Loglinear Models

Loglinear models are an alternative method of analyzing contingency tables, with the ability to test many of the same hypotheses we have discussed up to this point. It is different framework for conceptualizing categorical data that has many advantages because of its flexibility for constructing tests for more complex designs and ordinal variables.

Natural Logarithms

Before discussing loglinear models, we will need to review a few basic principles about logarithms, because natural log transformations are integral to the approach. Recall that the natural logarithm is complementary or inverse function to the exponential transformations. Though there are a several variants of the logarithm function, we will focus only on the natural logarithm which is the inverse of the exponent function. The natural logarithm is denoted either as $\log_e$ ($\log$ base e), $\ln$, or in many statistical texts just $\log$. I am not fond of just using "log" because of the potential ambiguity, but to remain consistent with your readings, I will always be referring to the natural logarithm ($\ln$) when using *log* below. The exponential function indicated by *e* (and also by exp) raises the constant named after the Swiss mathematician Leonhard Euler, $e \approx 2.71828$, to some power, say x. In general, we could state that $e^{\log x} = x$. If $x = 4.5$, then $e^x = \exp(x) = e^{4.5} = 90.017$. The natural log reverses this transformation, $\log_e(90.017) = \log(90.017) = 4.5$. The natural log of an integer will always be a positive number, and the log of a decimal less than 1 will always be negative. The log of a negative number is undefined.

There are a couple of special algebraic rules that will be used in discussing loglinear models. The log of a product is equal to the sum of the logs of each individual value or variable (product rule).

$$\log(xy) = \log(x) + \log(y)$$

Incidentally, this analysis is called log*linear*, because of this additive rule. The addition of terms on the right hand side of the equation implies a linear combination of the terms. In another rule, the log of a ratio is equal to the difference of the logs of each individual value or variable (quotient rule) in which the log of the numerator is subtracted from the log of the denominator.

$$\log\left(\frac{x}{y}\right) = \log(x) - \log(y)$$

We may also see the power rule,

$$\log\left(x^y\right) = y\log(x)$$

Odds, Log-Odds and Odds Ratios Revisited

Recall that the odds of an event occurring involves the ratio of the frequency of the event occurring to the frequency of the event not occurring (or the alternative occurring). Using the Quinnipiac poll data[1] from the "Analysis of Contingency Tables" handout, we had the following frequencies

	Trump	Biden	
Party affiliated	338	363	701
Independent	125	156	281
	463	519	982

[1] These results are based on a national Quinnipiac University poll from Oct 4-7, 2019, https://poll.qu.edu/national/release-detail?ReleaseID=3643. Methodological details are here https://poll.qu.edu/images/polling/us/us10082019_demos_uljv62.pdf/.

The odds that party affiliated voters favored Biden over Trump was $n_{12}/n_{11} = 363/338 = 1.07$. And we get the same result if we use the conditional proportions,[2] $p_{1|2}/p_{1|1} = (363/701)/(338/701) = .517/.482 = 1.07$ Loglinear models use log transformations of counts and proportions as the basis for the analysis, so it is useful to consider log transformation of the odds—the *log odds* or *logit*. The log odds for party affiliates favoring Biden over Trump was $\log(n_{12}/n_{11}) = \log(363/338) = \log(1.074) = .071$, which has an equivalent result if conditional proportions are used, $\log(p_{1|2}/p_{1|1}) = \log(.517/.482) = \log(1.074) = .071$. If the odds for independents favoring Biden are compared to the odds of party affiliates favoring Biden, we have the odds ratio, of course,

$$\theta = \frac{156/125}{363/338} = \frac{1.248}{1.074} = 1.162$$

which means that the odds that a respondent supported Biden if the voter was independent was 1.162 times the odds of supporting Biden if the voter was party affiliated. The log of the odds ratio, $\log(\theta) = \log(1.162) = .150$ is the logistic regression coefficient, β. Note that we could also arrive at this value by using the quotient rule (within rounding), which is the difference in logits,

$$\log\left(\frac{156/125}{363/338}\right) = \log(156/125) - \log(363/338) = .222 - .071 = .151$$

Using the exponential function of β, we get the odds ratio again, $e^{\beta} = e^{.151} = 1.162$. If the log of the odds ratio is 0, then the odds ratio is one, because $e^0 = 1.0$.

The Loglinear Model

We will begin by assuming nominal categories, but later we will see how the loglinear model can be extended to ordinal variables. Your text introduces a new notation for expected frequencies here. Instead of using E_{ij} for the expected count in a cell, the Greek letter "mu" is used, μ_{ij}. One way to express the computation for the expected frequencies for the Pearson χ^2 would be $\mu_{ij} = n\pi_{i+}\pi_{+j}$. The loglinear model expresses everything in terms of natural logs, so the log of the expected frequency for one cell is $\log(\mu_{ij}) = \log(n\pi_{i+}\pi_{+j})$. Using the product rule, we can see how one basic loglinear model, the independence model, partitions the expected frequencies for a cell into three components.

$$\log(\mu_{ij}) = \log(n) + \ln(\pi_{i+}) + \ln(\pi_{+j})$$

Above, μ_{ij} is the expected count for one cell, n is the total sample size, π_{i+} is the corresponding marginal row proportion, and π_{+j} is the corresponding column marginal proportion. The equation is true if X and Y are independent. In practice, we would use the observed marginal frequencies, p_{i+} and p_{+j}, to compute the expected frequency of a cell rather than known or population values as suggested by the use of π. Each natural log term is often re-expressed as a set of *parameters*, using the Greek symbol λ ("lambda").

$$\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y$$

The superscript is not an exponent, but just denotes that the parameter pertains to the X (row) or the Y (column) variable in the contingency table. Each one of the subscripts, i and j, refer to a particular row or column. Each lambda represents each log term in the previous equation, $\lambda = \log(n)$, $\lambda_i^X = \log(\pi_{i+})$, and $\lambda_j^Y = \log(\pi_{j+})$.

[2] Using the cell/total proportions, n_{ij}/n_{++}, produces the same number as long a common denominator is used.

As an example, we can compute the log of the expected frequency for the first cell in the Quinnipiac poll table,

$$\log(\mu_{ij}) = \log(n) + \log(\pi_{i+}) + \log(\pi_{+j})$$
$$\log(982) + \log(701/982) + \log(463/982)$$
$$= 6.890 + (-.337) + (-.752) = 5.801$$

This value does not mean too much by itself, but if we use the exponent function to undo the log, the $e^{5.801} = 330.630$. Within rounding, this value happens to be equal to the expected value using the expected value formula for the Pearson χ^2, which before we computed as $E_{ij} = (701*463)/982 = 330.512$.

Two-Way Contingency Table Tests

The independence model holds true under the null hypothesis that the conditional proportions in each row are equal (i.e., that the odds ratio is 1.0 and the two variables are independent). If the null hypothesis is not true, then there must be something added to the right hand side of the equation that would produce a value larger or smaller than $\log(\mu_{ij})$. The extra term (the *association term*)[3] on the righthand side of the equation, λ_{ij}^{XY}, represents the degree of non-independence, which quantifies the departure of the observed frequencies from the expected frequencies.

$$\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}$$

Adding this last term to the model produces the so-called *saturated model*, in which there is the same number of parameters as there are observed frequencies (i.e., the fit is perfect). Testing hypotheses in loglinear modeling involves the comparison of a model with more parameters to a model with fewer parameters. Although one could construct other tests by comparing the saturated model to a model that drops another parameter (e.g., λ_i^2), nearly always the interest is with the comparison of the saturated and the independence model in 2 × 2 contingency tables. Although a Pearson χ^2 could also be used, loglinear models are most often tested with the likelihood ratio test with $df = (I\text{-}1)(J\text{-}1)$.

$$G^2 = 2\sum_i^I \sum_j^J n_{ij} \log\left(\frac{n_{ij}}{\mu_{ij}}\right)$$

The loglinear model for the 2 × 2 contingency table is easily extended to $I \times J$ tables. For these larger designs, the table is usually divided up into a set of 2 × 2 subtables. For comparisons within larger tables, odds need to be defined as a comparison to a referent group. The odds ratios in the 2 × 2 case were based on odds of n_{i1} relative to n_{i2}. If there were three columns, we could choose any column as a referent. Usually, the referent is the last column (or sometimes the first). For example, with 3 × 3 table, we could compute the log odds ratio comparing the third and the first columns from the third and the second rows, $\log\left[(n_{33}/n_{31})/(n_{23}/n_{21})\right]$. In general, there are J -1 non-redundant comparisons needed to capture all of the differences. To identify the model, we are placing scaling (or normalization) constraints on the parameters that are usually in the form of dummy codes, just as with g -1 dummy codes used for categorical predictors in a regression model.[4]

Technically the independence model in the 2 × 2 case has parameters for each row and each column, $\log(\mu_{ij}) = \lambda + \lambda_1^X + \lambda_2^X + \lambda_1^Y + \lambda_2^Y$ and the referent row (e.g., the last) is removed (set to 0), so the

[3] This term is also commonly referred to as the interaction term, because the value of Y depends on the value of X. I think the analogy to the factorial ANOVA is potentially confusing, however, because we have a circumstance with only two variables rather than three variables.
[4] Alternatively, the sum of the parameters for cell pairs could be constrained to be 0, which is an effect coding scaling approach. The dummy coding approach seems to be the most common.

independence model for the 2 × 2 case is just $\log(\mu_{ij}) = \lambda + \lambda_1^X + \lambda_1^Y$. In the 3 × 3 case the independence model is $\log(\mu_{ij}) = \lambda + \lambda_1^X + \lambda_2^X + \lambda_3^X + \lambda_1^Y + \lambda_2^Y + \lambda_3^Y$, but a chosen referent category is dropped, so it becomes $\log(\mu_{ij}) = \lambda + \lambda_1^X + \lambda_2^X + \lambda_1^Y + \lambda_2^Y$. The saturated models for each case includes $(I\text{-}1)(J\text{-}1)$ association terms that are also dropped.

Loglinear models produce coefficients for one of the effects in the model. The λ^{XY} association term provides information about the X-Y relationship and is equal to a logistic regression with Y regressed on X. An odds ratio can be derived from any of the coefficients by using the exponential transformation, where $e^{b_{XY}}$ = OR for X predicting Y.

Three-Way Contingency Table Tests
The loglinear concepts presented above for two-way tables can be expanded to three-way tables, whether for the simpler 2 × 2 × 2 case or $I \times J \times K$ tables. The saturated loglinear model for three-way tables is

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{jk}^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}$$

Superscripts X, Y, and Z represent the three categorical variables, with their levels indexed by i, j, and k subscripts, respectively. The loglinear test of the homogeneous association hypothesis, which we previously analyzed with the Breslow-Day test, drops the three-way association term, λ_{ijk}^{XYZ}. The mutual independence model, analyzed earlier with the Cochran-Mantel-Haenszel test, would drop each of the two-way terms and the three-way term. The conditional independence model examining $X \times Y$ association within the Z strata, as with the test of the Mantel-Haenszel odds ratio, can be investigated by dropping the three-way term and the XY term, λ_{ij}^{XY}, because the null hypothesis assumes no three-way dependence (the $X \times Y$ association does not depend on Z) and there is no XY association at any level of Z. Of course, other types of hypotheses can be tested, but a hierarchical approach, which does not include higher order terms without all of the lower order terms is usually not recommended. Tests are usually based on the likelihood ratio test G^2 as given above, except that the particular three-way loglinear model being tested is used to generate the expected frequencies and summation occurs across the three dimensions. Conditional odds ratios for partial tables can be computed similar to the odds ratios described for two-way tables. Theus and Lauer (1999) discuss mosaic type plots for visualizing loglinear hypotheses.

Matched Pairs
Loglinear models can be applied to the matched pair design to test for marginal homogeneity (ala the McNemar test), for example. Recall that marginal homogeneity states the marginal proportions, p_{i+} and p_{j+} are equal, providing information about whether there is an increase or decrease in the proportion of one category of response (e.g., "yes") over time. For larger $I \times I$ square tables, the loglinear model is modified to specify symmetry and quasi-symmetry hypotheses in order to derive a test of marginal homogeneity. I use a and b subscripts here because of the special nature of the contingency table for matched pairs.

Symmetry
$$\log(\mu_{ab}) = \lambda + \lambda_a + \lambda_b + \lambda_{ab}$$

where $\mu_{ab} = n\pi_{ab}$ and the parameters above and below the diagonal are equal if the row and column numbers are switched, such that $\lambda_{ab} = \lambda_{ba}$. For example, if a question with three response options, "yes," "no," "maybe," are repeated, the test symmetry tests is whether there is the same log odds of "yes"-"maybe" as "maybe"-"yes" going from Time 1 and then Time 2. The lack of superscript indicates that the specific marginal effect terms cannot differ (i.e., the log of the marginal proportions must be the same for Time 1 and Time 2 responses).

Quasi-Symmetry

$$\log(\mu_{ab}) = \lambda + \lambda_a^X + \lambda_b^Y + \lambda_{ab}$$

The parameters are defined similarly to the symmetry model, but the quasi-symmetry only requires that odds ratios above and below the diagonal are equal, $\theta_{ab} = \theta_{ba}$, which allows the marginal proportions to differ. Marginal homogeneity can then be computed from the these two G^2 values, where the differences is also chi-square distributed.

$$G^2_{\text{marginal homogeneity}} = G^2_{\text{quasi-symmetry}} - G^2_{\text{symmetry}}$$

There are several specific hypotheses beyond these tests that can be conducted in which coding schemes are set up for $I \times J$ tables to investigate whether certain patterns of change occur. The levels or topological schemes are discussed by Hauser (1980).

Software Examples

The examples below reanalyze the 2 × 2 contingency table from the Quinnipiac poll data. Look back at the chi-square analyses in the prior handout for comparison.

SPSS

```
*Loglinear models.

*saturated model
loglinear ind(0,1) response(0,1)
  /criteria=delta(0)
  /print=default estim
  /design=ind response ind by response.
```

```
*** ML converged at iteration 2.
    Maximum difference between successive iterations =   .00000.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

  Observed, Expected Frequencies and Residuals

      Factor          Code              OBS. count & PCT.   EXP. count & PCT.      Residual   Std. Resid.   Adj. Resid.


  ind             affiliat
   response        Trump                    338.00 (34.42)      338.00 (34.42)         .0000         .0000         .0000
   response        Biden                    363.00 (36.97)      363.00 (36.97)         .0000         .0000         .0000

  ind             independ
   response        Trump                    125.00 (12.73)      125.00 (12.73)         .0000         .0000         .0000
   response        Biden                    156.00 (15.89)      156.00 (15.89)         .0000         .0000         .0000

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

 Goodness-of-Fit test statistics

    Likelihood Ratio Chi Square =      .00000    DF = 0  P =  .
             Pearson Chi Square =      .00000    DF = 0  P =  .

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 Estimates for Parameters

 ind

  Parameter         Coeff.        Std. Err.         Z-Value     Lower 95 CI     Upper 95 CI

        1      .4598197463          .03546        12.96565          .39031          .52933

 response

  Parameter         Coeff.        Std. Err.         Z-Value     Lower 95 CI     Upper 95 CI

        2     -.0732248022          .03546        -2.06474         -.14274         -.00371
```

```
 ind by response

  Parameter         Coeff.        Std. Err.         Z-Value      Lower 95 CI     Upper 95 CI

        3       .0375463328          .03546         1.05870         -.03196          .10706
  *** ML converged at iteration 2.

*independence model: model without the assocation term (ind by response).loglinear ind(0,1) response(0,1)
  /criteria=delta(0)
  /print=default estim
  /design=ind response.
```

(only the fit information is needed from the independence model so that is all that is included below)

```
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

 Goodness-of-Fit test statistics

    Likelihood Ratio Chi Square =     1.12346    DF = 1  P =  .289
             Pearson Chi Square =     1.12168    DF = 1  P =  .290

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

Note that the difference in fit of the two models is the same as the fit of the second independence model $LR\chi^2_{independence} - LR\chi^2_{saturated} = 1.2346 - 0 = 1.2346$ and that the odds ratio for the ind × response relationship is $e^{.038}$ = 1.46.

R

I use the MASS package function loglm, which requires a 2 x 2 table of counts and dimension names. It would be possible to compare two models using `anova(logm1,logm2)` but it does not provide any additional information here, so I omitted it.

```
> #library(MASS)
> tbl = table(mydata$ind, mydata$response)
> tbl
                    0   1
  party affiliate 491 629
  independent      63  48
> #dimnames assigns labels for categories
> dimnames(tbl) = list(ind = c("Party","Independent"),
+ response = c("Trump","Biden")
+ )
>
> #Loglinear models
> #two-way loglinear
> library(MASS)
> #independence model
> logm1 <- loglm( ~ ind + response, data=tbl)
> summary(logm1)

Statistics:
                      X^2 df  P(> X^2)
Likelihood Ratio 1.123457  1 0.2891754
Pearson          1.121677  1 0.2895576

> #obtain coefficients—use first column values and upper left quadrant from ind x response matrix
> coef(logm2)
$`(Intercept)`
[1] 5.39890
$ind
      Party Independent
  0.4598197  -0.4598197

$response
     Trump      Biden
```

```
-0.0732248  0.0732248

$ind.response
             response
ind                 Trump       Biden
  Party        0.03754633 -0.03754633
  Independent -0.03754633  0.03754633
```

SAS

The parameters for the saturated model appear in the first section (with the Wald test of the interaction) and the likelihood ratio test is conducted comparing to the independence model by default in the following section.

```
*the genmod procedure wants counts (frequencies as input) so they are read out into the data two data set;
proc freq data=one;
tables ind*response /out=two;
run;

proc genmod data=two;
class ind response;
model count = ind response ind*response / dist=poi link=log lrci type3 obstats;
run;
* type3 uses Type III sum of squares, which, as in ANOVA, uses SS for all effects partialing out all other effects;
* I changed the linesize in the options at the top, ls=240 to get all of the obstats in one line without wrapping;
```

The GENMOD Procedure

Analysis Of Maximum Likelihood Parameter Estimates

Parameter			DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept			1	4.8283	0.0894	4.6477	4.9986	2914.08	<.0001
ind	affiliate		1	0.9947	0.1047	0.7926	1.2033	90.29	<.0001
ind	independent		0	0.0000	0.0000	0.0000	0.0000	.	.
response	Biden		1	0.2215	0.1200	-0.0130	0.4581	3.41	0.0650
response	Trump		0	0.0000	0.0000	0.0000	0.0000	.	.
ind*response	affiliate	Biden	1	-0.1502	0.1419	-0.4291	0.1273	1.12	0.2897
ind*response	affiliate	Trump	0	0.0000	0.0000	0.0000	0.0000	.	.
ind*response	independent	Biden	0	0.0000	0.0000	0.0000	0.0000	.	.
ind*response	independent	Trump	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale			0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
ind	1	186.64	<.0001
response	1	4.28	0.0385
ind*response	1	1.12	0.2892

Sample Write-Up

A loglinear model was used to test the association between identification as a political independent and candidate choice. A larger proportion of independents, 156 out of 519 (30.1%), supported Biden, whereas a smaller proportion of independents, 125 out of 463 (27.0%), supported Trump. The odds ratio was 1.46 indicating a 46% increase in likelihood of supporting Biden if the voter was an independent. The difference was not significant, however, as indicted by the likelihood ratio test of the independence model, $G^2(1) = 1.12$, $p = .29$.

(Note: you could also report parameter estimate for the association term, the Wald chi-squared, and the confidence intervals if you like. And, as with the chi-square analysis, the row per cents could be reported, describing the percentage of party affiliates vs. the percentage of independence who supported Biden, for instance).

Three-Way Loglinear Examples

This three-way example comes from the Quinnipiac poll analyzed in the previous handout "Three-Way Contingency Tables" using chi-squared analyses.[5] I use excerpts from the outputs to save space.

SPSS

```
* without three-way association variable.
loglinear state(0,1) ind(0,1) response(0,1)
  /criteria=delta(0)
  /print=default estim
  /design=state ind response state by ind state by response ind by response.

*saturated.
loglinear state(0,1) ind(0,1) response(0,1)
  /criteria=delta(0)
  /print=default estim
  /design=state ind response state by ind state by response ind by response state by ind by response.
```

From the non-saturated model (1-ways and all 2-ways)

```
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -


 Goodness-of-Fit test statistics

    Likelihood Ratio Chi Square =      .02913    DF = 1   P =  .864
             Pearson Chi Square =      .02913    DF = 1   P =  .864


- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

From the saturated model

```
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -


 Goodness-of-Fit test statistics

    Likelihood Ratio Chi Square =      .00000    DF = 0   P =  .
             Pearson Chi Square =      .00000    DF = 0   P =  .


- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -


 Estimates for Parameters

 state

  Parameter         Coeff.        Std. Err.         Z-Value      Lower 95 CI     Upper 95 CI

        1       .0838086485          .02557         3.27769           .03369          .13392

 ind

  Parameter         Coeff.        Std. Err.         Z-Value      Lower 95 CI     Upper 95 CI

        2       .4755168410          .02557        18.59709           .42540          .52563

 response

  Parameter         Coeff.        Std. Err.         Z-Value      Lower 95 CI     Upper 95 CI

        3      -.1246547375          .02557        -4.87515          -.17477         -.07454


state by ind

  Parameter         Coeff.        Std. Err.         Z-Value      Lower 95 CI     Upper 95 CI
```

[5] Data source: https://poll.qu.edu/georgia/release-detail?ReleaseID=3679. Note that the data extrapolated cell sample sizes and used some rounding, so the results should be taken as only approximate.

```
       4      -.0782703761          .02557       -3.06109         -.12839         -.02815

 state by response

  Parameter         Coeff.         Std. Err.         Z-Value     Lower 95 CI     Upper 95 CI

       5      -.0336785500          .02557       -1.31714         -.08379          .01644

 ind by response

  Parameter         Coeff.         Std. Err.         Z-Value     Lower 95 CI     Upper 95 CI

       6       .0220384578          .02557         .86191         -.02808          .07215

 state by ind by response

  Parameter         Coeff.         Std. Err.         Z-Value     Lower 95 CI     Upper 95 CI

       7       .0043640075          .02557         .17067         -.04575          .05448
```

R

```
#dimnames assigns labels for categories
> dimnames(tbl) = list(ind = c("Party","Independent"),
+                      response = c("Trump","Biden"))

> counts <-array(
+   c(308,401,132,191,323,374,103,128),
+   dim=c(2, 2, 2),
+   dimnames=list(state=c("OH", "GA"),
+                 ind =c("party aff", "independent"),
+                 response  =c("Biden", "trump"))
+ )


> #three-way loglinear model
> library(MASS)
> #nonsaturated-no three-way
> logmodel <- loglm( ~ state + ind + response + state*ind + state*response + ind*response,
+                     digits=4, data=counts)
> summary(logmodel)

Statistics:
                        X^2 df  P(> X^2)
Likelihood Ratio 0.02912809  1 0.8644836
Pearson          0.02912958  1 0.8644802

> #saturated
> logmodel2 <- loglm( ~ state + ind + response + state*ind + state*response + ind*response +
state*ind*response,
+         digits=4, data=counts)
> summary(logmodel2)

Statistics:
                 X^2 df P(> X^2)
Likelihood Ratio   0  0        1
Pearson            0  0        1


> coef(logmodel2)
```

Use lefthand column values and upper-left quadrant values

```
$`(Intercept)`
[1] 5.380975

$state
        OH         GA
-0.1246547  0.1246547

$ind
  party aff independent
  0.4755168  -0.4755168

$response
      Biden       trump
 0.08380865 -0.08380865

$state.ind
```

```
     ind
state   party aff independent
   OH  0.02203846 -0.02203846
   GA -0.02203846  0.02203846

$state.response
     response
state       Biden       trump
   OH -0.03367855  0.03367855
   GA  0.03367855 -0.03367855

$ind.response
             response
ind                Biden       trump
  party aff   -0.07827038  0.07827038
  independent  0.07827038 -0.07827038

$state.ind.response
, , response = Biden

     ind
state    party aff  independent
   OH  0.004364007 -0.004364007
   GA -0.004364007  0.004364007

, , response = trump

     ind
state    party aff  independent
   OH -0.004364007  0.004364007
   GA  0.004364007 -0.004364007
```

SAS

```
*/save out data table of counts */;

proc freq data=one;
      tables state*ind*response /out=two ;
      title 'Test of 3-Way Table';
run;
* Loglinear */
*the genmod procedure wants counts (frequencies as input) so they are read out into
the data two data set;

proc genmod data=two;
class state ind response;
model count = state ind response state*ind ind*response state*response / dist=poi
link=log lrci type3 obstats;
run;

proc genmod data=two;
class state ind response;
model count = state ind response state*ind ind*response state*response
state*ind*response / dist=poi link=log lrci type3 obstats;
run;
```

Analysis Of Maximum Likelihood Parameter Estimates

Parameter				DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept				1	5.7301	0.0570	5.6163	5.8397	10112.9	<.0001
state	GA			1	0.0476	0.0796	-0.1085	0.2038	0.36	0.5505
state	OH			0	0.0000	0.0000	0.0000	0.0000	.	.
ind	independent			1	-0.8473	0.1040	-1.0542	-0.6460	66.34	<.0001
ind	party affiliate			0	0.0000	0.0000	0.0000	0.0000	.	.
response	Biden			1	0.2639	0.0758	0.1158	0.4129	12.13	0.0005
response	Trump			0	0.0000	0.0000	0.0000	0.0000	.	.
state*ind	GA	independent		1	-0.2956	0.1537	-0.5982	0.0049	3.70	0.0544
state*ind	GA	party affiliate		0	0.0000	0.0000	0.0000	0.0000	.	.
state*ind	OH	independent		0	0.0000	0.0000	0.0000	0.0000	.	.
state*ind	OH	party affiliate		0	0.0000	0.0000	0.0000	0.0000	.	.
ind*response	independent	Biden		1	0.1056	0.1362	-0.1606	0.3736	0.60	0.4381
ind*response	independent	Trump		0	0.0000	0.0000	0.0000	0.0000	.	.
ind*response	party affiliate	Biden		0	0.0000	0.0000	0.0000	0.0000	.	.
ind*response	party affiliate	Trump		0	0.0000	0.0000	0.0000	0.0000	.	.
state*response	GA	Biden		1	-0.1173	0.1073	-0.3277	0.0930	1.19	0.2744
state*response	GA	Trump		0	0.0000	0.0000	0.0000	0.0000	.	.
state*response	OH	Biden		0	0.0000	0.0000	0.0000	0.0000	.	.
state*response	OH	Trump		0	0.0000	0.0000	0.0000	0.0000	.	.
state*ind*response	GA	independent	Biden	1	-0.0349	0.2046	-0.4359	0.3662	0.03	0.8645
state*ind*response	GA	independent	Trump	0	0.0000	0.0000	0.0000	0.0000	.	.
state*ind*response	GA	party affiliate	Biden	0	0.0000	0.0000	0.0000	0.0000	.	.
state*ind*response	GA	party affiliate	Trump	0	0.0000	0.0000	0.0000	0.0000	.	.
state*ind*response	OH	independent	Biden	0	0.0000	0.0000	0.0000	0.0000	.	.
state*ind*response	OH	independent	Trump	0	0.0000	0.0000	0.0000	0.0000	.	.
state*ind*response	OH	party affiliate	Biden	0	0.0000	0.0000	0.0000	0.0000	.	.
state*ind*response	OH	party affiliate	Trump	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale				0	1.0000	0.0000	1.0000	1.0000		

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
state	1	10.82	0.0010
ind	1	388.34	<.0001
response	1	23.98	<.0001
state*ind	1	9.43	0.0021
ind*response	1	0.74	0.3883
state*response	1	1.73	0.1879
state*ind*response	1	0.03	0.8645

Sample Write-Up

A loglinear model was used to investigate whether the association between identification as an independent (identified with a major party vs. independent) and candidate preference (Biden vs. Trump) was the same in Ohio and Georgia. Among Ohioans, 56.56% of respondents identifying with a major party favored Biden, whereas 59.13% of respondents identifying as independent favored Biden, OR = 1.05. Among Georgians, 53.66% of respondents identifying with a major party favored Biden, whereas 55.41% of respondents identifying with a major party favored Biden, OR = 1.03. The likelihood ratio test of the three-way association term indicated that the association between party identification and the candidate preference did not differ significantly between Ohioans and Georgians, $G^2(1) = .03$, p = .86.

(Notes: The odds ratios within states can be computed in the usual way as a ratio between two ratios. The results above focus on the homogenous association hypothesis, but several other loglinear tests, including the conditional independence of the ind × response association controlling for state and other types of odds ratios, such as the Mantel-Haenszel would be possible.)

References and Further Reading

Agresti, A. (2013). *Categorical data analysis, third edition*. New York: Wiley.

Hauser, R. M. (1980). Some exploratory methods for modeling mobility tables and other cross-classified data. *Sociological methodology*, *11*, 413-458.

Powers, D. A., Xie, Y., & Xie, Y. (2000). “Chapter 4: Loglinear models for statistical tables” in *Statistical methods for categorical data analysis* (Vol. 106). New York: Academic Press.

Theus, M., and Lauer, S.R.W. (1999). Visualizing loglinear models. *Journal of Computational and Graphical Statistics,* 8, 396-412.

Wickens, T.D. (1989). *Multiway contingency tables analysis for the social sciences*. Hillsdale, NJ: Erlbaum.