ABSTRACT

In the epidemiological world, we often encounter the following analytical question: is the relationship between exposure and outcome different for different strata? The Breslow-Day test (with or without the Tarone adjustment) can be used in PROC FREQ to assess homogeneity across a series of $2 \times 2$ tables, but what if your tables are not $2 \times 2$? One alternative is to fit a log-linear model to a $k \times R \times C$ table and test the fit with the three-way interaction removed. This can be done easily in PROC CATMOD or, with some additional effort, in PROC GENMOD and PROC GLIMMIX. A real-world data example will demonstrate the details of different methods to analyze $k \times R \times C$ tables and tests for homogeneity.

INTRODUCTION

One frequent question that epidemiologists come across is whether the relationship between exposure and outcome is different for different strata. For example, in the study of treatment responses, we may want to test the hypothesis that the distribution between the study drug and a traditional therapy dose change is the same at each level of disease severity. In this case, the distribution we are interested in encompasses not only an exposure-outcome relationship but also the homogeneity across different strata. Homogeneity is defined as: the conditional relationship between any pair of variables given a third one is the same at each level of that third variable.

Many of us are familiar with PROC FREQ as a tool for assessing homogeneity in a contingency table using count data. Specifically, the Breslow-Day (BD) Statistic produced using the Cochran Mantel Haenzel (CMH) option in PROC FREQ is commonly used for this purpose. If we add the BDT option, which requests the Breslow Day Statistic with the Tarone adjustment, SAS® will give an adjusted, asymptotically chi-squared result.

If we are interested in studying the distribution between exposures and outcomes that have more than two levels, we cannot use PROC FREQ to assess homogeneity. The Breslow-Day Statistic has not been generalized for this kind of $k \times R \times C$ table and can only be used for $k \times 2 \times 2$ tables. Fitting a log-linear model can solve this problem. On one hand, the GENMOD procedure and the GLIMMIX procedure can derive the homogeneity statistic in this situation with pooled data. On the other hand, PROC CATMOD is set up to achieve the goal without pooling the data first. In addition, the PROC CATMOD method does not require manual calculation of the likelihood ratio P values. We will provide examples to demonstrate the details of adopting different methods for analyzing $k \times R \times C$ tables.

SAMPLE DATA

In this paper, we will be using a sample dataset containing patients with medication information at baseline and 1 year follow-up, all of whom are receiving traditional therapy. Our exposure component is the treatment group: placebo, newly on study drug, and continuing on study drug. The outcome is the traditional therapy dose increase, decrease, or remaining the same from baseline to one-year follow-up. Finally, all the patients were stratified by physicians into moderate or severe disease categories. Our hypothesis is that moderate patients who start the study drug will decrease their traditional therapy use to a greater extent than severe patients who start the study drug.

$K \times 2 \times 2$ TABLES

In order to compare the PROC CATMOD output to the Breslow-Day output, we first collapsed the treatment into 2 groups: placebo and study drug groups. In addition, we grouped the ‘same’ and ‘increase’ traditional dose together. Figure 1 shows the collapsed data. A slight difference was shown between the distribution among the ‘decrease’ and ‘same/increase’ group in the placebo patients, i.e. 42% and 58% versus 46% and 54%. However, the row percents for the medicated patients are similar.
PROC FREQ is SAS’s basic procedure for the analysis of count data. The Breslow-Day (BD) Statistic that comes with the Cochran Mantel Haenszel (CMH) option in PROC FREQ is commonly used to test for homogeneity. It is a test of differences of the odds ratios from each stratum, and has an approximate chi-square distribution with k-1 degrees of freedom.

```sas
proc freq data=pt_meds01;
title4 "PROC FREQ With Breslow-Day Option";
table severity2 * grp2 * dose_cng2 /cmh;
run;
```

However, for the Breslow-Day test to be valid, the sample size should be relatively large in each stratum, and at least 80% of the expected cell counts should be greater than 5. Even when the Breslow-Day test is valid, it might not be very powerful against certain alternatives, as discussed in Breslow and Day (1980). If we add the BDT option after the CMH option, the SAS output will include the Breslow-Day statistic with the Tarone adjustment. Tarone derived an adjustment factor that is subtracted from the Breslow-Day statistic resulting in an asymptotically chi-squared statistic. This will relax the stricter sample size requirement. The Tarone adjustment was added to SAS in version 9.0.

```sas
proc freq data=pt_meds01;
title4 "PROC FREQ With Breslow-Day Option - Tarone Adjusted";
table severity2 * grp2 * dose_cng2 /cmh bdt;
run;
```

Here is part of the PROC FREQ outputs generated with and without the BDT option.

![Table](image)

Note that the Tarone adjusted chi-square value is slightly different. The $P$ value is 0.0833, and as a result, we conclude that there is no difference in traditional therapy dose change by study drug treatment when comparing patients with different disease severity. However, is this still true when we look at the expanded $2 \times 3 \times 3$ table sets?

**PROC CATMOD**

Since the Breslow-Day test is based on testing the homogeneity by comparing the actual odds ratios, it only applies to $k \times 2 \times 2$ tables. In our specific case, we need a generalized Breslow-Day test for the $k \times R \times C$ tables. Let’s now look at the original question without having to collapse our variables.
H0: The distribution between treatment group and traditional medication dosage change are the same in different severity groups
H1: There is an overall association between treatment group and traditional medication dosage change in different severity groups

This is where the CATMOD procedure is useful. First, let’s run it on the 2 x 2 x 2 data to compare with PROC FREQ.

```sas
proc catmod data=pt_meds01 order=data;
  title4 "Breslow-Day Analogue from PROC CATMOD";
  model severity2 * grp2 * dose_cng2 = _response_ /
    ml noiter noresponse nodesign nolrs NOPROFILE ZERO=sampling;
  loglin severity2|grp2|dose_cng2@2 /title="No 3-way association";
run;
quit;
```

The model statement has the variables disease severity, treatment group, and traditional therapy dose change in the same order as before. The keyword _RESPONSE_ tells PROC CATMOD that you want to model the variation among the dependent variables. In other words, it indicates that this is a log-linear model. The effects to be included in the log-linear model are specified by the LOGLIN statement. We specify all two-way interactions to be included in the LOGLIN statement. By doing so, we are asking the question in the null hypothesis: is the three-way interaction equal to zero? That is, does leaving out the three-way interaction still result in a model that fits? If it does, then we have homogeneity. The options NOITER, NORESPONSE, NODESIGN, NOLRS, and NOPROFILE suppress extra printed output that is not always needed for log-linear models. Figure 3 shows the result from PROC CATMOD.

Note that one drawback of using PROC CATMOD is that there’s no REF= option in the MODEL or CLASS statement as in other modeling procedures. Accordingly, when obtaining estimate from a parameter, be sure to code the variables to cater to your specific solution.

**Figure 3.**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>severity2</td>
<td>1</td>
<td>412.72</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>grp2</td>
<td>1</td>
<td>676.65</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>severity2*grp2</td>
<td>1</td>
<td>474.32</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>dose_cng2</td>
<td>1</td>
<td>74.72</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>severity2*dose_cng2</td>
<td>1</td>
<td>16.01</td>
<td>0.0722</td>
</tr>
<tr>
<td>grp2*dose_cng2</td>
<td>1</td>
<td>6.28</td>
<td>0.0122</td>
</tr>
<tr>
<td><strong>Likelihood Ratio</strong></td>
<td>1</td>
<td>2.99</td>
<td>0.0836</td>
</tr>
</tbody>
</table>

Here is a summary of the three methods: Breslow-Day, Tarone adjusted Breslow-Day, and PROC CATMOD.

**Figure 4.**

<table>
<thead>
<tr>
<th>Method</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breslow-Day</td>
<td>0.0833 (X²=2.9992)</td>
</tr>
<tr>
<td>Tarone Adjusted Breslow-Day</td>
<td>0.0833 (X²=2.9991)</td>
</tr>
<tr>
<td>PROC CATMOD</td>
<td>0.0836 (X²=2.9932)</td>
</tr>
</tbody>
</table>

The results from PROC FREQ are slightly different from the results from PROC CATMOD. The reason is that the methodologies used in these two procedures are different. In other words, the two methods do not exactly agree because they are not computing the same test statistic. PROC CATMOD fits a model to the data via maximum likelihood. It then tests for the parameter of the three-way interaction using a likelihood ratio test statistic for the parameter. PROC FREQ does not fit a model but uses a statistic to test the homogeneity of the odds ratio hypothesis. Consequently, the statistic in PROC FREQ is not algebraically equivalent to the test of the three-way interaction parameter in PROC CATMOD. We expect them to be very similar, but they are two different approaches to testing the same hypothesis.
K × R × C TABLES
Here is the 2 x 3 x 3 data. Comparing to the 2 x 2 x 2 tables, it appears to be more variability across severity levels. Note that in the placebo group, the distribution for the ‘decrease’ and the ‘same’ cells seem to be different for moderate versus severe patients: 42% and 46% versus 46% and 41% respectively.

Figure 5.

<table>
<thead>
<tr>
<th>Severity=MODERATE</th>
<th>Frequency</th>
<th>Decrease</th>
<th>Same</th>
<th>Increase</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Row Pct</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo</td>
<td></td>
<td>830</td>
<td>907</td>
<td>245</td>
<td>1982</td>
</tr>
<tr>
<td></td>
<td></td>
<td>41.88</td>
<td>45.76</td>
<td>12.36</td>
<td></td>
</tr>
<tr>
<td>Med: Continue</td>
<td></td>
<td>892</td>
<td>762</td>
<td>306</td>
<td>1960</td>
</tr>
<tr>
<td></td>
<td></td>
<td>45.51</td>
<td>38.88</td>
<td>15.61</td>
<td></td>
</tr>
<tr>
<td>Med: New</td>
<td></td>
<td>141</td>
<td>93</td>
<td>25</td>
<td>259</td>
</tr>
<tr>
<td></td>
<td></td>
<td>54.44</td>
<td>35.91</td>
<td>9.65</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1863</td>
<td>1762</td>
<td>576</td>
<td>4201</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Severity=SEVERE</th>
<th>Frequency</th>
<th>Decrease</th>
<th>Same</th>
<th>Increase</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Row Pct</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo</td>
<td></td>
<td>285</td>
<td>254</td>
<td>85</td>
<td>624</td>
</tr>
<tr>
<td></td>
<td></td>
<td>45.67</td>
<td>40.71</td>
<td>13.62</td>
<td></td>
</tr>
<tr>
<td>Med: Continue</td>
<td></td>
<td>915</td>
<td>886</td>
<td>246</td>
<td>2047</td>
</tr>
<tr>
<td></td>
<td></td>
<td>44.70</td>
<td>43.28</td>
<td>12.02</td>
<td></td>
</tr>
<tr>
<td>Med: New</td>
<td></td>
<td>136</td>
<td>91</td>
<td>28</td>
<td>255</td>
</tr>
<tr>
<td></td>
<td></td>
<td>53.39</td>
<td>35.69</td>
<td>10.98</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1336</td>
<td>1231</td>
<td>359</td>
<td>2926</td>
</tr>
</tbody>
</table>

The code and output below produces the final results to answer our study question.

```r
proc catmod data=pt_meds01 order=data;
    title4 "PROC CATMOD General Test of Homogeneity";
    model severity * grp * dose_cng = _response_ /
        ml noiter noresponse nodelay nolinks NOLR PROFILE ZERO=sampling;
    loglin severity|grp|dose_cng@2 /title="No 3-way association";
run;
quit;
```
Figure 6.

### Maximum Likelihood Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>severity</td>
<td>1</td>
<td>113.69</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>grp</td>
<td>2</td>
<td>1499.23</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>severity*grp</td>
<td>2</td>
<td>476.48</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>dose_cng</td>
<td>2</td>
<td>577.14</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>severity*dose_cng</td>
<td>2</td>
<td>4.48</td>
<td>0.1067</td>
</tr>
<tr>
<td>grp*dose_cng</td>
<td>4</td>
<td>27.09</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

The result shows a $P$ value of 0.0048. Previously, we were forced to collapse down our data to a 2 x 2 x 2 table. Using the Breslow-Day method in this case resulted in lost of information and non-significant results ($P$ value=.0833). By keeping the original categories, we now have sufficient power to reject the null hypothesis and conclude that there is an overall association between treatment group and traditional therapy dose change in different severity groups.

### CODING ALTERNATIVES

The coding methods described above are not the only way to generate homogeneity test. PROC GENMOD and PROC GLIMMIX have the ability to output the homogeneity statistics with the type 3 analysis option. Both methods fit generalized linear models using maximum-likelihood methods.

In order to test for the 3-way association using PROC GENMOD and PROC GLIMMIX, we need to first collapse the data into 'count' level. In other words, we summarized the data to obtain all the counts in each 2 x 3 x 3 combination as in Figure 5.

```latex
\begin{verbatim}
data pt_meds02;
  set pt_meds01;
  count=1;
run;

proc summary data=pt_meds02 nway;
  class severity grp dose_cng;
  var count;
  output out=pt_meds03 sum=;
run;
\end{verbatim}
```

**PROC GENMOD**

To fit a log-linear model with PROC GENMOD, Poisson regression was used. Here, the observed count is the response variable, and the log-linear effects appear on the right-hand side of the model equation.

```latex
\begin{verbatim}
proc genmod data=pt_meds03;
  class severity grp dose_cng / param=effect;
  model count=severity|grp|dose_cng /link=log dist=poisson type3;
run;
\end{verbatim}
```

The TYPE3 option requests that statistics for Type III contrasts be computed for each effect specified in the MODEL statement. The default analysis is to compute likelihood ratio statistics for the contrasts. This is an analog to Type III sums of squares in the GLM procedure. A Type III analysis does not depend on the order in which the terms for the model are specified here. A GENMOD Type III analysis consists of a table that contains the likelihood ratio statistics, degrees of freedom, and p-values based on the limiting chi-square distributions for each effect in the model. This analysis is intended to test the significance of the main effect in the presence of interactions. Therefore, when we fit a saturated model, the analysis for the term 'severity*grp*dose_cng' will give us the test for homogeneity, i.e. if the three-way interaction equal to zero.

The resulting $P$ value of 0.0048 from the effect 'severity*grp*dose_cng' draws the same conclusion as previously noted using PROC CATMOD.
PROC GLIMMIX
Similar code could be used in PROC GLIMMIX. Again, by default, PROC GLIMMIX computes the Type III Tests by first constructing a Type III L matrix for each fixed effect. In other words, a Type III (partial) test is created without using any keyword in the procedure, unlike Type I (sequential) test or Type II (adjusted) test. However, these tests are Wald type tests, not likelihood ratio tests, which generally are not quite as good. In order to obtain homogeneity estimates, the CHISQ option must be added which requests that chi-square tests be performed for all contrasts in addition to any F tests.

```plaintext
proc glimmix data=pt_meds03;
   class severity grp dose_cng;
   model count=severity|grp|dose_cng /link=log dist=poisson chisq;
run;
```

The same P value of 0.0048 from the effect ‘severity*grp*dose_cng’ from all three methods leads us to conclude that there is an overall association between new treatment and traditional therapy dose change among patients with different disease severity.

When using these different methods, beware that PROC CATMOD treats all explanatory (independent) variables as classification variables by default. Conversely, PROC GENMOD and PROC GLIMMIX treat covariates as continuous variables by default, and you must specify the classification variables in the CLASS statement. Furthermore, specifying a reference group by REF= option is only available in PROC GENMOD and PROC GLIMMIX.

**CONCLUSION**

The Breslow-Day test (with or without the Tarone adjustment) can be used in PROC FREQ to assess homogeneity across a series of k x 2 x 2 tables. Nonetheless, when one or two of the explanatory variables are more than two levels, we need to explore alternative solutions. Fitting a log-linear model to a k x R x C table and testing the fit with the three-way interaction removed using CATMOD is a creative way to achieve the goal. Specifying the actual model in the LOGLIN statement is an advantage when using PROC CATMOD. Other alternatives require some additional effort. For example, using PROC GENMOD specifying the TYPE3 option with pooled data could achieve the same goal by fitting a saturated model. Finally, a similar method to PROC GENMOD, a homogeneity test can be performed with a Type III analysis using PROC GLIMMIX with the CHISQ option.
REFERENCES


CONTACT INFORMATION
Your comments and questions are valued and encouraged. Contact the authors at:

Ginny P. Lai
ICON Late Phase & Outcomes Research
4350 La Jolla Village Dr., Suite 400
San Diego, CA 92122
(858) 795-8232
ginny.lai@iconplc.com
www.iconplc.com/

David R. Mink
ICON Late Phase & Outcomes Research
188 Embarcadero, Suite 200
San Francisco, CA 94105
(415) 371-2108
david.mink@iconplc.com
www.iconplc.com/

David J. Pasta
ICON Late Phase & Outcomes Research
188 Embarcadero, Suite 200
San Francisco, CA 94105
(415) 371-2111
david.pasta@iconplc.com
www.iconplc.com/

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.