Partition (Decompose) a 5 x 2 Contingency Table using R

W. Gregory Alvord

We partition, or decompose a 5 x 2 contingency table of psychiatric patients cros classified as to their diagnostic category and whether they were prescribed drugs. We will discover that for some diagnostic groups, psychiatrists prescribe drugs more often than not. For other groups they prescribe, or do not prescribe drugs, about equally. For one group, psychiatrists are unlikely to prescribe drugs.

All of the R code required for the solution to this problem is supplied in the Appendix.

The example dataset (Table 3.10, Problem 3.6, page 72) is taken from Alan Agresti's ***Categorical Data Analysis*** (1990, 2nd ed.). 276 psychiatric patients were cross classified as to their diagnosis in one of five psychiatric groups: (1 Schizophrenia, (2) Affective Disorder, (3) Neurosis, (4) Personality Disorder, and (5 Special Symptoms and as to whether (or not) they were prescribed drugs in their treatment regimens. Here is the table as displayed in the R Statistical Environment (R Core Development Team, 2012).

```
> Ag.3.10.table # Observed Frequencies
                      Drugs.Rx
Diagnosis              Yes  No
  Schizophrenia        105   8
  Affective.Disorder    12   2
  Neurosis              18  19
  Personality.Disorder  47  52
  Special.Symptoms       0  13
```

The problem here is to partition, or decompose, the table in a statistically rigorous way to "describe similarities and differences among the diagnoses in terms of the relative frequencies of the prescribed drugs," (Agresti, page 72). The decomposition involves the partitioning of the contingency table and its corresponding Likelihood Ratio Chi-Square statistic, LR χ^2, into orthogonal, additive components (Agresti, pp 50-54). The advantage to partitioning a contingency table into orthogonal components is that independent inferences can be drawn for each component involved in the partitioning. "A [correct] partitioning may show than an association primarily reflects differences between certain categories or groupings of categories," (Agresti, page 50). Rules for partitioning the table are provided in the Appendix.

Our first general question is this: Do the data reveal a *relationship* between the patients' diagnostic class (**Diagnosis**) and whether or not drugs were prescribed (**Drugs.Rx**)? Under the assumption of *independence* (i.e., **Diagnosis** and **Drugs.Rx** are unrelated) we would *expect* the following frequencies (counts):

```
> round(Ag.3.10.chisq.test$exp,1) # Expected under
                        Drugs.Rx  # Independence
Diagnosis                Yes   No
  Schizophrenia         74.5 38.5
  Affective.Disorder     9.2  4.8
  Neurosis              24.4 12.6
  Personality.Disorder  65.3 33.7
  Special.Symptoms       8.6  4.4

> Ag.3.10.table # Observed Frequencies
                      Drugs.Rx
Diagnosis              Yes  No
  Schizophrenia        105   8
  Affective.Disorder    12   2
  Neurosis              18  19
  Personality.Disorder  47  52
  Special.Symptoms       0  13
```

The observed and expected frequencies appear to be different, but how different?

We can compute a formal statistical test for the differences between the observed and expected frequencies with a Chi-Square test. The Chi-Square test with which most are familiar is called the Pearson's Chi-Squared test (see Appendix). In many applications it is approximately equivalent to the less familiar, but more important, Likelihood Ratio Chi-Squared test. For technical reasons, we use the Likelihood Ratio Chi-Squared test value, also known as LR χ^2 (LR X^2, G^2) (se Appendix). A huge advantage associated with the LR χ^2 is that it can be broken down into statistically independent, orthogonal components, which are additive. This allows independent statistical inferences to be drawn for each specific component.

Unfortunately, there is no mechanical, automatic method for partitioning or 'decomposing' an I x J contingency table (here, a 5 x 2 table) into meaningful orthogonal components. Each problem involving Chi-Square decomposition is approached differently. This requires an active process of thought with a goal in mind; the 'goal' sometime emerges more clearly as one progresses through the analysis.

First, we perform a 'global' test for the hypothesis of independence (n association) between the variables **Diagnosis** and **Drugs.Rx**. We specify the null hypothesis, $\mathbf{H_0}$, and display some R output for the current problem.

$\mathbf{H_0}$: **Diagnosis** and **Drugs.Rx** are independent (statistically unrelated).

```
> Ag.3.10.global.loglm
Call:
loglm(formula = ~ Diagnosis + Drugs.Rx, data =
Ag.3.10.table)

Statistics:
                      X^2 df P(> X^2)
Likelihood Ratio 96.53689  4           ← p << 0.0001
```

The LR X^2 value is 96.53689 on 4 degrees of freedom. The probability of obtaining this statistic 'by chance' is extremely low, $p << 0.0001$. We reject the null hypothesis of independence and conclude that the variables **Diagnosis** and **Drugs.Rx** are ***not*** independent. They are highly related, or associated, statistically.

The purpose of *partitioning* (decomposing) the LR Chi-Square statistic is to gain further insight into the nature of the relationship between specific sets of diagnoses for patients and whether drugs were prescribed.

We review the original table.

```
 > Ag.3.10.table # Observed
                      Drugs.Rx
Diagnosis              Yes  No
  Schizophrenia        105   8
  Affective.Disorder    12   2
  Neurosis              18  19
  Personality.Disorder  47  52
  Special.Symptoms       0  13
```

Our next task is to identify two rows of this table (i.e., two psychiatric diagnostic groups) that appear to have comparable proportions (percentages) of cases classified as **Yes** (or alternatively as **No**). Displayed below is a table of percentages that, for each diagnostic group, sum to 100% across the two categories of whether or not drugs were prescribed (**Yes** or **No**).

```
> round(Ag.3.10.percent.mar.1.table, 1) # Percentage
                        Drugs.Rx
Diagnosis                Yes    No
  Schizophrenia         92.9   7.1 # Yes + No = 100%, etc.
  Affective.Disorder    85.7  14.3 # Yes + No = 100%, etc.
  Neurosis              48.6  51.4
  Personality.Disorder  47.5  52.5
  Special.Symptoms       0.0 100.0
```

We note that for **Neurosis**, 48.6% of patients were prescribed drugs (**Yes** while 51.4% were not (**No**). For **Personality.Disorder**, 47.5% were prescribed drugs while 52.5% were not. The percentage of patients who were prescribed drugs for **Neurosis** (48.6%) appears to be approximately comparable to that for patients who were prescribed drugs for **Personality.Disorder** (47.5%). From the original 5 x 2 table of observed frequencies, we extract this 2 x 2 sub-table of interest.

```
> Ag.3.10.rows.34.table # Observed
                     Drugs.Rx
Diagnosis              Yes No
  Neurosis              18 19
  Personality.Disorder  47 52
```

For this 2 x 2 sub-table, under the assumption that **Diagnosis** and **Drugs.Rx** are independent, we would *expect* to observe the following counts.

```
> round(Ag.3.10.rows.34.chisq.test$exp, 1) # Expected
                       Drugs.Rx  # under Independence
Diagnosis                Yes   No
  Neurosis             17.7 19.3
  Personality.Disorder 47.3 51.7
```

The expected values, under the null hypothesis of independence are very close to those observed for **Neurosis** and **Personality.Disorder** of the 2 x 2 table.

Statistically, we ask whether **Diagnosis** and **Drugs.Rx** are independent (i.e., uncorrelated) for these two diagnostic classes alone. We use the Likelihood Ratio Chi-Square test. Again we state the null hypothesis and show a portion of the R output.

H_0: Diagnosis (restricted to Neurosis and Personality Disorder) and **Drugs.Rx** are independent (uncorrelated).

```
> Ag.3.10.rows.34.loglm
Call:
loglm(formula = ~Diagnosis + Drugs.Rx, data =
Ag.3.10.rows.34.table)

Statistics:
                        X^2 df  P(> X^2)
Likelihood Ratio 0.01487122  1 0.9029405 ← p = 0.90 n.s.
```

The value of the LR X^2 statistic is **`0.01487122 = 0.015`**. (This is the *test* statistic, *not* the *probability* value associated with the test.) The LR test value is 'low' for 1 degree of freedom. The probability value for observing this outcome by chance is 'high', p = 0.9029, i.e., highly likely, p >> 0.05. We do not reject $\mathbf{H_0}$. We infer that **Diagnosis** and **Drugs.Rx** are independent, or uncorrelated, in this particular sub-table. Another way of phrasing this is to say that the entries across the cells of the table are *homogeneous*. Interpretation: The proportion (percentage) of patients classified with **Neurosis** who were prescribed drugs is comparable (statistically equivalent) to the proportion of patients classified with **Personality.Disorder** who also were prescribed drugs. [Alternatively, the proportion of patients in the class **Neurosis** who *were not* prescribed drugs is comparable to the proportion of patients classified with **Personality.Disorder** who *were not* prescribed drugs.]

With experience, one learns to distinguish other homogeneous patterns in the contingency tables. We review the frequency and percentage tables, restricting our attention to diagnostic levels for Schizophrenia, Affective Disorder and Special Symptoms to determine if we can further detect additional homogeneous sets of counts.

```
> Ag.3.10.table[c(1,2,4),] # Observed
                      Drugs.Rx
Diagnosis              Yes No
  Schizophrenia        105  8
  Affective.Disorder    12  2
  Personality.Disorder  47 52
```

Again, our task is to identify two or three rows of this table (i.e., two 'diagnoses') that appear to have comparable proportions of cases classified as **Yes** (or alternatively as **No**).

```
> round(Ag.3.10.percent.mar.1.table[c(1,2,5),],1) # Percent
                      Drugs.Rx
Diagnosis              Yes    No # Yes + No = 100%
  Schizophrenia       92.9   7.1 # = 100%, etc.
  Affective.Disorder 85.7  14.3
  Special.Symptoms     0.0 100.0
```

Scrutinizing the 'Percentage' table, for patients who were prescribed drugs (**Yes**) we note that 92.9% of patients diagnosed with **Schizophrenia** may be approximately comparable to 85.7% of patients diagnosed with **Affective.Disorder**. [Alternatviely, 7.1% (**Schizophrenia**) may be approximately comparable to 14.3% (**Affective.Disorder**) on **No**.] These two categories display substantially greater proportions for the **Yes** category than does **Special.Symptoms**, which display 0.0% on **Yes** and 100.0% on **No**. The **Yes/No** trend for the **Special.Symptoms** lies in the opposite direction of those for **Schizophrenia** and **Affective.Disorder**.

We extract the 2 x 2 sub-table considering only those cases associated with **Schizophrenia** and **Affective.Disorder**.

```
> Ag.3.10.rows.12.table # Observed
                    Drugs.Rx
Diagnosis            Yes  No
  Schizophrenia      105   8
  Affective.Disorder  12   2
```

We proceed as before. Under the assumption that **Diagnosis** and **Drugs.Rx** are independent, we would expect the following counts for rows 1 and 2 of the original table.

```
> round(Ag.3.10.rows.12.table.chisq.test$exp, 1) # Expected
                      Drugs.Rx         # under Independence
Diagnosis               Yes  No
  Schizophrenia      104.1 8.9
  Affective.Disorder  12.9 1.1
```

Again, under the assumption of independence, the expected values are quite close to those observed. We ask the same question: Are **Diagnosis** and **Drugs.Rx** independent for these two diagnostic classes (**Schizophrenia** and **Affective.Disorder**). We test this with Likelihood Ratio χ^2 as before. Here is null hypothesis and R output.

H_0: Diagnosis (restricted to **Schizophrenia** and **Affective.Disorder**) and **Drugs.Rx** are independent.

```
> Ag.3.10.rows.12.loglm
Call:
loglm(formula = ~Diagnosis + Drugs.Rx, data =
Ag.3.10.rows.12.table)

Statistics:
                      X^2 df  P(> X^2)
Likelihood Ratio 0.7529516  1 0.3855433 ← p = 0.39 n.s.
```

The LR X^2 statistic is `0.7529516 = 0.753`. It is 'low'. The probability value for observing this outcome by chance is p = 0.3855 ≈ 0.39, which means it is likely that it did occur by chance (the p value is > 0.05). We do *not* reject the null hypothesis, $\mathbf{H_0}$. For this sub-table **Diagnosis** and **Drugs.Rx** are statistically independent, or uncorrelated (homogeneous). Interpretation: The proportions of patients who were prescribed drugs in **Schizophrenia** and **Affective.Disorder** diagnostic classes are statistically equivalent. [Alternatively, the proportion of patients classified with **Schizophrenia** who *were not* prescribed drugs is statistically equivalent to the proportion of patients classified with **Affective.Disorder** who *were not* prescribed drugs.]

We have identified two 2 x sub-tables from the original that are homogeneous. In one, we found that **Neurosis** and **Personality.Disorder** were homogeneous. In another, we found that **Schizophrenia** and **Affective.Disorder** were homogeneous. When this occurs with a sub-table, the counts in the sub-table can be combined or 'collapsed', i.e., summed over its margins, without loss of information. The original 5 x 2 table can now be collapsed (combined) into 3 x 2 'Observed' and 'Percentage' tables.

```
> Ag.3.10.collapsed.table # Observed
                  Drugs.Rx
Diagnosis          Yes  No
  Schiz.or.Aff.Dis 117  10 # combining Schiz & Aff.Dis
  Neur.or.Pers.Dis  65  71 # combining Neur & Pers.Dis
  Special.Symptoms   0  13 # original Observed counts
```

The observed counts for **Schizophrenia** and **Affective.Disorder** are combined into a single category now labeled **Schiz.or.Aff.Dis**. Similarly, the observed counts for **Neurosis** and **Personality.Disorder** are combined into a single category now labeled **Neur.or.Pers.Dis**. Since the counts associated with **Special.Symptoms** have not been used in a previous sub-table, they are repeated here. Again, we present a table of percentages that, for each diagnostic group, sum to 100% across the two categories of whether or not drugs were prescribed.

```
> round(Ag.3.10.collapsed.percent.mar.1.table, 1)# Percent
                  Drugs.Rx
Diagnosis            Yes    No
  Schiz.or.Aff.Dis  92.1   7.9 # Yes + No = 100%, etc.
  Neur.or.Pers.Dis  47.8  52.2 # Yes + No = 100%, etc.
  Special.Symptoms   0.0 100.0 # Yes + No = 100%.
```

The pattern for how drugs were prescribed emerges more clearly. For the combined class of Schizophrenia or Affective Disorder (**Schiz.or.Aff.Dis**), relatively more patients (92.1%) were prescribed drugs than were not (7.9%). For the combined class of Neurosis or Personality Disorder (**Neur.or.Pers.Dis**) approximately equal numbers of patients (47.8% vs. 52.2%), were either prescribed drugs or not. For patients with Special Symptoms in this sample, 0.0% were prescribed drugs while 100% were not.

Again, we examine the observed counts in the 3 x 2 collapsed table and the expected counts under the assumption of independence.

```
> Ag.3.10.collapsed.table # Observed
                  Drugs.Rx
Diagnosis          Yes  No
  Schiz.or.Aff.Dis 117  10 # combining Schiz & Aff.Dis
  Neur.or.Pers.Dis  65  71 # combining Neur & Pers.Dis
  Special.Symptoms   0  13 # original Observed counts
```

Under the assumption of independence, we would expect to see the following counts:

```
> round(Ag.3.10.collapsed.table.chisq.test$exp,1)# Expected
                    Drugs.Rx           # under Independence
Diagnosis            Yes   No
  Schiz.or.Aff.Dis 83.7 43.3
  Neur.or.Pers.Dis 89.7 46.3
  Special.Symptoms  8.6  4.4
```

Under ‘independence’, we would *expect* to see approximately twice as many patients to be prescribed drugs as not, regardless of the diagnosis. We perform the LR Chi-square test again.

H_0: Diagnosis and **Drugs.Rx** (with some categories combined) are independent.

```
> Ag.3.10.collapsed.loglm
Call:
loglm(formula = ~Diagnosis + Drugs.Rx, data =
Ag.3.10.collapsed.table)

Statistics:
                      X^2 df P(> X^2)
Likelihood Ratio 95.76907  2            ← p << 0.0001
```

For this sub-table the LR X^2 value is 96.76907 on 2 degrees of freedom. The probability of obtaining this statistic by chance is p << 0.0001. We conclude that variables **Diagnosis** and **Drugs.Rx**, for these subsets of psychiatric categories, are ***not*** independent. They are highly related.

To this point, we have made several discoveries. We first determined that for the entire original table, the variables **Diagnosis** and **Drugs.Rx** were not independent; they were highly related. Next, we found that patients diagnosed with either Schizophrenia or Affective Disorder, were homogeneous with respect to the proportions that were or were not prescribed drugs. We also found this to be true of patients who were diagnosed with Neurosis or Personality Disorder. When the five diagnostic classes were combined (collapsed) into three diagnostic groups, we again found that the variable **Diagnosis** and **Drugs.Rx** were not independent, but were, in fact, highly related.

In the beginning of this document we indicated that the general problem was to partition, or decompose the table in a statistically rigorous way to describe differences and similarities among the diagnoses in terms of the relative frequencies of the prescribed drugs. The decomposition involves the partitioning of the Likelihood Ratio Chi-Square statistic, LR X^2, into orthogonal, additive components.

When the partitioning is performed in the correct way, the LR X^2 values of the sub-tables sum, exactly, to the LR X^2 value for the original table. Similarly, the degrees of freedom associated with each test sum to the degrees of freedom associated with the test from the original table. We have correctly followed the rules for decomposing the original 5 x 2 table into two 2 x 2 tables and a 3 x 2 table. We again display the original 5 x 2 table.

```
> Ag.3.10.table # Observed
                      Drugs.Rx
Diagnosis              Yes  No
  Schizophrenia        105   8
  Affective.Disorder    12   2
  Neurosis              18  19
  Personality.Disorder  47  52
  Special.Symptoms       0  13
```

The LR X^2 value for the original table was 96.53689, on 4 degrees of freedom:

```
> Ag.3.10.global.loglm$lr
[1] 96.53689
```

Next we re-display the 2 x 2 table that was restricted to **Schizophrenia** and **Affective.Disorder**.

```
> Ag.3.10.rows.12.table
                    Drugs.Rx
Diagnosis             Yes  No
  Schizophrenia       105   8
  Affective.Disorder   12   2
```

The LR X^2 value for this 2 x 2 table was 0.75295, on 1 degree of freedom:

```
> Ag.3.10.rows.12.loglm$lr
[1] 0.7529516
```

Next we re-display the 2 x 2 table that was restricted to **Neurosis** and **Personality.Disorder**.

```
> Ag.3.10.rows.34.table
                      Drugs.Rx
Diagnosis               Yes No
  Neurosis               18 19
  Personality.Disorder   47 52
```

The LR X^2 value for this 2 x 2 table was 0.01487, on 1 degree of freedom:

```
> Ag.3.10.rows.34.loglm$lr
[1] 0.01487122
```

We re-display the 3 x 2 table that showed entries for which (1) Schizophrenia and Affective Disorder were combined (**Schiz.or.Aff.Dis**), (2) Neurosis and Personality Disorder were combined (**Neur.or.Pers.Dis**), and (3) Special Symptoms (**Special.Symptoms)** remained as it was in the original table.

```
> Ag.3.10.collapsed.table
                  Drugs.Rx
Diagnosis           Yes  No
  Schiz.or.Aff.Dis 117  10
  Neur.or.Pers.Dis  65  71
  Special.Symptoms   0  13
```

The LR X^2 value for this 3 x 2 table was 95.76907, on 2 degrees of freedom:

```
> Ag.3.10.collapsed.loglm$lr
[1] 95.76907
```

Add the three LR X^2 values:

```
> Ag.3.10.rows.12.loglm$lr + Ag.3.10.rows.34.loglm$lr +
Ag.3.10.collapsed.loglm$lr
[1] 96.53689
```

... and compare to LR X^2 value for the original 5 x 2 table, on 4 degrees of freedom,

```
> Ag.3.10.global.loglm$lr
[1] 96.53689
```

They are equal. (A further, more exacting test to demonstrate this result is shown in the Appendix.) Also, the degrees of freedom for each component are, respectively, 1 1, and 2, which sum to 4 degrees of freedom associated with the original table.

Summary and Interpretation

We began with a 5 x 2 contingency table of 276 psychiatric patients that were cros classified as to their diagnosis in one of five psychiatric diagnostic groups b whether they were prescribed drugs in their treatment regimens.

```
> Ag.3.10.table # Observed
                       Drugs.Rx
Diagnosis               Yes  No
  Schizophrenia         105   8
  Affective.Disorder     12   2
  Neurosis               18  19
  Personality.Disorder   47  52
  Special.Symptoms        0  13
```

The patterns for prescribing drugs were homogeneous for the **Schizophrenia** and **Affective.Disorder** diagnostic groups. These categories were combined. Similarly, the patterns for prescribing drugs were homogeneous for **Neurosis** and **Personality.Disorder**, but differed from those for **Schizophrenia** and **Affective.Disorder**. The categories for **Neurosis** and **Personality.Disorder** were also combined. The pattern for prescribing drugs for **Special.Symptoms** differed from all the other diagnostic groups. Following the rules for partitioning contingency tables, the original 5 x 2 contingency table was decomposed into a 3 x 2 table that showed three distinct patterns for prescribing drugs depending on patients' respective diagnoses. The 'collapsed' 3 x 2 table is shown here.

```
> Ag.3.10.collapsed.table # Observed
                  Drugs.Rx
Diagnosis          Yes  No
  Schiz.or.Aff.Dis 117  10
  Neur.or.Pers.Dis  65  71
  Special.Symptoms   0  13

> round(Ag.3.10.collapsed.percent.mar.1.table, 1) # Percent
                  Drugs.Rx
Diagnosis            Yes    No
  Schiz.or.Aff.Dis  92.1   7.9
  Neur.or.Pers.Dis  47.8  52.2
  Special.Symptoms   0.0 100.0
```

Psychiatric patients were relatively more or less likely to be prescribed drugs depending on their respective diagnoses. Patients diagnosed with Schizophrenia or Affective Disorder were *more* likely to be prescribed drugs than not (92.1% vs. 7.9%). Patients diagnosed with Neurosis or Personality Disorder were about *equally* likely to be prescribed drugs or not (47.8% vs. 52.2%). And patients with Special Symptoms were *not* likely to be prescribed drugs; in fact, no drugs were prescribed for these patients in this sample (0.0% vs. 100.0%).

References

Agresti A, *Categorical Data Analysis, 2nd ed.*, Wiley, New York, 1990.

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/

Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

John Fox and Sanford Weisberg (2011). An {R} Companion to Applied Regression, Second Edition. Thousand Oaks CA: Sage. URL: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Appendix

Chi-Square Formulas

The Pearson Chi-Squared test value is

$$\text{Pearson } \chi^2 = \sum\sum \{ [(O_{ij} - E_{ij})^2] / E_{ij} \}$$

where O_{ij} and E_{ij} are Observed and Expected values in the i,j^th cell, respectively, and the symbol $\sum\sum$ indicates summation over the (i,j) cells in a given table, i = 1, …, I (number of rows in the table) and j = 1, …, J (number of columns in the table).

The Likelihood Ratio (LR) Chi-Square test value has the form

$$\text{LR } \chi^2 = 2 \bullet \sum\sum O_{ij} \bullet \log (O_{ij}/E_{ij})$$

where O_{ij} and E_{ij} are Observed and Expected values, 'log' denotes the natural (Naperian) logarithm, and the symbol $\sum\sum$ indicates summation over the (i,j) cells in a given table, i = 1, …, I (number of rows in the table) and j = 1, …, J (number of columns in the table).

Rules for Partitioning

Here is a brief list of some of the **Rules for Partitioning** the contingency table (Agresti, page 53):

1. The degrees of freedom for the sub-tables must sum to the degrees of freedom for the original table.
2. Each cell count in the original table must be a cell count in one and only one sub-table.
3. Each marginal total of the original table must be a marginal total for one and only one sub-table.

Appendix

```
##
## Consider problem of decomposition of I * J (here 5 x 2)
 matrix.
## See Agresti, page 72, Table 3.10
## Load libraries
##
library(MASS) # Venables and Ripley
library(car) # Fox and Sanford
##
## Set up the table
##
Ag.3.10.table.entries <- c(105, 12, 18, 47, 0, 8, 2, 19,
 52, 13)
Ag.3.10.mat <- matrix(Ag.3.10.table.entries, nrow = 5,
 byrow = FALSE, dimnames = list(Diagnosis =
 c('Schizophrenia', 'Affective.Disorder', 'Neurosis',
 'Personality.Disorder', 'Special.Symptoms'), Drugs.Rx =
 c('Yes', 'No')))
Ag.3.10.table <- as.table(Ag.3.10.mat)
Ag.3.10.df <- as.data.frame(Ag.3.10.table)
## We can easily get the LR G^2 statistic with
## either loglm() or glm().
## For loglm(), use Ag.3.10.table
Ag.3.10.global.loglm <- loglm( ~ Diagnosis + Drugs.Rx, data
 = Ag.3.10.table)
Ag.3.10.global.loglm
## Execute glm model.
## We can get this from glm as well, ... family =
 poisson...
## For glm(), use Ag.3.10.df
Ag.3.10.global.glm <- glm(Freq ~ Diagnosis + Drugs.Rx,
data = Ag.3.10.df, family = poisson)
Anova(Ag.3.10.global.glm, type = 'II')
summary(Ag.3.10.global.glm)
## Compute expected values under H0: Independence
Ag.3.10.chisq.test <- chisq.test(Ag.3.10.table)
```

```
## Expected values in 'exp' attribute
round(Ag.3.10.chisq.test$exp, 1)
## Use prop table to get proportions by row (Diagnosis)
Ag.3.10.prop.mar.1.table = prop.table(Ag.3.10.table, margin
 = 1)
Ag.3.10.prop.mar.1.table
##
Ag.3.10.percent.mar.1.table = 100*Ag.3.10.prop.mar.1.table
round(Ag.3.10.percent.mar.1.table, 1)
## Plot the 'percentage' table
Ag.3.10.percent.mar.1.table.df <-
 as.data.frame(Ag.3.10.percent.mar.1.table, responseName =
 'Percentage.within.Diagnosis')
##
with(Ag.3.10.percent.mar.1.table.df,
 interaction.plot(Diagnosis, Drugs.Rx,
 Percentage.within.Diagnosis, lty = c(1,1,1,1,1), pch =
 c(7,7,7,7,7), col = c('blue', 'red', 'black', 'orange',
 'turquoise'), lwd = 3, type = 'b', xlab = 'Diagnosis', ylab
 = 'Percentage within Diagnosis', ylim = c(0, 100), main =
 'Percentages (adding to 100) within Diagnosis'))
##
## Use prop table to get proportions by column
 (Drug:Yes/No)
##
Ag.3.10.prop.mar.2.table = prop.table(Ag.3.10.table, margin
 = 2)
Ag.3.10.prop.mar.2.table
##
Ag.3.10.percent.mar.2.table = 100*Ag.3.10.prop.mar.2.table
round(Ag.3.10.percent.mar.2.table, 2)
##
## Now hunt for homogeneous subsets. . .
##
## Create table for rows 3 & 4
Ag.3.10.rows.34.table <- as.table(Ag.3.10.table[3:4,])
Ag.3.10.rows.34.table
Ag.3.10.rows.34.chisq.test <-
 chisq.test(Ag.3.10.rows.34.table)
round(Ag.3.10.rows.34.chisq.test$exp, 1)
##
Ag.3.10.rows.34.loglm <- loglm( ~ Diagnosis + Drugs.Rx,
 data = Ag.3.10.rows.34.table)
Ag.3.10.rows.34.loglm
##
## Create rows 1 & 2 subtable
Ag.3.10.rows.12.table <- as.table(Ag.3.10.table[1:2,])
Ag.3.10.rows.12.table
```

```
##
## Get expected values under H0: Indpendence for this
 subtable
Ag.3.10.rows.12.table.chisq.test <-
 chisq.test(Ag.3.10.rows.12.table)
round(Ag.3.10.rows.12.table.chisq.test$exp, 1)
## Compute loglm object for this 2 x 2 subset
Ag.3.10.rows.12.loglm <- loglm( ~ Diagnosis + Drugs.Rx,
 data = Ag.3.10.rows.12.table)
Ag.3.10.rows.12.loglm
## perform some checks
apply(Ag.3.10.table[1:2,], 2, sum)
apply(Ag.3.10.table[3:4,], 2, sum)
##
## Re-display whole table for clarity...
Ag.3.10.table
##
Ag.3.10.collapsed.entries <- c(117, 65, 0, 10, 71, 13)
Ag.3.10.collapsed.mat <- matrix(Ag.3.10.collapsed.entries,
 nrow = 3, byrow = FALSE, dimnames = list(Diagnosis =
 c('Schiz.or.Aff.Dis', 'Neur.or.Pers.Dis',
 'Special.Symptoms'), Drugs.Rx = c('Yes', 'No')))
Ag.3.10.collapsed.table <- as.table(Ag.3.10.collapsed.mat)
Ag.3.10.collapsed.df <-
 as.data.frame(Ag.3.10.collapsed.table)
##
Ag.3.10.collapsed.prop.mar.1.table =
 prop.table(Ag.3.10.collapsed.table, margin = 1)
Ag.3.10.collapsed.prop.mar.1.table
round(Ag.3.10.collapsed.prop.mar.1.table, 1)
## turn proportions into percentages
Ag.3.10.collapsed.percent.mar.1.table =
 100*Ag.3.10.collapsed.prop.mar.1.table
round(Ag.3.10.collapsed.percent.mar.1.table, 1)
##
Ag.3.10.collapsed.loglm <- loglm( ~ Diagnosis + Drugs.Rx,
 data = Ag.3.10.collapsed.table)
Ag.3.10.collapsed.loglm
##
## get expected values...
Ag.3.10.collapsed.table.chisq.test <-
 chisq.test(Ag.3.10.collapsed.table)
round(Ag.3.10.collapsed.table.chisq.test$exp, 1)
##
## We follow the rules of partitioning when we
## take the sum across several runs (use apply())
## and insert them into a new table.
## The object Ag.3.10.global.loglm fit has 4 df
```

```
## and yields a LR G^2 of 96.53689.
## The object Ag.3.10.rows.12.loglm fit has 1 df
## and yields a LR G^2 of 0.7678228.
## The object Ag.3.10.rows.34.loglm fit has 1 df
## and yields a LR G^2 of 0.01487122.
## the object Ag.3.10.collapsed.loglm fit has 2 df
## and yields a LR G^2 of 95.76907
##
## The following command shows this to be true:
## test for equivalence
round(Ag.3.10.rows.12.loglm$lr + Ag.3.10.rows.34.loglm$lr +
 Ag.3.10.collapsed.loglm$lr, 12) ==
 round(Ag.3.10.global.loglm$lr, 12)
```