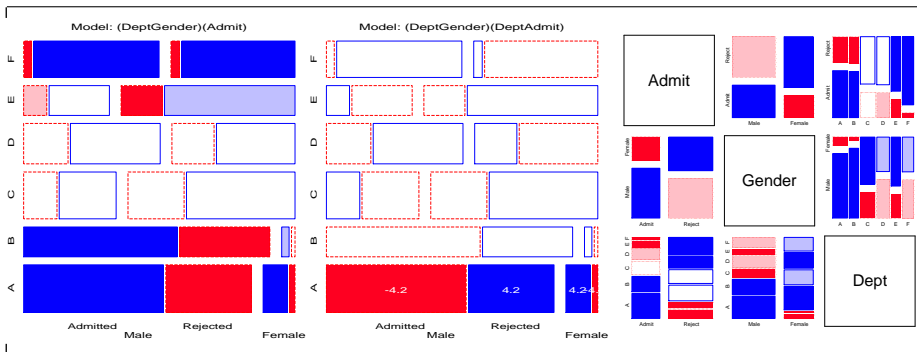


Loglinear Models and Mosaic Displays

Michael Friendly

Psych 6136

October 5, 2017

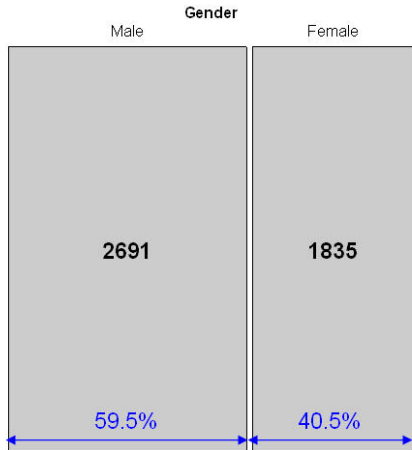


Mosaic displays: Basic ideas

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 \sim relative frequencies $|V1, n_{ij+}/n_{i++}$
 - V3: width \sim relative frequencies $| (V1, V2), n_{ijk}/n_{ij+}$
 - ...
- \Rightarrow area \sim cell frequency, n_{ijk}

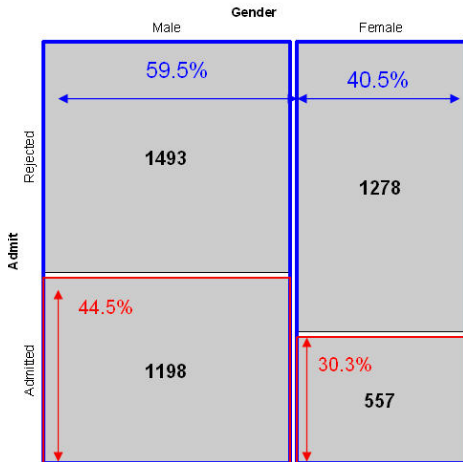


Mosaic displays: Basic ideas

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** \sim relative frequencies $|V1$, n_{ij+}/n_{i++}
 - V3: **width** \sim relative frequencies $|(V1, V2)$, n_{ijk}/n_{ij+}
 - ...
- \Rightarrow area \sim cell frequency, n_{ijk}

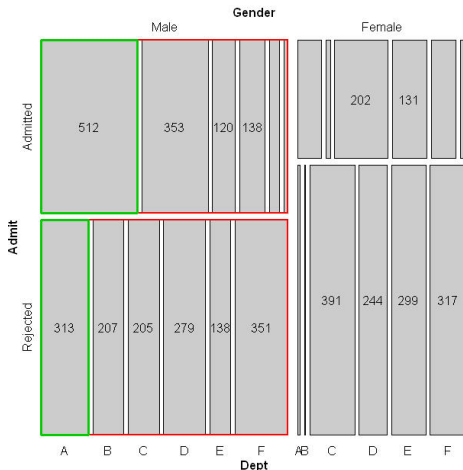


Mosaic displays: Basic ideas

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** \sim relative frequencies $|V1, n_{ij+}/n_{i++}$
 - V3: **width** \sim relative frequencies $|(V1, V2), n_{ijk}/n_{ij+}$
 - ...
- \Rightarrow area \sim cell frequency, n_{ijk}



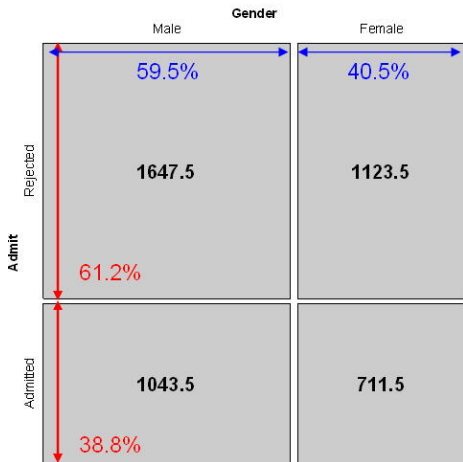
Mosaic displays: Basic ideas

- Independence: Two-way table
- Expected frequencies:

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

- \Rightarrow rows & columns align when variables are independent

Independence: Expected frequencies

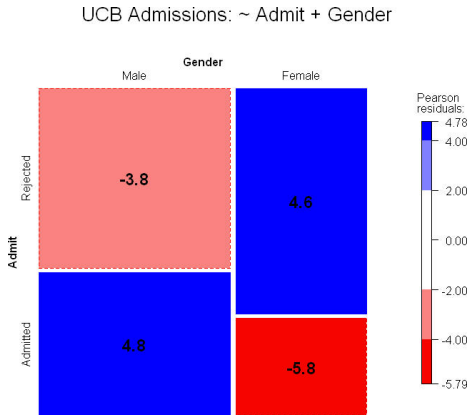


Mosaic displays: Residuals & shading

- Pearson residuals:

$$d_{ij} = \frac{n_{ij} - \hat{m}_{ij}}{\sqrt{\hat{m}_{ij}}}$$

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



Mosaic displays: Animation

- A 3×2 table, of answers to a question (Yes, ?, No), by sex.
- Marginal proportions of answers is fixed at (.40, .25, .35)
- Proportion of M, F is varied from frame to frame

Loglinear models: Perspectives

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 $\log(\text{frequency})$.

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

- 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
- Fit using **MASS: loglm()**

```
loglm(Freq ~ A + B + C, data=) loglm(Freq ~ A + B * C,  
data=)
```


Loglinear models: Perspectives

GLM approach

More generally, loglinear models are also **generalized linear models** (GLMs) for $\log(\text{frequency})$, with a **Poisson** distribution for the cell counts.

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

- 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**
- Fit using **glm()**, with `family=poisson` → a model for $\log(\text{Freq})$

```
glm(Freq ~ A + B + C, family = poisson)
glm(Freq ~ A + B * C, family = poisson)
```

Loglinear models: Perspectives

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).

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

- $\log(m_{1jk}/m_{2jk})$ 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 $[AB] \rightarrow \beta_j^B$, $[AC] \rightarrow \beta_k^C$
- Fit using

```
glm(outcome=="survived" ~ B + C, family=binomial
```

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, $m_{ij} = n\pi_{ij}$ is

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

- This multiplicative model can be expressed in additive form as a model for $\log m_{ij}$,

$$\log m_{ij} = -\log n + \log m_{i+} + \log m_{+j} . \quad (1)$$

Loglinear models: Overview

Independence model

By analogy with ANOVA models, the independence model (??) can be expressed as

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

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

Loglinear models: Overview

Saturated model

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

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

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

Example: Independence

Generate a table of Education by Party preference, strictly independent

```
educ <- c(50, 100, 50) # row marginal frequencies
names(educ) <- c("Low", "Med", "High")

party <- c(20, 50, 30) # col marginal frequencies
names(party) <- c("NDP", "Liberal", "Cons")

table <- outer(educ, party) / sum(party) # row x col / n
names(dimnames(table)) <- c("Education", "Party")
table
```

```
##           Party
## Education NDP  Liberal  Cons
##      Low    10      25    15
##      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
```

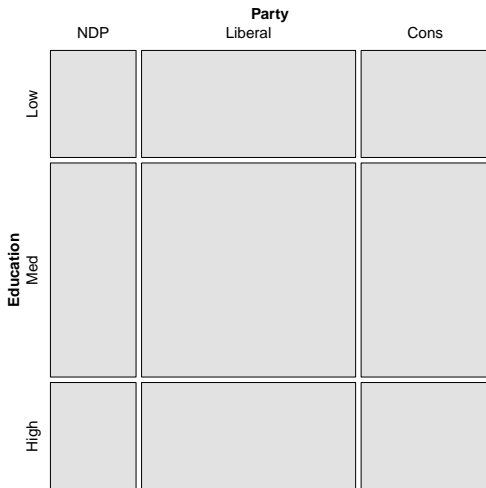
All statistics are 0:

```
vcd::assocstats(table)
```

```
##           X^2 df P(> X^2)
## Likelihood Ratio    0  4      1
## Pearson              0  4      1
##
## Phi-Coefficient      : NA
## Contingency Coeff.: 0
## Cramer's V           : 0
```

Mosaic plot shows equal row and column proportions:

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



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 $\mathbf{m} = \{m_{ij}\}$ given by

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

- \mathbf{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, $\log \mathbf{m}$ 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 (??) with the usual zero-sum constraints can be represented as

$$\log \begin{pmatrix} m_{11} \\ m_{12} \\ m_{21} \\ m_{22} \end{pmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{pmatrix} \mu \\ \lambda_1^A \\ \lambda_1^B \\ \lambda_{11}^{AB} \end{pmatrix}$$

- 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 ,

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

or the Pearson X^2 ,

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

with degrees of freedom $df = \# \text{ cells} - \# \text{ estimated parameters}$.

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

$$AIC = G^2 - 2df \text{ or } AIC = G^2 + 2 \# \text{ parameters}$$

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: $\sim A + B + \dots$ (independence); $\sim A*B + C$ (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

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 2 2 ...
## $ Sex      : Factor w/ 2 levels "Female","Male": 2 2 2 2 2 2 2 2 2 2 ...
## $ Age      : 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))
arth.tab

##           Improved
## Treatment None  Some  Marked
##   Placebo   29    7    7
##   Treated   13    7   21
```

Example: Arthritis treatment

Fit the independence model, \sim Treatment + Improved

```
library(MASS)
(arth.mod <- loglm(~ Treatment + Improved, data=arth.tab, fitted=TRUE))

## Call:
## 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

round(residuals(arth.mod), 3)

##           Improved
## Treatment  None   Some Marked
## Placebo    1.535 -0.063 -2.152
## Treated   -1.777  0.064  1.837

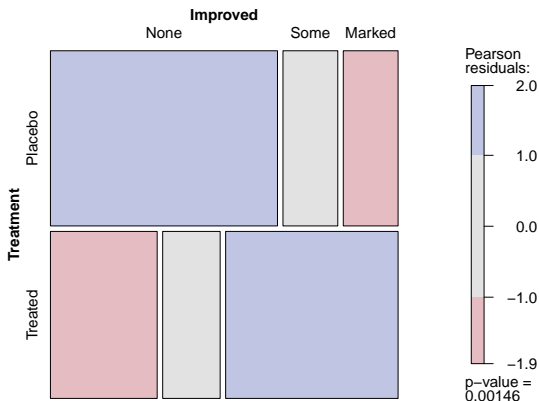
sum(residuals(arth.mod)^2) # Pearson chisquare

## [1] 13.53
```

Example: Arthritis treatment

Visualize association: `mosaic()` or `plot()` the model *or* table

```
mosaic(arth.mod, shade=TRUE, gp_args=list(interpolate=1:4))
```



Example: Hair color and eye color

```

haireye <- margin.table(HairEyeColor, 1:2)
(HE.mod <- loglm(~ Hair + Eye, data=haireye))

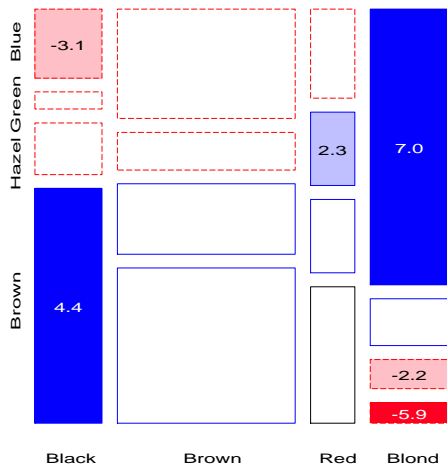
## Call:
## loglm(formula = ~Hair + Eye, data = haireye)
##
## Statistics:
##
##              X^2 df P(> X^2)
## Likelihood Ratio 146.44  9      0
## Pearson          138.29  9      0

round(residuals(HE.mod), 2)

## Re-fitting to get frequencies and fitted values
##
##      Eye
## Hair   Brown  Blue Hazel Green
## Black  4.00 -3.39 -0.49 -2.21
## Brown  1.21 -2.02  1.31 -0.35
## Red    -0.08 -1.85  0.82  2.04
## Blond -7.33  6.17 -2.47  0.60

```


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

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_{ijk}^{ABC}

$$\log m_{ijk} = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC} . \quad (6)$$

- One-way terms ($\lambda_i^A, \lambda_j^B, \lambda_k^C$): 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.
- The three-way term, λ_{ijk}^{ABC} 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 $\lambda_{ijk}^{ABC} \rightarrow$ **all** low-order relatives are automatically included.
- Thus, a short-hand notation for a loglinear model lists only the **high-order** terms,
- i.e., the saturated model (**??**) $\equiv [ABC]$, 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

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	$[AC][BC]$	$(A \perp B) C$
All two-way associations	$[AB][AC][BC]$	homogeneous assoc.
Saturated model	$[ABC]$	ABC interaction

Three-way tables: Model types

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

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

- **Conditional independence:** $A \perp B$, controlling for C

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

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

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

Collapsibility: Marginal and Conditional Associations

- Q: When can we legitimately collapse a table, ABC over some variable (C)?
- A: When the **marginal** association of AB is the same as the **conditional** association, $AB | C$
- Recall: Berkeley data
 - The marginal association of Admit, Gender ignoring Dept showed a strong association
 - The partial associations within Dept were mostly NS
 - This is an example of **Simpson's paradox**
- Three-way tables: The AB marginal and $AB | C$ conditional associations are the same, if either
 - A and C are conditionally independent, $A \perp C | B \equiv [AB][CB]$
 - B and C are conditionally independent, $B \perp C | A \equiv [AB][AC]$
 - \implies no three-way interaction

Response vs. Association models

- In **association models**, the interest is just on *which* variables 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 \implies 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 ,

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

with residual degrees of freedom $\nu = \# \text{ cells} - \# \text{ estimated parameters}$.

- These measure the **lack of fit** of a given model— a large value \implies a poor model
- Both are distributed as $\chi^2(\nu)$ (in large samples: all $\hat{m}_i > 5$)
- $\mathcal{E}(\chi^2(\nu)) = \nu$, 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?

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 : $G^2(M_2) \geq G^2(M_1)$
- 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 a χ^2 distribution with df = $\nu_2 - \nu_1$.

$$\begin{aligned} \Delta G^2 \equiv G^2(M_2 | M_1) &= G^2(M_2) - G^2(M_1) \\ &= 2 \sum n_i \log(\hat{m}_{i1} / \hat{m}_{i2}) \end{aligned} \quad (7)$$

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

Type	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 in R

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

```

1 data(UCBAdmissions)
2   ## conditional independence (AD, DG) in Berkeley data
3 mod.1 <- loglm(~ (Admit + Gender) * Dept, data=UCBAdmissions)
4   ## all two-way model (AD, DG, AG)
5 mod.2 <- loglm(~ (Admit + Gender + Dept)^2, data=UCBAdmissions)

```

`glm()` - data in frequency form

```

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

```

- `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()`

```
data("UCBAdmissions")
library(MASS)
berk.loglm0 <- loglm(~ Dept + Gender + Admit, data=UCBAdmissions)
berk.loglm0

## Call:
## loglm(formula = ~Dept + Gender + Admit, data = UCBAdmissions)
##
## Statistics:
##
##              X^2 df P(> X^2)
## Likelihood Ratio 2097.7 16      0
## Pearson          2000.3 16      0
```

Example: Berkeley admission data

Fit other models with `loglm()`

```
# conditional independence [AD] [AG]
berk.loglm1 <- loglm(~ Admit * (Dept + Gender), data=UCBAdmissions)

# joint independence [A] [DG]
berk.loglm2 <- loglm(~ Admit + (Dept * Gender), data=UCBAdmissions)
berk.loglm2

## Call:
## loglm(formula = ~Admit + (Dept * Gender), data = UCBAdmissions)
##
## Statistics:
##
##              X^2 df P(> X^2)
## Likelihood Ratio 877.06 11      0
## Pearson          797.70 11      0

# all two-way model [AD] [AG] [DG]
berk.loglm3 <- loglm(~ (Admit+Dept+Gender)^2, data=UCBAdmissions)
```

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         5         0.00000
## Model 3      20.204  5     856.852         6         0.00000
## Saturated    0.000  0      20.204         5         0.00114
```

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.loglm0 2273 2282   2098 16   <2e-16 ***
## berk.loglm1 1336 1352   1149 10   <2e-16 ***
## berk.loglm2 1062 1077   877 11   <2e-16 ***
## berk.loglm3  217  240    20  5   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

$$AIC = G^2 + 2 \times \# \text{ parameters}$$

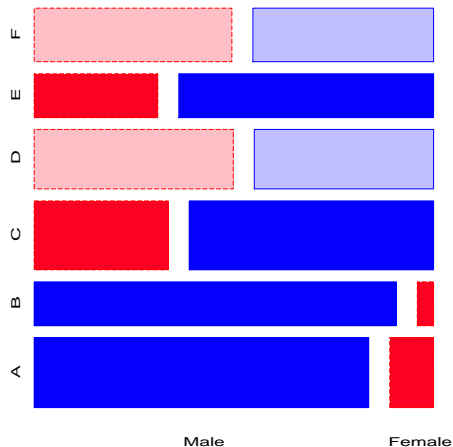
$$BIC = G^2 + 2 \log(n) \times \# \text{ parameters}$$

Mosaic displays: Predictor variables

Berkeley data: Departments \times Gender (ignoring Admit):

- Did departments differ in the total number of applicants?
- Did men and women apply differentially to departments?

Model: (Dept)(Gender)

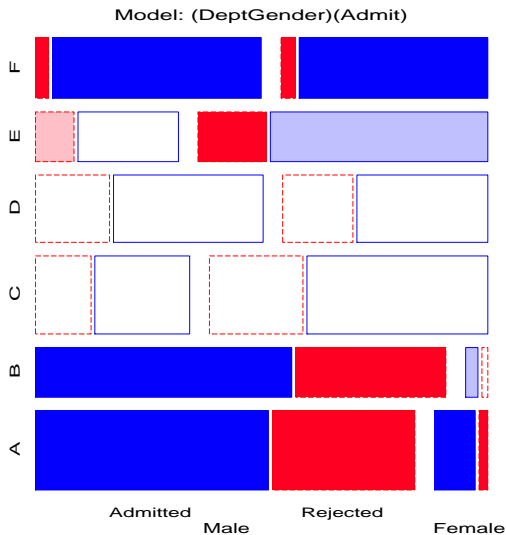


- Model [Dept] [Gender]: $G_{(5)}^2 = 1220.6$.
- *Note*: Departments ordered A–F by overall rate of admission.
- Men more likely to apply to departments A,B; women more likely in depts C–F

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 mosaic shows:
 - **DATA** (size of tiles)
 - (some) **marginal** frequencies (spacing \rightarrow visual grouping)
 - **RESIDUALS** (shading) — what associations have been omitted?
- Visual fitting:
 - Pattern of lack-of-fit (residuals) \rightarrow “better” model— smaller residuals
 - “cleaning the mosaic” \rightarrow “better” model— empty cells
 - best done interactively!

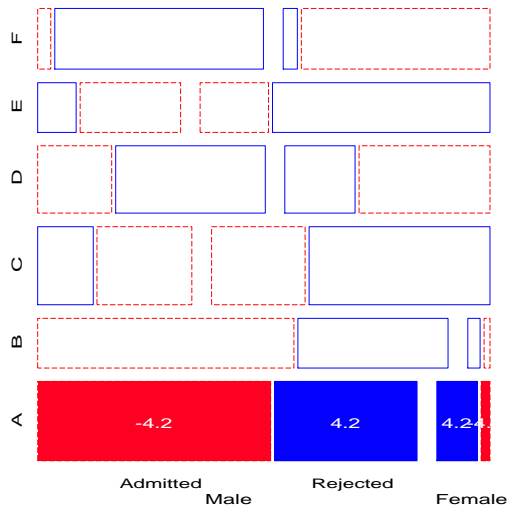
Joint independence, [DG][A] (null model, Admit as response) [$G^2_{(11)} = 877.1$]:



Mosaic displays for multiway tables

Conditional independence, [AD] [DG]:

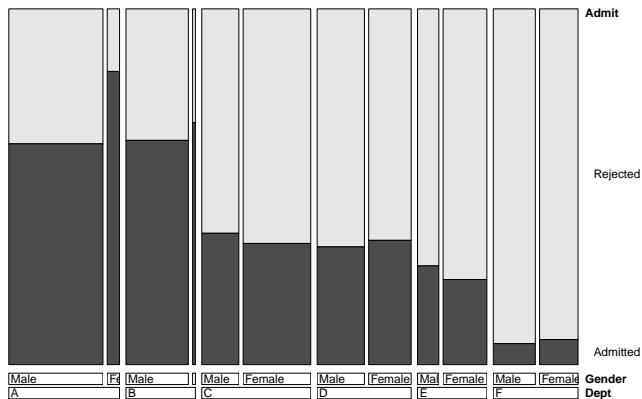
Model: (DeptGender)(DeptAdmit)



- 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.

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



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
```

Zero cells

```
structable(Titanic)
```

##		Sex	Male		Female	
##		Survived	No	Yes	No	Yes
##	Class	Age				
##	1st	Child	0	5	0	1
##		Adult	118	57	4	140
##	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
##		Adult	670	192	3	20

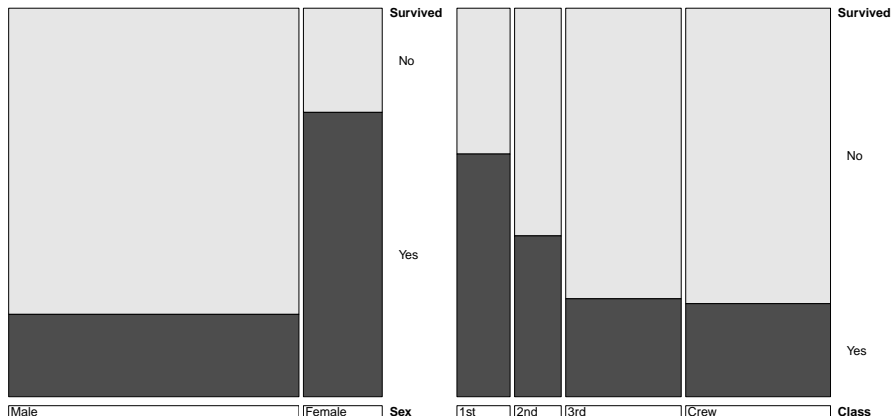
Two types of zero cells:

- **structural zeros** — could not occur (children in crew)
- **sampling zeros** — did not happen to occur (children in 1st & 2nd who died)
- Structural zeros can cause problems — loss of df; $0/0 = \text{NaN}$ in χ^2 tests

Exploratory plots

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

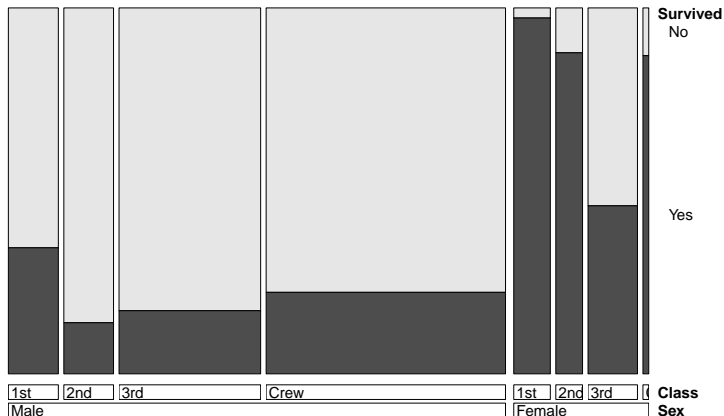
```
doubledecker(Survived ~ Sex, data=Titanic)
doubledecker(Survived ~ Class, data=Titanic)
```



Exploratory plots

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

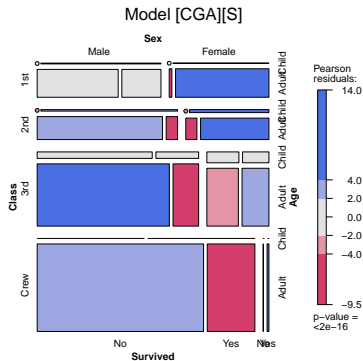
```
doubledecker(Survived ~ Sex + Class, data=Titanic)
```



Fitting and visualizing models

In the model formulas for `loglm()`, I use the variable numbers 1–4, and letters *Class*, *Gender*, *Age* and *Survived*

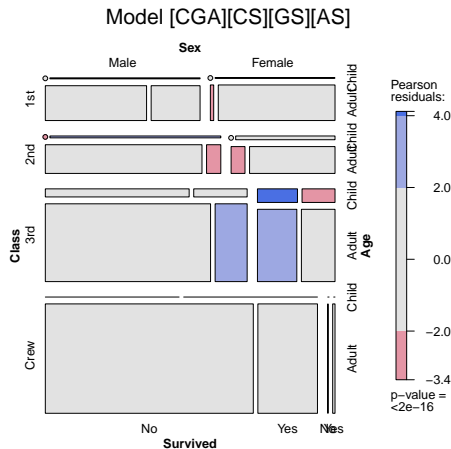
```
# mutual independence [C][G][A][S]
mod0 <- loglm(~ 1 + 2 + 3 + 4, data=Titanic)
# baseline (null) model
mod1 <- loglm(~ 1*2*3 + 4, data=Titanic)
mosaic(mod1, main="Model [CGA][S]")
```



- 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

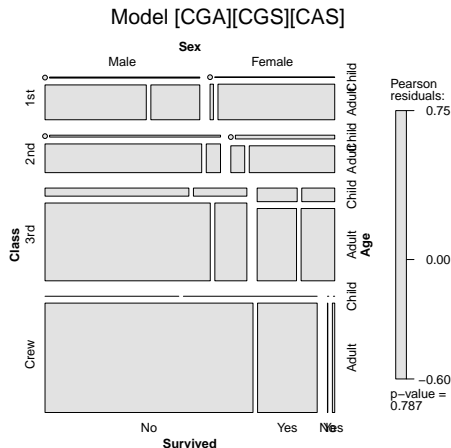
```
# main effects of C, G, A on survival: [CGA][CS][GS][AS]
mod2 <- loglm(~ 1*2*3 + (1+2+3)*4, data=Titanic)
mosaic(mod2, main="Model [CGA][CS][GS][AS]")
```



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

Final model

```
mod3 <- loglm(~ 1*2*3 + (1*2)*4 + (1*3)*4, data=Titanic)
mosaic(mod3, main="Model [CGA][CGS][CAS]")
```



Comparing models

As usual, `anova()` gives compact comparisons of a set of **nested** models.

```
anova(mod0, mod1, mod2, mod3)

## LR tests for hierarchical log-linear models
##
## Model 1:
## ~1 + 2 + 3 + 4
## Model 2:
## ~1 * 2 * 3 + 4
## Model 3:
## ~1 * 2 * 3 + (1 + 2 + 3) * 4
## Model 4:
## ~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      10      0.00000
## Model 3     112.5666  10    559.3956       5      0.00000
## Model 4       1.6855   4    110.8811       6      0.00000
## Saturated    0.0000   0      1.6855       4      0.79335
```

Comparing models

LRstats() gives compact summaries of a set of models

```
LRstats(mod0, mod1, mod2, mod3)
```

```
## Likelihood summary table:
##      AIC   BIC LR Chisq Df Pr(>Chisq)
## mod0 1385 1395    1244 25    <2e-16 ***
## mod1  833  858    672 15    <2e-16 ***
## mod2  284  316    113 10    <2e-16 ***
## mod3  185  226     2  4      0.79
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

mod3, [CGA][CGS][CAS], looks best by AIC and BIC, and also shows NS lack of fit!

Model interpretation

- Regardless of Gender and Age, lower Class \implies 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 children, ordered by class and 1st class men”
 -

Sequential plots and models

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

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

$\{v_1 v_2\}$

- 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: $X_1, X_2|X_1, X_3|X_1X_2, \dots$
- The **order of variables** matters for interpretation

Sequential plots and models

Sequential models of *joint independence* → additive decomposition of the total association, $G^2_{[v_1][v_2]...[v_p]}$ (mutual independence),

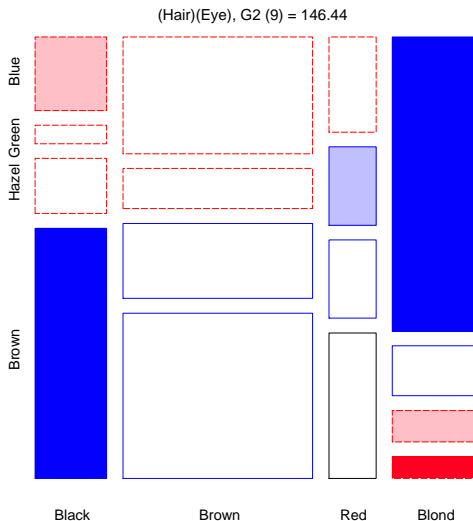
$$G^2_{[v_1][v_2]...[v_p]} = G^2_{[v_1][v_2]} + G^2_{[v_1 v_2][v_3]} + G^2_{[v_1 v_2 v_3][v_4]} + \dots + G^2_{[v_1 \dots v_{p-1}][v_p]}$$

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

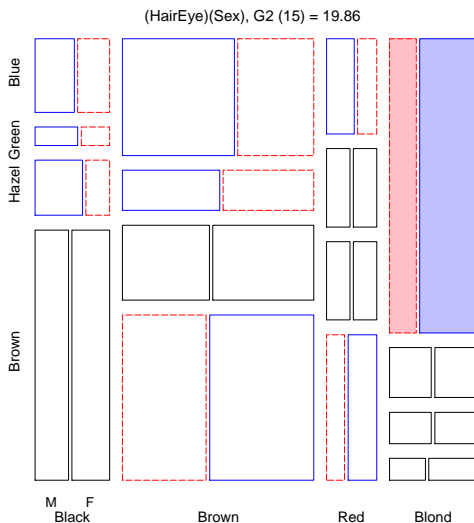
Sequential plots and models: Example

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



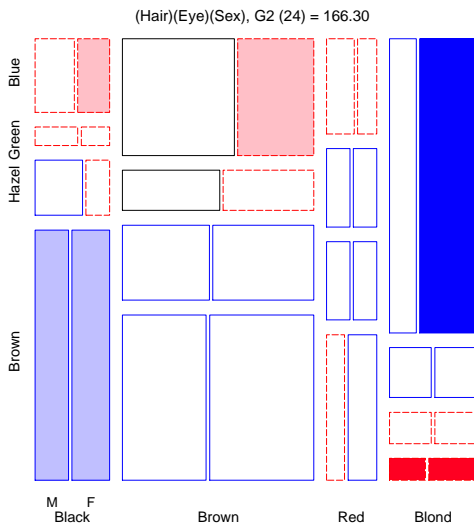
Sequential plots and models: Example

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

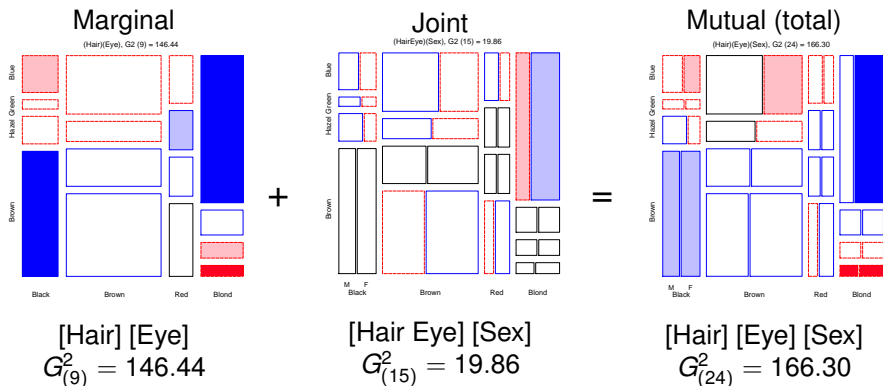


Sequential plots and models: Example

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



Sequential plots and models: Example



Applications

Response models

- When one variable, R , is a response and E_1, E_2, \dots are explanatory, 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 $[E_i R], [E_i E_j R] \dots$

Causal models

- Sometimes there is an assumed **causal ordering** of variables:

$$A \rightarrow B \rightarrow C \rightarrow D$$

- 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

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

? studied divorce patterns in relation to premarital and extramarital sex, a 2^4 table, PreSex in `vcd`

```
data("PreSex", package="vcd")
structable(Gender+PremaritalSex+ExtramaritalSex ~ MaritalStatus, PreSex)
```

##	Gender	Women				Men			
##	PremaritalSex	Yes		No		Yes		No	
##	ExtramaritalSex	Yes	No	Yes	No	Yes	No	Yes	No
##	MaritalStatus								
##	Divorced	17	54	36	214	28	60	17	68
##	Married	4	25	4	322	11	42	4	130

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?

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

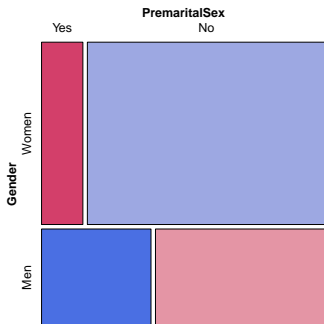
```
mod.1 <- loglm(~ Gender + PremaritalSex, data=PreSex)
mod.2 <- loglm(~ Gender * PremaritalSex + ExtramaritalSex, data=PreSex)
...
```

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

Mosaic plots:

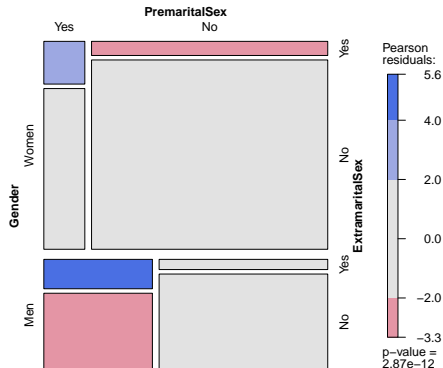
```
# (Gender Pre)
mosaic(margin.table(PreSex, 1:2), shade=TRUE,
       main = "Gender and Premarital Sex")
# (Gender Pre) (Extra)
mosaic(margin.table(PreSex, 1:3),
       expected = ~Gender * PremaritalSex + ExtramaritalSex,
       main = "Gender*Pre + ExtramaritalSex")
```

Gender and Premarital Sex



Pearson residuals:
6.3
4.0
2.0
0.0
-2.0
-4.0
-4.6
p-value = <2e-16

Gender*Pre + ExtramaritalSex

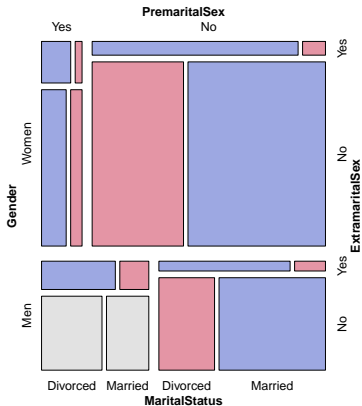


Pearson residuals:
5.6
4.0
2.0
0.0
-2.0
-3.3
p-value = 2.87e-12

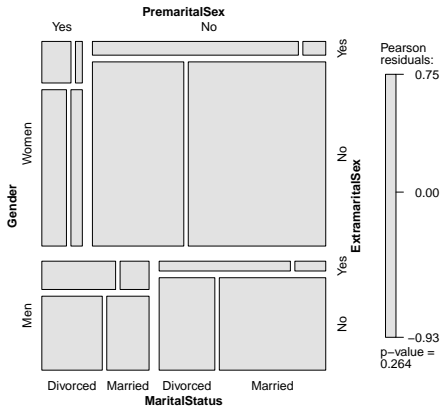
Mosaic plots:

```
mosaic(PreSex,  
       expected = ~Gender*PremaritalSex*ExtramaritalSex  
         + MaritalStatus,  
       main = "Gender*Pre*Extra + MaritalStatus")  
# (GPE) (PEM)  
mosaic(PreSex,  
       expected = ~ Gender * PremaritalSex * ExtramaritalSex  
         + MaritalStatus * PremaritalSex * ExtramaritalSex,  
       main = "G*P*E + P*E*M")
```

Gender*Pre*Extra + MaritalStatus

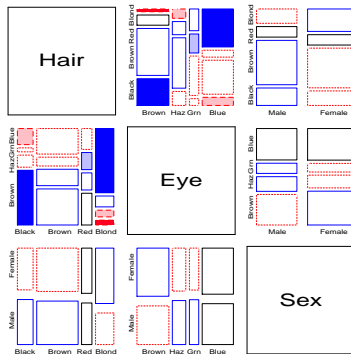


G*P*E + P*E*M

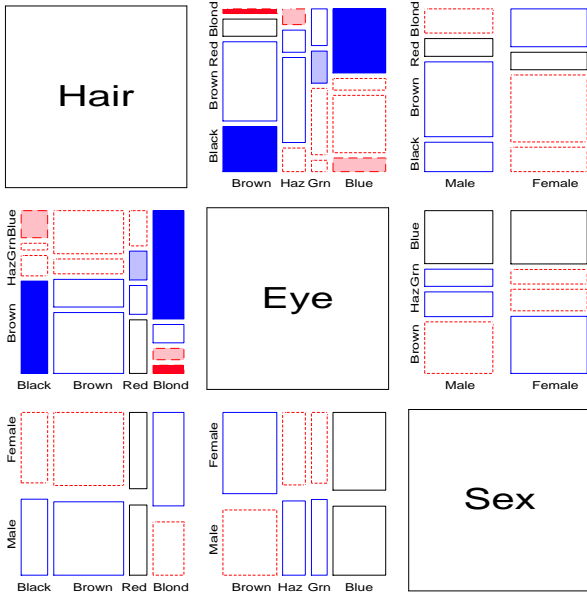


Mosaic matrices

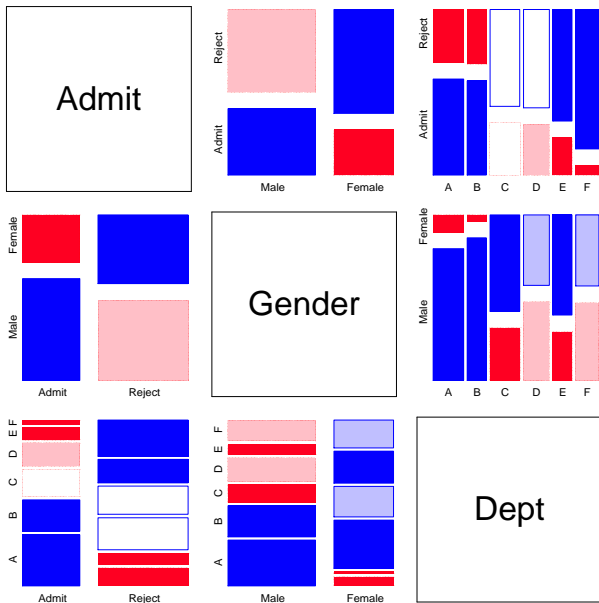
- Analog of *scatterplot matrix* for categorical data (?)
 - Shows all $p(p - 1)$ pairwise views in a coherent display
 - Each pairwise mosaic shows bivariate (marginal) relation
 - Fit: marginal independence
 - Residuals: show **marginal** associations



Hair, Eye, Sex data:



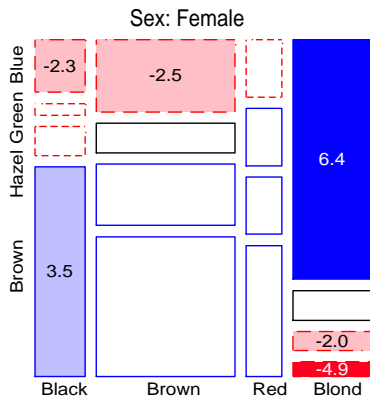
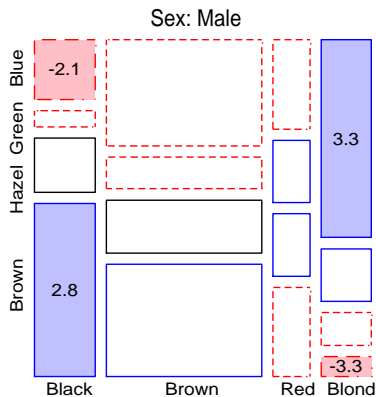
Berkeley data:



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



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, $Hair \perp Eye \mid Sex$

Model	df	G^2	p -value
$[Hair][Eye] \mid \text{Male}$	9	44.445	0.000
$[Hair][Eye] \mid \text{Female}$	9	112.233	0.000
$[Hair][Eye] \mid \text{Sex}$	18	156.668	0.000

References I