Laboratory #9a.
Problem-Solving with the GLM. I. Setting up the Analysis.

The purpose of this lab is to gain practice in setting up statistical analyses. This is the most important part of statistical analysis, as it determines the analytic path. At the same time it is the most difficult. It is difficult because it requires unfamiliar forms of abstract thinking. It requires identification of variables, including inventing the name of a variable to include multiple categories. It requires a decision as to type of measurement scale–nominal (categorical), ordinal (ranks), interval, or ratio. This is by no means obvious. For example, variables such as time can be treated as nominal, ordinal, or ratio. Setting up an analysis requires separating response from explanatory variables. While this is usually clear for experimental studies, it is not clearcut where there are many variables with complex or unknown causal relations among them.

In lab 9a you will gain experience in setting up model based analysis of data. There are three data sets, each more challenging than the previous. The first example is a sophisticated experimental design called a latin square. Instead of being asked to look up the prescription for a latin square, you will be asked to use what you have learned to set up the analysis. The second example is an extension of the first. It does not have a name and so cannot be found by name in a text. With the model based approach, you will be able to set up this analysis. The third example is data for which there is again no named test. For this example, there are several possible models depending on how the response variable is defined. Upon completing Lab 9 you will have a greater capacity to undertake statistical analysis than if you had spent an entire year learning a battery of tests and their names.

All three data sets consist of counts. They are thus further challenging in that heterogeneous errors may be a problem, as is often the case with count data. You will be asked to use judgement in deciding whether heterogeneity is present and if so, how severe. And you will be expected to judge what action to take, in light of the severity of heterogeneity, sample size, and whether a recomputed p-value (by randomization) is necessary.

The 3 data sets are available on the website for this course
www.mun.ca/biology/schneider/b4605/Labs/Data
They are printed for you, below, with descriptive statistics.

wworm1.txt
wworm2.txt
leprosy.txt

You are encouraged to work in groups to decide how to analyze each data set.

The three examples use data from:

> Snedecor, G.W. and Cochran, W.G. (1980) Statistical Methods 7th ed. Iowa State University Press: Ames, Iowa

Honours students with a data set of their own can substitute their data set for one of the three sets in Lab 9 analyses. If you decide to use your own data set, be sure to describe (in a short paragraph) how the data were collected.

The write-up for lab 9a (step 1 in the generic recipe for statistical analysis with the GLM) and the write-up for lab 9b in the next lab period (steps 2-10 in the generic recipe) are both due together after lab 9b.

```
4 3    0 2    2 5    3 1    1 4
1 6    3 0    0 6    2 4    4 4
0 4    1 9    3 1    4 6    2 5
2 17   4 8    1 8    0 9    3 0
3 4    2 4    4 2    1 4    0 8

T  N   T N    T N    T N    T N
```

wworm1.txtt
Snedecor and Cochran (1980) p 289

N = number of wire worms per 9 x 9 x 5 inch plot
T = treatment: control (0) and 4 soil fumigants.
Plots are arranged in a 5 x 5 array.

Each treatment occurs in each row and in each column. This is called a Latin Square

```
MTB > print c11-c14
  ROW  trtmnt  count   col   rows
    1       4      3     1      1
    2       1      6     1      2
    3       0      4     1      3
    4       2     17     1      4
    5       3      4     1      5
    6       0      2     2      1
    7       3      0     2      2
    8       1      9     2      3
    9       4      8     2      4
   10       2      4     2      5
   11       2      5     3      1
   12       0      6     3      2
   13       3      1     3      3
   14       1      8     3      4
   15       4      2     3      5
   16       3      1     4      1
   17       2      4     4      2
   18       4      6     4      3
   19       0      9     4      4
   20       1      4     4      5
   21       1      4     5      1
   22       4      4     5      2
   23       2      5     5      3
   24       3      0     5      4
   25       0      8     5      5
```

continued... (over)

```
MTB > plot 'count' 'trtmnt'                                          wworm1.txt
     18.0+
         _                                *
 count   _
         _
         _
         _
     12.0+
         _
         _
         _   2            2                                   *
         _
      6.0+   *            *                                   *
         _                                2
         _   *            2               2            *      *
         _   *                                                2
         _                                           2
         _                                           2
      0.0+
         _
           --+---------+---------+---------+---------+---------+----trtmnt
          0.00      0.80      1.60      2.40      3.20      4.00

 MTB > describe 'count';
 SUBC> by 'rows'.
              rows        N     MEAN   MEDIAN   TRMEAN    STDEV   SEMEAN
 count           1        5    3.000    3.000    3.000    1.581    0.707
                 2        5     4.00     4.00     4.00     2.45     1.10
                 3        5     5.00     5.00     5.00     2.92     1.30
                 4        5     8.40     8.00     8.40     6.02     2.69
                 5        5    4.400    4.000    4.400    2.191    0.980

 MTB > describe 'count';
 SUBC> by 'col'.
               col        N     MEAN   MEDIAN   TRMEAN    STDEV   SEMEAN
 count           1        5     6.80     4.00     6.80     5.81     2.60
                 2        5     4.60     4.00     4.60     3.85     1.72
                 3        5     4.40     5.00     4.40     2.88     1.29
                 4        5     4.80     4.00     4.80     2.95     1.32
```

Count variation relative to both random factors. Pct variance = $(SS_{rows} + SS_{cols})/SS_{total}$ = _____

```
MTB > anova 'count' = 'rows' 'col';
SUBC> residuals c15.

 Analysis of Variance for count

 Source      DF         SS         MS       F      P
 rows         4      84.56     21.140    1.61  0.220
 col          4      22.16      5.540    0.42  0.791
 Error       16     210.24     13.140
 Total       24     316.96     13.207
```

wworm2.txt

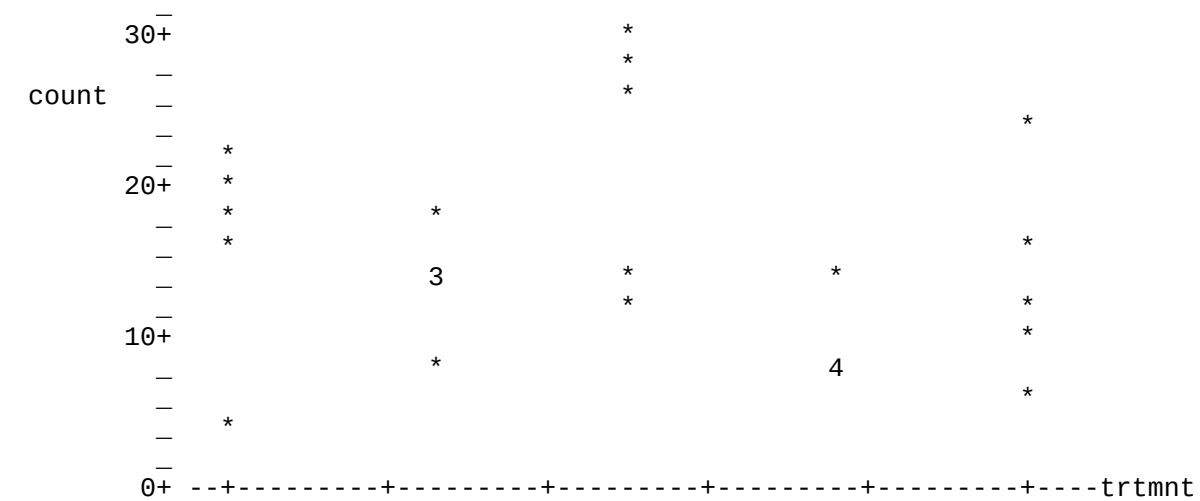
```
4  6    0  3     2 29     3  8     1 17
1  8    3 13     0 18     2 12     4 16
0 15    1 13     3  7     4 10     2 28
2 14    4 11     1 13     0 22     3  7
3  7    2 26     4 24     1 14     0 20

T  N    T  N     T  N     T  N     T  N
```

Snedecor and Cochran (1980)
p 273

T = treatment: 0 = contrl 1,2,3,4 = 4 soil fumigants
N = number of wire worms per 9 x 9 x 5 inch plot
Each treatment occurs in each row, and in each column of the 5 x 5 array.
This is called a Latin Square layout.
This is a repeat (1 year later) of the experiment reported as 'WWORM1.dat'

wworm2.txt

```
MTB > plot 'count' 'trtmnt'
```



continued... (over)

wworm2.txtt (continued)

```
MTB > describe 'count';
SUBC> by 'rows'.
               rows       N      MEAN    MEDIAN    TRMEAN     STDEV    SEMEAN
 count            1       5     12.60      8.00     12.60     10.55      4.72
                  2       5     13.40     13.00     13.40      3.85      1.72
                  3       5     14.60     13.00     14.60      8.08      3.61
                  4       5     13.40     13.00     13.40      5.50      2.46
                  5       5     18.20     20.00     18.20      7.76      3.47

MTB > describe 'count';
SUBC> by 'col'.

               col        N      MEAN    MEDIAN    TRMEAN     STDEV    SEMEAN
 count            1       5     10.00      8.00     10.00      4.18      1.87
                  2       5     13.20     13.00     13.20      8.26      3.69
                  3       5     18.20     18.00     18.20      8.70      3.89
                  4       5     13.20     12.00     13.20      5.40      2.42
                  5       5     17.60     17.00     17.60      7.57      3.39
```

leprosy.txt

```
11  6   6  0   16  13
 8  0   6  2   13  10
 5  2   7  3   11  18
14  8   8  1    9   5
19 11  18 18   21  23
 6  4   8  4   16  12
10 13  19 14   12   5
 6  1   8  9   12  16
11  8   5  1    7   1
 3  0  15  9   12  20

 B  A   B  A    B   A
 TRI    TRII  Control
```

Snedecor and Cochran (1980) p 368

Scores for Leprosy bacilli before (B) and after (A) treatment with two types of antibiotics (TRI and TRII), with one control (C)

```
MTB > read 'leprosy.dat' c1-c6;
MTB > describe c1-c6
                N     MEAN   MEDIAN   TRMEAN    STDEV   SEMEAN
 bTRI          10     9.30     9.00     8.88     4.76     1.51
 aTRI          10     5.30     5.00     5.00     4.64     1.47
 bTRII         10    10.00     8.00     9.50     5.25     1.66
 aTRII         10     6.10     3.50     5.37     6.15     1.95
 bCONT         10    12.90    12.00    12.62     3.96     1.25
 aCONT         10    12.30    12.50    12.37     7.15     2.26

 MTB > plot c2 c1
    15.0+
        _
 aTRI   _                          *
        _                                                    *
        _
    10.0+
        _                                  *         *
        _
        _                                  *
        _
     5.0+
        _             *
        _
        _          *
        _             *
     0.0+   *               *
         --+---------+---------+---------+---------+---------+---------+----bTRI
           3.0       6.0       9.0      12.0      15.0      18.0
```

continued... (over)

leprosy.txt (continued)

```
MTB > plot c4 c3

    18.0+                                                           *
        _
aTRII   _
        _                                                           *
        _
        _
    12.0+
        _
        _
        _               *                           *
        _
     6.0+
        _
        _               *
        _       *   *
        _  *            *
        _
     0.0+       *
        _
          +---------+---------+---------+---------+---------+------bTRII
         5.0       7.5      10.0      12.5      15.0      17.5

MTB > plot c6 c5
    24.0+                                                           *
        _
aCONT   _                       *
        _
        _                   *
        _
    16.0+                       *
        _
        _                                       2
        _
        _                           *
        _
     8.0+
        _
        _           *           *
        _
        _  *
        _
     0.0+
        _
          --+---------+---------+---------+---------+---------+----bCONT
            7.5      10.0      12.5      15.0      17.5      20.0
```