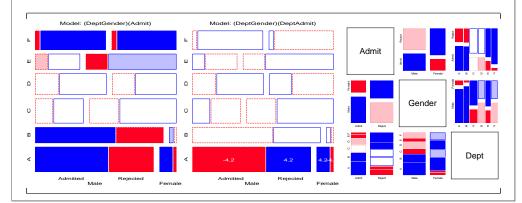
Loglinear Models and Mosaic Displays

Michael Friendly

Psych 6136

December 17, 2014



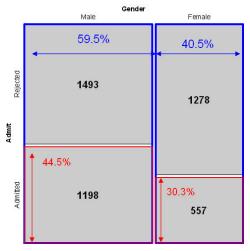
Mosaic displays: Basic ideas

Mosaic displays: Basic ideas

Hartigan and Kleiner (1981), Friendly (1994, 1999)

UCB Admissions: Observed frequencies

- Area-proportional display of frequencies in an *n*-way table
- Tiles (cells): recursive splits of a unit square—
 - V1: width \sim marginal frequencies, n_{i++}
 - V2: height ∼ relative frequencies $|V1, n_{ii+}/n_{i++}|$
 - V3: width ~ relative frequencies
- \Rightarrow area \sim cell frequency, n_{iik}

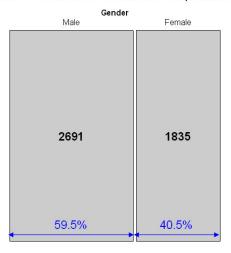


Mosaic displays: Basic ideas

Hartigan and Kleiner (1981), Friendly (1994, 1999)

UCB Admissions: Observed frequencies

- Area-proportional display of frequencies in an *n*-way table
- Tiles (cells): recursive splits of a unit square—
 - V1: width \sim marginal frequencies, n_{i++}
 - V2: height ~ relative frequencies
 - V3: width ∼ relative frequencies
- \Rightarrow area \sim cell frequency, n_{iik}



Mosaic displays: Basic ideas

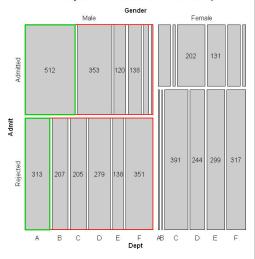
n-way tables

Mosaic displays: Basic ideas

Hartigan and Kleiner (1981), Friendly (1994, 1999)

3 way table: Gender x Admit x Dept

- Area-proportional display of frequencies in an *n*-way table
- Tiles (cells): recursive splits of a unit square—
 - V1: width \sim marginal frequencies, n_{i++}
 - V2: height ∼ relative frequencies $|V1, n_{ij+}/n_{i++}|$
 - V3: width ∼ relative frequencies $| (V1, V2), n_{ijk}/n_{ii+} |$
- \Rightarrow area \sim cell frequency, n_{iik}



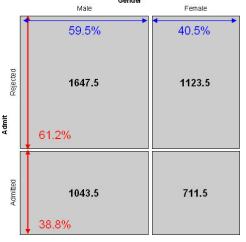
Mosaic displays: Basic ideas

- Independence: Two-way table
- Expected frequencies:

$$\widehat{m}_{ij} = \frac{n_{i+}n_{+j}}{n_{++}} = n_{++}$$
row %col %

 ⇒ rows & columns align when variables are independent





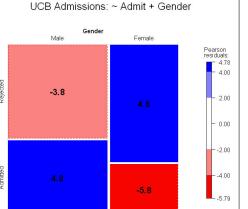
Independence: Expected frequencies

Pearson residuals:

$$d_{ij} = rac{n_{ij} - \widehat{m}_{ij}}{\sqrt{\widehat{m}_{ij}}}$$

Mosaic displays: Residuals & shading

- Pearson $\chi^2 = \Sigma \Sigma d_{ij}^2 = \Sigma \Sigma \frac{(n_{ij} \hat{m}_{ij})^2}{\hat{m}_{ii}}$
- Other residuals: deviance (LR), Freeman-Tukey (FT), adjusted (ADJ), ...
- Shading:
 - Sign: negative in red; + positive in blue
 - Magnitude: intensity of shading: $|d_{ii}| > 0, 2, 4, \dots$
- ⇒ Independence: rows align, or cells are empty!



Overview Loglinear models

Overview

Loglinear models

Loglinear models: Perspectives I

Loglinear approach

Loglinear models were first developed as an analog of classical ANOVA models, where multiplicative relations (under independence) are re-expressed in *additive* form as models for log(frequency).

$$\log m_{ij} = \mu + \lambda_i^{A} + \lambda_j^{B} \equiv [A][B] \equiv \sim A + B$$

- This expresses the model of independence for a two-way table (no A*B association)
- The notations $[A][B] \equiv \sim A + B$ are shorthands

Loglinear models: Perspectives II

GLM approach

More generally, loglinear models are also generalized linear models (GLMs) for log(frequency), with a Poisson distribution for the cell counts.

$$\log \boldsymbol{m} = \boldsymbol{X}\boldsymbol{\beta}$$

- This looks just like the general linear ANOVA, regression model, but for log frequency
- This approach allows quantitative predictors and special ways of treating ordinal factors

Loglinear models: Perspectives III

Logit models

When one table variable is a binary response, a logit model for that response is equivalent to a loglinear model (as discussed later).

$$\log(m_{1jk}/m_{2jk}) = \alpha + \beta_i^B + \beta_k^C \equiv [AB][AC][BC]$$

- $\log(m_{1ik}/m_{2ik})$ represents the log odds of response category 1 vs. 2
- The model formula includes only terms for the effects on A of variables B and C
- The equivalent loglinear model is [AB] [AC] [BC]
- The logit model assumes [BC] association, and [AB] $\rightarrow \beta_i^B$, [AC] $\rightarrow \beta_k^C$

Loglinear models: Overview

Two-way tables: Loglinear approach

For two discrete variables, A and B, suppose a multinomial sample of total size n over the IJ cells of a two-way $I \times J$ contingency table, with cell frequencies n_{ij} , and cell probabilities $\pi_{ij} = n_{ij}/n$.

• The table variables are statistically independent when the cell (joint) probability equals the product of the marginal probabilities, $Pr(A = i \& B = j) = Pr(A = i) \times Pr(B = j)$, or,

$$\pi_{ij} = \pi_{i+}\pi_{+j} .$$

• An equivalent model in terms of expected frequencies, $m_{ii} = n\pi_{ii}$ is

$$m_{ij} = (1/n) m_{i+} m_{+j}$$
.

 This multiplicative model can be expressed in additive form as a model for log m_{ii} ,

$$\log m_{ij} = -\log n + \log m_{i+} + \log m_{+j} . \tag{1}$$

Loglinear models

Loglinear models: Overview

Independence model

By anology with ANOVA models, the independence model (1) can be expressed as

$$\log m_{ij} = \mu + \lambda_i^A + \lambda_i^B , \qquad (2)$$

- μ is the grand mean of log m_{ii}
- the parameters λ_i^A and λ_j^B express the marginal frequencies of variables A and B "main effects"
- typically defined so that $\sum_i \lambda_i^A = \sum_i \lambda_i^B = 0$ as in ANOVA

Overview Loglinear models

Loglinear models: Overview

Saturated model

Dependence between the table variables is expressed by adding association parameters, λ_{ii}^{AB} , giving the *saturated model*,

$$\log m_{ij} = \mu + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB} \equiv [AB] \equiv \sim A * B .$$
 (3)

- The saturated model fits the table perfectly $(\hat{m}_{ii} = n_{ii})$: there are as many parameters as cell frequencies. Residual df = 0.
- A global test for association tests $H_0: \lambda_{ii}^{AB} = \mathbf{0}$.
- If reject H_0 , which $\lambda_{ii}^{AB} \neq 0$?
- For ordinal variables, the λ_{ii}^{AB} may be structured more simply, giving tests for ordinal association.

Example: Independence

Generate a table of Education by Party preference, strictly independent

```
educ \leftarrow c(50, 100, 50) # row marginal frequencies
names(educ) <- c("Low", "Med", "High")</pre>
party <- c(20, 50, 30) # col marginal frequencies
names(party) <- c("NDP", "Liberal", "Cons")</pre>
table <- outer(educ, party) / sum(party) # row x col / n
names(dimnames(table)) <- c("Education", "Party")</pre>
table
            Party
## Education NDP Liberal Cons
                      25 15
        Low 10
        Med 20
                      50 30
##
        High 10
                      25 15
```

Example: Independence

All row (and column) proportions are the same:

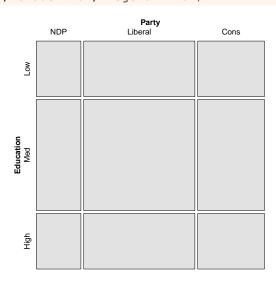
```
prop.table(table,1)
           Party
## Education NDP Liberal Cons
       Low 0.2
                    0.5 0.3
##
       Med 0.2
                    0.5 0.3
       High 0.2
##
                    0.5 0.3
```

All statistics are 0:

```
vcd::assocstats(table)
                 X^2 df P(> X^2)
## Likelihood Ratio 0 4
## Pearson
## Phi-Coefficient : 0
## Contingency Coeff.: 0
## Cramer's V : 0
```

Mosaic plot shows equal row and column proportions:

```
library(vcd)
mosaic(table, shade=TRUE, legend=FALSE)
```



Overview Loglinear models

Overview Loglinear models

Two-way tables: GLM approach

In the GLM approach, the vector of cell frequencies, $\mathbf{n} = \{n_{ij}\}$ is specified to have a Poisson distribution with means $\mathbf{m} = \{m_{ii}\}$ given by

$$\log \boldsymbol{m} = \boldsymbol{X}\boldsymbol{\beta}$$

- X is a known design (model) matrix, expressing the table factors
- β is a column vector containing the unknown λ parameters.
- This is the same as the familiar matrix formulation of ANOVA/regression, except that
 - The response, log *m* makes multiplicative relations additive
 - The distribution is taken as Poisson rather than Gaussian (normal)

Example: 2 x 2 table

For a 2 × 2 table, the saturated model (3) with the usual zero-sum constraints can be represented as

- only the linearly independent parameters are represented. $\lambda_2^A = -\lambda_1^A$, because $\lambda_1^A + \lambda_2^A = 0$, and so forth.
- association is represented by the parameter λ_{11}^{AB}
- can show that $\lambda_{11}^{AB} = \frac{1}{4} \log(\theta)$ (log odds ratio)
- Advantages of the GLM formulation: easier to express models with ordinal or quantitative variables, special terms, etc. Can also allow for over-dispersion.

Assessing goodness of fit

Goodness of fit of a specified model may be tested by the likelihood ratio G^2 ,

$$G^2 = 2\sum_i n_i \log \left(\frac{n_i}{\widehat{m}_i}\right) , \qquad (4)$$

or the Pearson X^2 .

$$X^2 = \sum_i \frac{(n_i - \widehat{m}_i)^2}{\widehat{m}_i} , \qquad (5)$$

with degrees of freedom df = # cells - # estimated parameters.

- E.g., for the model of independence, [A][B], df = |J - [(I - 1) - (J - 1)] = (I - 1)(J - 1)
- The terms summed in (4) and (5) are the squared *cell residuals*
- Other measures of balance goodness of fit against parsimony, e.g., Akaike's Information Criterion (smaller is better)

$$AIC = G^2 - 2df$$
 or $AIC = G^2 + 2$ # parameters

Overview Loglinear models

R functions for loglinear models

- chisq.test() and vcd::assocstats() only χ^2 tests for two-way tables, not a model (no parameters, no residuals)
- MASS::loglm() general loglinear models for *n*-way tables

loglm(formula, data, subset, na.action, ...)

- glm() all generalized linear models; loglinear with family=poisson glm(formula, family = poisson, data, weights, subset, ...)
- Formulas have the form: ~ A + B + ... (independence); ~ A*B + C (allow A*B association)
- Both return an R object, with named components use names (object)
- Both have print(), summary(), coef(), residuals(), plot() and other methods

Twoway tables

Example: Arthritis treatment

Data on effects of treatment for rheumatoid arthritis (in case form)

```
data(Arthritis, package="vcd")
str(Arthritis)
## 'data.frame': 84 obs. of 5 variables:
  $ ID : int 57 46 77 17 36 23 75 39 33 55 ...
## $ Treatment: Factor w/ 2 levels "Placebo", "Treated": 2 2 2 2 2 2 2 2
              : Factor w/ 2 levels "Female", "Male": 2 2 2 2 2 2 2 2 2
              : int 27 29 30 32 46 58 59 59 63 63 ...
## $ Improved : Ord.factor w/ 3 levels "None"<"Some"<...: 2 1 1 3 3 3 1 3
```

For now, examine the 2 × 3 table of Treatment and Improved

```
arth.tab <- with(Arthritis, table(Treatment, Improved))</pre>
arth.tab
           Improved
## Treatment None Some Marked
## Placebo 29 7
## Treated 13 7
```

Twoway tables

Example: Arthritis treatment

```
Fit the independence model, ~ Treatment + Improved
```

```
library (MASS)
(arth.mod <- loglm(~Treatment + Improved, data=arth.tab, fitted=TRUE))</pre>
## loglm(formula = Treatment + Improved, data = arth.tab, fitted = TRUE)
## Statistics:
##
                       X^2 df P(> X^2)
## Likelihood Ratio 13.530 2 0.0011536
## Pearson
                    13.055 2 0.0014626
residuals (arth.mod)
           Improved
                 None
## Treatment
                            Some
                                    Marked
    Placebo 1.534955 -0.062501 -2.152482
    Treated -1.776974 0.063501 1.837108
sum(residuals(arth.mod)^2) # Pearson chisquare
## [1] 13.53
```

1/68

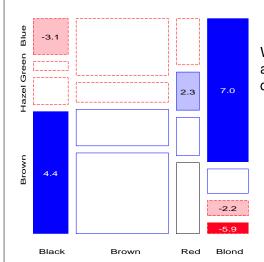
Twoway tables

Example: Hair color and eye color

```
haireye <- margin.table(HairEyeColor, 1:2)
(HE.mod <- loglm(~ Hair + Eye, data=haireye))
## Call:
## loglm(formula = ~Hair + Eye, data = haireye)
## Statistics:
                       X^2 df P(> X^2)
## Likelihood Ratio 146.44
## Pearson
                   138.29 9
round (residuals (HE.mod), 2)
## Re-fitting to get frequencies and fitted values
          Eye
          Brown Blue Hazel Green
    Black 4.00 -3.39 -0.49 -2.21
    Brown 1.21 -2.02 1.31 -0.35
         -0.08 -1.85 0.82 2.04
     Blond -7.33 6.17 -2.47 0.60
```

Twoway tables

Mosaic displays: Hair color and eye color



We know that hair color and eye color are associated ($\chi^2(9) = 138.29$). The question is how?

- Dark hair goes with dark eyes, light hair with light eyes
- Red hair, hazel eyes an exception?
- Effect ordering: Rows/cols permuted by CA Dimension 1
- ⇒ Opposite corner pattern

23/68

Three-way tables

Saturated model

For a 3-way table, of size $I \times J \times K$ for variables A, B, C, the saturated loglinear model includes associations between all pairs of variables, as well as a 3-way association term, λ_{iik}^{ABC}

$$\log m_{ijk} = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{ABC} + \lambda_{ik}^{ABC} + \lambda_{ik}^{ABC} + \lambda_{ik}^{ABC}$$
(6)

- One-way terms $(\lambda_i^A, \lambda_i^B, \lambda_k^C)$: differences in the *marginal frequencies* of the table variables
- Two-way terms $(\lambda_{ij}^{AB}, \lambda_{ik}^{AC}, \lambda_{jk}^{BC})$ pertain to the *partial association* for each pair of variables, *controlling* for the remaining variable.
- ullet The three-way term, λ^{ABC}_{ijk} allows the partial association between any pair of variables to vary over the categories of the third variable.
- Fits perfectly, but doesn't *explain* anything, so we hope for a simpler model!

Three-way tables: Reduced models

Reduced models

- Loglinear models are usually hierarchical: a high-order term, such as $\lambda_{iik}^{ABC} \rightarrow all$ low-order relatives are automatically included.
- Thus, a short-hand notation for a loglinear model lists only the high-order
- i.e., the saturated model (6) \equiv [ABC], and implies all two-way and one-way terms
- The usual goal is to fit the smallest model (fewest high-order terms) that is sufficient to explain/describe the observed frequencies.
- This is similar to ANOVA/regression models with all possible interactions

Three-way tables

Reduced models

Three-way tables: Reduced models

Reduced models

- For a 3-way table there are a variety of models between the mutual independence model, [A][B][C], and the saturated model, [ABC]
- Each such model has an independence interpretation: $A \perp B$ means an hypothesis that A is independent of B.

Table: Log-linear Models for Three-Way Tables

Model	Model symbol	Interpretation
Mutual independence	[A][B][C]	$A \perp B \perp C$
Joint independence	[AB][C]	$(A B) \perp C$
Conditional independence	[AC][BC]	$(A \perp B) \mid C$
All two-way associations	[AB][AC][BC]	homogeneous assoc.
Saturated model	[ABC]	ABC interaction

Three-way tables Reduced models

Three-way tables: Model types

• **Joint independence**: $(AB) \perp C$, allows A*B association, but asserts no A*C and B*C associations

$$[AB][C] \equiv \log m_{ijk} = \mu + \lambda_i^A + \lambda_i^B + \lambda_k^C + \lambda_{ij}^{AB}$$

Conditional independence: A ⊥ B, controlling for C

$$[AC][BC] \equiv \log m_{ijk} = \mu + \lambda_i^A + \lambda_i^B + \lambda_k^C + \lambda_{ik}^{AC} + \lambda_{ik}^{BC}$$

• Homogeneous association: All two-way, but each two-way is the same over the other factor

$$[AB][AC][BC] \equiv \log m_{ijk} = \mu + \lambda_i^A + \lambda_i^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{ik}^{BC}$$

Response vs. Association models

- In association models, the interest is just on *which* variabels are associated, and how
 - Hair-eye data: [Hair Eye]? [Hair Sex]? [Eye Sex]
 - \implies fit the homogeneous association model (or the saturated model)
 - Test the individual terms, delete those which are NS
- In response models, the interest is on which predictors are associated with the response
 - The minimal (null or baseline) model is the model of joint independence of the response (say, A) from all predictors, [A] [B C D ...]
 - Associations among the predictors are fitted exactly (not analyzed)
 - Similar to regression, where predictors can be arbitrarily correlated
 - e.g., Berkeley data: fit the baseline model [Admit] [Gender Dept]
 - lack-of-fit ⇒ associations [Admit Gender] and/or [Admit Dept]

Goodness of fit tests

As noted earlier, overall goodness of fit of a specified model may be tested by the likelihood ratio G^2 , or the Pearson X^2 ,

$$G^2 = 2\sum_i n_i \log\left(\frac{n_i}{\widehat{m}_i}\right) \qquad X^2 = \sum_i \frac{(n_i - \widehat{m}_i)^2}{\widehat{m}_i} ,$$

with residual degrees of freedom ν = # cells – # estimated parameters.

- These measure the lack of fit of a given model— a large value ⇒ a poor model
- Both are distributed as $\chi^2(\nu)$ (in large samples: all $\hat{m}_i > 5$)
- $\mathcal{E}(\chi^2) = \nu$, so G^2/ν (or X^2/ν) measures lack of fit per degree of freedom (overdispersion)
- But: how to compare or test competing models?

Three-way tables GOF & ANOVA tests

Nested models and ANOVA-type tests

Nested models

Two models, M_1 and M_2 are nested when one (say, M_2) is a special case of the other

- Model M_2 (with ν_2 df) fits a subset of the parameters of M_1 (with ν_1 df)
- M_2 is more restrictive cannot fit better than M_1 : $G^2(M_2) > G^2(M_1)$
- The least restrictive model is the saturated model [ABC...] with $G^2 = 0$ and $\nu = 0$

Therefore, we can test the difference in G^2 as a specific test of the added restrictions in M_2 compared to M_1 . This test has has a χ^2 distribution with df = $\nu_2 - \nu_1$.

$$\Delta G^{2} \equiv G^{2}(M_{2} | M_{1}) = G^{2}(M_{2}) - G^{2}(M_{1})$$

$$= 2 \sum_{i} n_{i} \log(\widehat{m}_{i1}/\widehat{m}_{i2})$$
(7)

Three-way tables GOF & ANOVA tests

Example: Berkeley admissions data

For the UC Berkeley data, with table variables [A]dmit, [D]ept and [G]ender the following models form a nested chain

$$[A][D][G] \subset [A][DG] \subset [AD][AG][DG] \subset [ADG]$$

Table: Hierarchical G^2 tests for loglinear models fit to the UC Berkeley data

Туре	LLM terms	G^2	df	$\Delta(G^2)$	$\Delta(df)$	$Pr(>\Delta(G^2))$
Mutual ind	[A][D][G]	2097.67	16			
Joint	[A][DG]	877.06	11	1220.62	5	0.0000
All 2-way	[AD][AG][DG]	20.20	5	1128.70	5	0.0000
Saturated	[ADG]	0.0	0	20.20	5	0.0011

- Only testing decrease in G^2 from one model to the next
- Here, each model is significantly better than the previous
- Joint vs. all 2-way: does Admit depend on Dept and/or Gender?
- Absolute fit of the all 2-way model is not terrible. Investigate further!

Fitting loglinear models

Fitting loglinear models in R

loglm() - data in contingency table form (MASS package)

```
data(UCBAdmissions)

## conditional independence (AD, DG) in Berkeley data

mod.1 <- loglm(~ (Admit + Gender) * Dept, data=UCBAdmissions)

## all two-way model (AD, DG, AG)

mod.2 <- loglm(~ (Admit + Gender + Dept)^2, data=UCBAdmissions)
```

glm() - data in frequency form

```
berkeley <- as.data.frame(UCBAdmissions)
mod.3 <- glm(Freq ~ (Admit + Gender) * Dept, data=berkeley,
family='poisson')</pre>
```

- loglm() simpler for nominal variables
- glm() allows a wider class of models and quantitative predictors (covariates)
- gnm () fits models for structured association and generalized non-linear models
- vcdExtra package provides visualizations for all.

Example: Berkeley admission data

Fit the model of mutual independence using loglm()

Fitting loglinear models

34/68

Fitting loglinear models

Example: Berkeley admission data

Fit other models with loglm()

Fitting loglinear models

Example: Berkeley admission data

Compare nested models with anova ()

```
anova(berk.loglm0, berk.loglm2, berk.loglm3, test="Chisq")
## LR tests for hierarchical log-linear models
## Model 1:
## ~Dept + Gender + Admit
## Model 2:
## ~Admit + (Dept * Gender)
## Model 3:
## ~ (Admit + Dept + Gender) ^2
##
            Deviance df Delta(Dev) Delta(df) P(> Delta(Dev)
## Model 1 2097.671 16
## Model 2 877.056 11 1220.615
                                                   0.00000
                                          6
## Model 3 20.204 5
                         856.852
                                                   0.00000
## Saturated 0.000 0
                           20.204
```

35/6

Fitting loglinear models

Example: Berkeley admission data

LRStats () in vcdExtra gives one line summaries of a collection of models

```
LRstats(berk.loglm0, berk.loglm1, berk.loglm2, berk.loglm3)
## Likelihood summary table:
              AIC BIC LR Chisq Df Pr(>Chisq)
## berk.log1m0 2273 2282
                        2098 16
                          1149 10
                                      <2e-16 ***
## berk.loglm1 1336 1352
## berk.loglm2 1062 1077
                        877 11
                        20 5
## berk.loglm3 217 240
                                      0.0011 **
## Signif. codes: 0 '*** 0.001 '** 0.01 '* 0.05 '.' 0.1 ' ' 1
```

- AIC and BIC are GOF measures adjusted for model parsimony
- Not not significance tests, but smaller is better
- Also apply to non-nested models

$$AIC = G^2 + 2 \times \#$$
 parameters
 $BIC = G^2 + 2 \log(n) \times \#$ parameters

Mosaic displays: Predictor variables

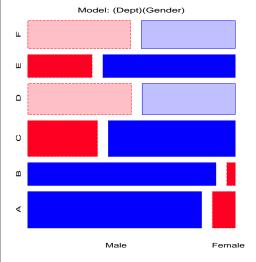
Berkeley data: Departments × Gender (ignoring Admit):

• Did departments differ in the total number of applicants?

Fitting loglinear models

• Did men and women apply differentially to departments?

Fitting loglinear models



- Model [Dept] [Gender]: $G_{(5)}^2 =$ 1220.6.
- Note: Departments ordered A-F by overall rate of admission.
- Men more likely to apply to departments A,B; women more likely in depts C-F

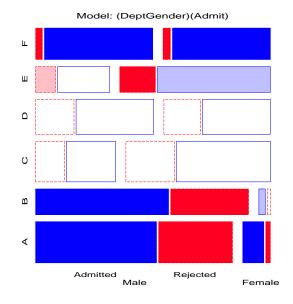
Fitting loglinear models Mosaic displays

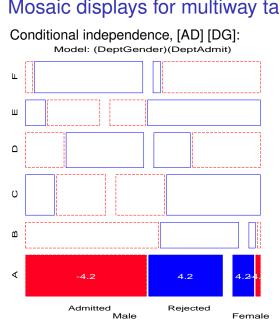
Mosaic displays for multiway tables

- Generalizes to *n*-way tables: divide cells recursively
- Can fit any log-linear model (e.g., 2-way, 3-way, ...),
 - For a 3-way table: [A][B][C], [AB][C], [AB][AC], ..., [ABC]
- Each mosaics shows:
 - DATA (size of tiles)
 - (some) marginal frequencies (spacing → visual grouping)
 - RESIDUALS (shading) what associations have been omitted?
- Visual fitting:
 - Pattern of lack-of-fit (residuals) → "better" model— smaller residuals
 - "cleaning the mosaic" → "better" model— empty cells
 - best done interactively!

Joint independence, [DG][A] (null model, Admit as response) [$G_{(11)}^2 = 877.1$]:

Mosaic displays





Mosaic displays for multiway tables

Fitting loglinear models

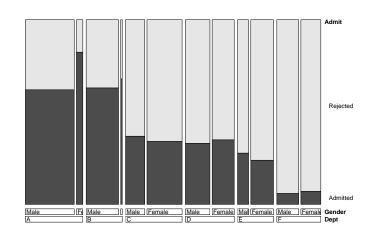
• E.g., Add [Admit Dept] association → Conditional independence:

- Fits poorly: $(G_{(6)}^2 = 21.74)$
- But, only in Department A!
- GLM approach allows fitting a special term for Dept. A
- Note: These displays use standardized residuals: better statistical properties.

Fitting loglinear models Mosaic displays

Other variations: Double decker plots

- Visualize dependence of one categorical (typically binary) variable on predictors
- Formally: mosaic plots with vertical splits for all predictor dimensions, highlighting the response by shading



Fitting loglinear models

4-way example

4-way example: Survival on the *Titanic*

Data on the fate of passengers and crew on the HMS *Titanic*, a $4 \times 2 \times 2 \times 2$ frequency table:

```
data (Titanic)
str(Titanic)
   table [1:4, 1:2, 1:2, 1:2] 0 0 35 0 0 0 17 0 118 154 ...
   - attr(*, "dimnames") = List of 4
    ..$ Class : chr [1:4] "1st" "2nd" "3rd" "Crew"
    ..$ Sex : chr [1:2] "Male" "Female"
    ..$ Age : chr [1:2] "Child" "Adult"
    ..$ Survived: chr [1:2] "No" "Yes"
```

What proportion survived? Ans: 711 / 2201 = 32.3 %

```
margin.table(Titanic, 4)
## Survived
     No Yes
## 1490
       711
```

Fitting loglinear models 4-way example

Zero cells

structable (Titanic)

Sex Male Female Survived No Yes No Yes ## Class Age ## 1st Child 0 0 4 140 ## Adult 118 57 ## 2nd Child 0 11 0 13 ## Adult 154 14 13 80 ## 3rd Child 35 13 17 14 ## Adult 387 75 89 76 ## Crew Child 0 0 0 0

Two types of zero cells:

Adult

- structural zeros could not occur (children in crew)
- sampling zeros did not happen to occur (children in 1st & 2nd who died)

670 192

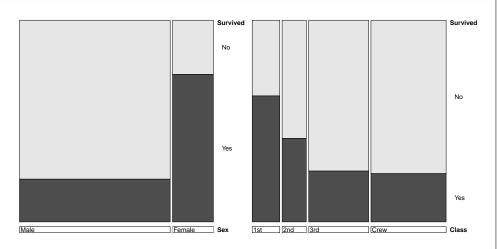
• Structural zeros can cause problems — loss of df; 0/0 = NaN in χ^2 tests

Exploratory plots

One-way doubledecker plots against survival show what might be expected:

Fitting loglinear models

doubledecker(Survived ~ Sex, data=Titanic)
doubledecker(Survived ~ Class, data=Titanic)



Exploratory plots
Two-way doubledecker plot against survival shows different effects of Class for men and women:

doubledecker (Survived ~ Sex + Class, data=Titanic)

Survived No Yes

Fitting loglinear models

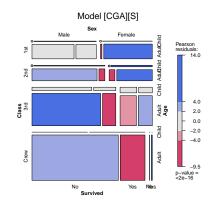
Fitting loglinear models

4-way example

Fitting and visualizing models

In the model formulas for <code>loglm()</code>, I use the variable numbers 1–4, and letters Class, Gender, Age and Survived

mutual independence [C][G][A][S]
mod0 <- loglm(~ 1 + 2 + 3 + 4, data=Titanic)
baseline (null) model
mod1 <- loglm(~ 1*2*3 + 4, data=Titanic)
mosaic(mod1, main="Model [CGA][S]")</pre>



- With S as response, the baseline model includes all associations among [CGA]
- But this model asserts no associations of these with survival
- $G^2(15) = 671.96$, a very poor fit

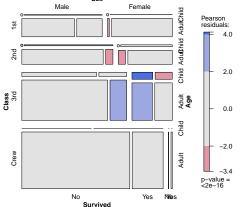
Adding associations

main effects of C, G, A on survival: [CGA][CS][GS][AS]
mod2 <- loglm(~ 1*2*3 + (1+2+3)*4, data=Titanic)
mosaic(mod2, main="Model [CGA][CS][GS][AS]")</pre>

Fitting loglinear models

4-way example

Model [CGA][CS][GS][AS]



- This model allows associations of each of C, G, A with Survived
- $G^2(10) = 112.57$, still not good
- Pattern of residuals suggests2-way interactions (3-way terms):
- "Women & children first": suggests a term [GAS]
- Allow interactions of Class with Gender [CGS] and Class with Age [CAS]

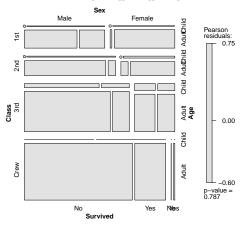
47/68

Fitting loglinear models 4-way example

Final model

mod3 <- loglm($\sim 1 \times 2 \times 3 + (1 \times 2) \times 4 + (1 \times 3) \times 4$, data=Titanic) mosaic(mod3, main="Model [CGA][CGS][CAS]")

Model [CGA][CGS][CAS]



Fitting loglinear models

4-way example

Model interpretation

- Regardless of Gender and Age, lower Class ⇒ decreased survival
- Differences in survival by Class were moderated by both Gender and Age
- term [CGS]: Women in 3rd class did not have an advantage, while men in 1st did vs. other classes
- term [CAS]: No children in 1st or 2nd class died, but nearly 2/3 in 3rd class did
- Summary:
 - Not so much "women and children first" as
 - "women and chilren, ordered by class and 1st class men"

Fitting loglinear models

Comparing models

As usual, anova () (and LRstats ()) give compact comparisons of a set of models

```
anova (mod0, mod1, mod2, mod3)
## LR tests for hierarchical log-linear models
## Model 1:
  ^{\sim}1 + 2 + 3 + 4
## Model 2:
  ~1 * 2 * 3 + 4
## Model 3:
   ^{\sim}1 * 2 * 3 + (1 + 2 + 3) * 4
## Model 4:
    ^{\sim}1 * 2 * 3 + (1 * 2) * 4 + (1 * 3) * 4
              Deviance df Delta(Dev) Delta(df) P(> Delta(Dev)
## Model 1 1243.6632 25
## Model 2
              671.9622 15
                            571.7010
                                                        0.00000
## Model 3
              112.5666 10
                            559.3956
                                              5
                                                        0.00000
## Model 4
                1.6855 4 110.8811
                                                        0.00000
## Saturated
                0.0000 0
```

Sequential plots and models

Sequential plots and models

- Mosaic for an n-way table \rightarrow hierarchical decomposition of association
- Joint cell probabilities are decomposed as

$$p_{ijk\ell \dots} = \underbrace{\overbrace{p_i \times p_{j|i} \times p_{k|ij}}_{\{v_1 v_2 v_3\}} \times p_{\ell|ijk} \times \cdots \times p_{n|ijk \dots}}_{\{v_1 v_2 v_3\}}$$

- First 2 terms \rightarrow mosaic for v_1 and v_2
- First 3 terms \rightarrow mosaic for v_1 , v_2 and v_3
- Roughly analogous to sequential fitting in regression: X_1 , $X_2|X_1$, $X_3|X_1X_2$,
- The order of variables matters for interpretation

Sequential plots and models

Sequential models of *joint independence* \to additive decomposition of the total association, $G^2_{[\nu_1][\nu_2]\dots[\nu_\rho]}$ (mutual independence),

$$G^2_{[\nu_1][\nu_2]\dots[\nu_\rho]} = G^2_{[\nu_1][\nu_2]} + G^2_{[\nu_1\nu_2][\nu_3]} + G^2_{[\nu_1\nu_2\nu_3][\nu_4]} + \dots + G^2_{[\nu_1\dots\nu_{\rho-1}][\nu_\rho]}$$

e.g., for Hair Eye color data

Model	Model symbol	df	G^2
Marginal	[Hair] [Eye]	9	146.44
Joint	[Hair, Eye] [Sex]	15	19.86
Mutual	[Hair] [Eye] [Sex]	24	166.30

Sequential plots and models: Example

• Hair color x Eye color marginal table (ignoring Sex)

(Hair)(Eye), G2 (9) = 146.44

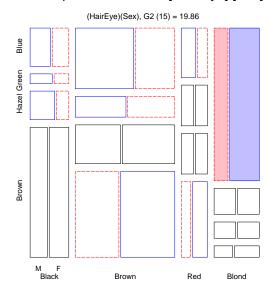
| Particular of the color of the colo

53/68

Sequential plots and models

Sequential plots and models: Example

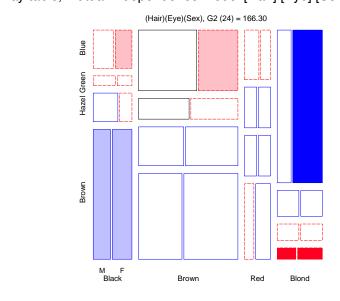
• 3-way table, Joint Independence Model [Hair Eye] [Sex]



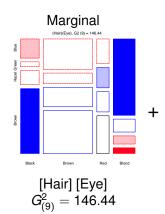
Sequential plots and models

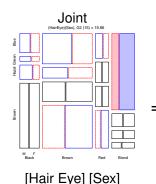
Sequential plots and models: Example

• 3-way table, Mutual Independence Model [Hair] [Eye] [Sex]

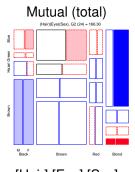


Sequential plots and models: Example





 $G_{(15)}^2 = 19.86$



[Hair] [Eye] [Sex] $G_{(24)}^2 = 166.30$

Response models

Applications

- When one variable, R_1 is a response and E_1, E_2, \ldots are explantory, the baseline model is the model of joint independence, $[E_1, E_2, \dots][R]$
- Sequential mosaics then show the associations among the predictors
- The last mosaic shows all associations with R
- Better-fitting models will need to add associations of the form $[E_iR], [E_iE_iR]...$

Causal models

Sometimes there is an assumed causal ordering of variables:

- Each path of arrows: $A \rightarrow B$, $A \rightarrow B \rightarrow C$ is a sequential model of joint independence: [A][B], [AB] [C], [ABC] [D].
- Testing these decomposes all joint probabilities

Sequential plots and models

Example: Marital status, pre- and extra-marital sex

• Order the table variables as $G \rightarrow P \rightarrow E \rightarrow M$

PreSex <- aperm(PreSex, 4:1) # order variables G, P, E, M

• Fit each sequential model to the marginal sub-table

mod.1 <- loglm(~ Gender + PremaritalSex, data=PreSex)</pre> mod.2 <- loglm(~ Gender * PremaritalSex + ExtramaritalSex, data=PreSex</pre>

Model	df	G^2
[G] [P]	1	75.259
[GP] [E]	3	48.929
[GPE] [M]	7	107.956
[G] [P] [E] [M]	11	232.142

Thornes and Collard (1979) studied divorce patterns in relation to premarital

Example: Marital status, pre- and extra-marital sex

and extramarital sex, a 24 table, PreSex in vcd

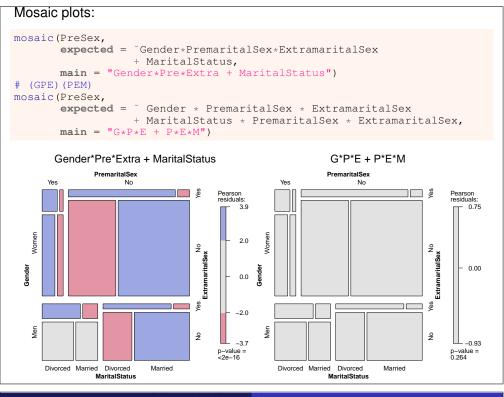
Sub-models:

• [G][P]: do men and women differ in pre-marital sex?

• [GP][E]: given G & P, are there differences in extra-marital sex?

• [GPE][M]: given G, P & E, are there differences in divorce?

Mosaic plots: # (Gender Pre) mosaic (margin.table (PreSex, 1:2), shade=TRUE, main = "Gender and Premarital Sex") # (Gender Pre) (Extra) mosaic (margin.table (PreSex, 1:3), expected = ~Gender * PremaritalSex + ExtramaritalSex, main = "Gender*Pre + ExtramaritalSex") Gender*Pre + ExtramaritalSex Gender and Premarital Sex PremaritalSex Pearson residuals: PremaritalSex Yes 4.0 S Women 4.0 2.0 2.0 0.0 0.0 -2.0 -2.0 -4.0 -4.6

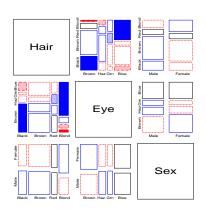


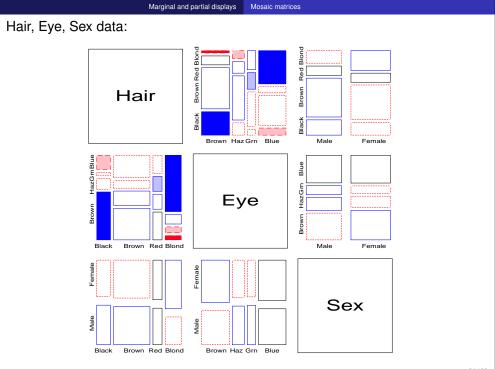
Marginal and partial displays

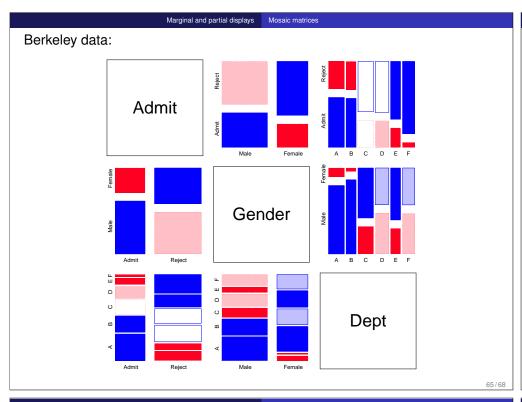
Mosaic matrices

Mosaic matrices

- Analog of *scatterplot matrix* for categorical data (Friendly, 1999)
 - Shows all p(p-1) pairwise views in a coherent display
 - Each pairwise mosaic shows bivariate (marginal) relation
 - Fit: marginal independence
 - Residuals: show marginal associations







Marginal and partial displays

tial displays Partial association

Partial association, Partial mosaics

Stratified analysis: conditional decomposition of G^2

- Fit models of partial (conditional) independence, $A \perp B \mid C_k$ at each level of (controlling for) C.
- \Rightarrow partial G^2 s add to the overall G^2 for conditional independence, $A \perp B \mid C$

$$G^2_{A\perp B\mid C}=\sum_k G^2_{A\perp B\mid C(k)}$$

Table: Partial and Overall conditional tests, $\textit{Hair} \perp \textit{Eye} \mid \textit{Sex}$

Model	df	G^2	<i>p</i> -value
[Hair][Eye] Male	9	44.445	0.000
[<i>Hair</i>][<i>Eye</i>] Female	9	112.233	0.000
[Hair][Eye] Sex	18	156.668	0.000

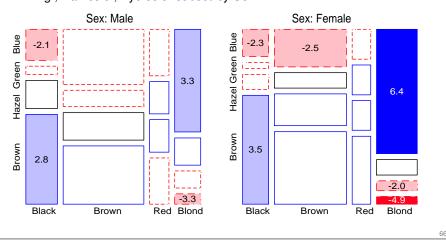
Marginal and partial displays

Partial association

Partial association, Partial mosaics

Stratified analysis:

- How does the association between two (or more) variables vary over levels of other variables?
- Mosaic plots for the main variables show partial association at each level of the other variables.
- E.g., Hair color, Eye color subset by Sex



Marginal and partial displays

Partial association

References I

Friendly, M. Mosaic displays for multi-way contingency tables. *Journal of the American Statistical Association*, 89:190–200, 1994.

Friendly, M. Extending mosaic displays: Marginal, conditional, and partial views of categorical data. *Journal of Computational and Graphical Statistics*, 8(3):373–395, 1999.

Hartigan, J. A. and Kleiner, B. Mosaics for contingency tables. In Eddy, W. F., editor, *Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface*, pp. 268–273. Springer-Verlag, New York, NY, 1981.

Thornes, B. and Collard, J. *Who Divorces?* Routledge & Kegan, London, 1979.

67/68