEST. In addition, the total number of observations (1's and 0's) in each cultivar/treatment combination is calculated and placed in the variable TOTAL. For the analysis, PROC GENMOD can not use the proportion directly, but instead requires the number of successes. This can be calculated from the proportions as COUNT = PROP*TOTAL as part of a data step in the following manner:

```
data test;
     set test;
     count = prop*total;
     if prop = 1 then count = count - .5;
     if prop = 0 then count = .5;

proc print data=test;
```

This data step also checks for the condition p=1 or p=0 and adjusts the value of COUNT by adding or subtracting a small amount. The data are then printed as:

```
 Obs    cult    treat    _TYPE_    _FREQ_      prop      total     count


   1    Mum      0.0        0        24      0.75000      24       18.0
   2    Mum      0.2        0        24      1.00000      24       23.5
   3    Mum      0.4        0        24      0.95652      23       22.0
   4    Mum      0.6        0        24      1.00000      24       23.5
   5    Mum      1.0        0        24      0.75000      24       18.0
   6    PJM      0.0        0        24      0.35000      20        7.0
   7    PJM      0.2        0        24      0.17647      17        3.0
   8    PJM      0.4        0        24      0.20000      15        3.0
   9    PJM      0.6        0        24      0.06250      16        1.0
  10    PJM      1.0        0        24      0.04348      23        1.0
```

PROC GENMOD can then be called with:

```
proc genmod;
     class treatment;
     model count/total = treatment/dist=binomial link = logit
type3;
     lsmeans treatment/diff;
     contrast 'Osmotic Effect' treat -4 1 1 1 1;
     by cultivar;
```

This model is the same as that given earlier with the exception of the response. In this case, we give the ratio of successes (COUNT) to the total number of observations (TOTAL). The resulting output for the cultivar MUM is:

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


                                     The GENMOD Procedure

                                   Class Level Information

                            Class      Levels    Values

                            treat           5    0 0.2 0.4 0.6 1



                               LR Statistics For Type 3 Analysis

                                                     Chi-
                         Source             DF     Square     Pr > ChiSq

                         treat               4      14.27         0.0065
```

```
                              Least Squares Means


                                   Standard              Chi-
      Effect    treat  Estimate       Error      DF    Square    Pr > ChiSq


      treat     0        1.0986      0.4714       1      5.43        0.0198
      treat     0.2      3.8501      1.4292       1      7.26        0.0071
      treat     0.4      3.0910      1.0225       1      9.14        0.0025
      treat     0.6      3.8501      1.4292       1      7.26        0.0071
      treat     1        1.0986      0.4714       1      5.43        0.0198
```

```
                       Differences of Least Squares Means


                                         Standard              Chi-
Effect    treat    _treat    Estimate       Error      DF    Square    Pr > ChiSq


treat     0        0.2        -2.7515      1.5049       1      3.34        0.0675
treat     0        0.4        -1.9924      1.1259       1      3.13        0.0768
treat     0        0.6        -2.7515      1.5049       1      3.34        0.0675
treat     0        1           0.0000      0.6667       1      0.00        1.0000
treat     0.2      0.4         0.7591      1.7573       1      0.19        0.6658
treat     0.2      0.6         0.0000      2.0212       1      0.00        1.0000
treat     0.2      1           2.7515      1.5049       1      3.34        0.0675
treat     0.4      0.6        -0.7591      1.7573       1      0.19        0.6658
treat     0.4      1           1.9924      1.1259       1      3.13        0.0768
treat     0.6      1           2.7515      1.5049       1      3.34        0.0675
```

```
                                Contrast Results


                                         Chi-
          Contrast              DF     Square    Pr > ChiSq    Type


          Osmotic Effect         1       6.93        0.0085    LR
```

The first part of the output lists the 5 levels that SAS found for the TREAT variable. Following this is the ANOVA - like table produced from the TYPE3 option. In this case, the effect of TREAT had 4 degrees of freedom (5 levels minus 1) and a chi-square value of 14.27 which was significant with a p-value of .0065.

The results of the LSMEANS statement are given next. The numbers reported are for the logit transformation, not the proportions. For example, in TREAT=0, the proportion of successes was .75 and, hence, the corresponding logit value would be ln(.75) - ln(1-.75) = 1.0986. It is important to remember that with the binomial, the chi-square tests and associated p-values refer to these logit values, not the proportions themselves. In all treatments for this data set, the tests were significant. Therefore, we would conclude that the osmotic treatment appears to have shifted the proportions of success and failure away from the equal distribution of 50/50.

The next table has the results of the DIFF option. Here, all possible pair-wise comparisons of logit values are produced. For example, the first line of the table shows the estimated difference between TREAT=0 and TREAT=0.2. The value of the difference (logits) is -2.75 and the corresponding chi-square is 3.34. The p-value of 0.0675 indicates only a marginal difference in the success rates for these osmotic pressures. For TREAT=0 and TREAT=1.0, however, the p-value is 1.0, confirming the previous printout which showed that the two treatments have exactly the same value of PROP, 0.75.

The last portion of the printout presents another method of comparing treatment levels, the CONTRAST statement. In the SAS program, the statement:

```
contrast 'Osmotic Effect' treat -4 1 1 1 1;
```

is used to compare the first osmotic treatment, 0, to the average of the other treatments. This is essentially designed to assess any overall effect of osmotic pressure on the success rate. The contrast results in a chi-square value of 6.93 and an associated p-value of 0.0085. Thus, we can conclude that there is an overall effect on the success rate from the application of osmotic pressure.