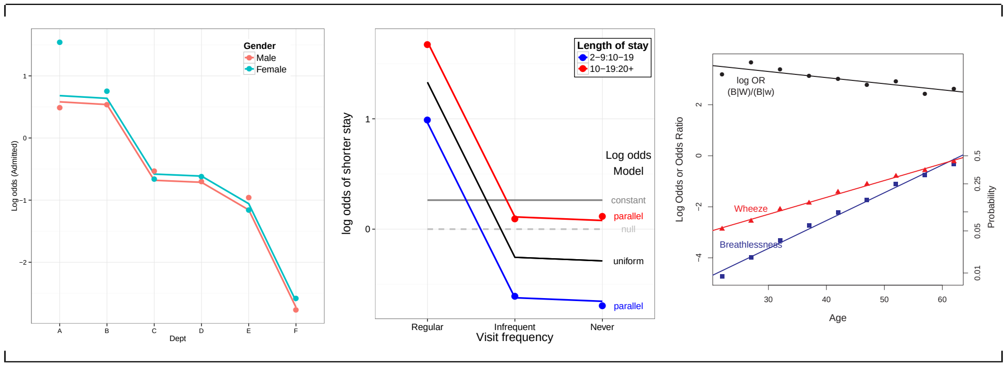
General Models and Graphs for Log Odds and Log Odds Ratios

Michael Friendly
York University, Toronto

David Meyer
UAS Technium, Vienna

CARME 2015
Naples, September 20–23, 2015
Slides: `http://datavis.ca/papers/CARME2015-2x2.pdf`



Outline

General ideas

- **Topic**: analysis and interpretation of multi-way frequency tables
  - How to visualize and understand associations?
  - How to test or compare competing explanations?
  - How to allow for special circumstances: ordinal variables, square tables, that provide simplified descriptions?
- **Loglinear models** provide one, very general approach
  - `loglm()` and Poisson `glm()` frameworks
  - Special models for ordinal variables, square tables, non-linear terms (RC models), etc.
  - A wide range of associated visualization methods: mosaic plots and family
  - Full-data plots: maybe these plot too much?

General ideas

- **CA and MCA**
  - Two-way tables: CA; *n*-way tables: MCA, JCA, etc. (but only bivariate associations)
  - Simple visualizations: 2D (3D?) plots of category points
  - Principally descriptive: Hard to specify or test specific hypotheses
  - Model plots: maybe these plot too little?
- **Odds and odds ratios**
  - Odds and odds ratios are natural summaries for quantities of interest
  - Some familiar models can be recast as models for odds or odds ratios
  - Model-based plots can provide simpler interpretation
  - Data + Model plots: maybe these are just right!

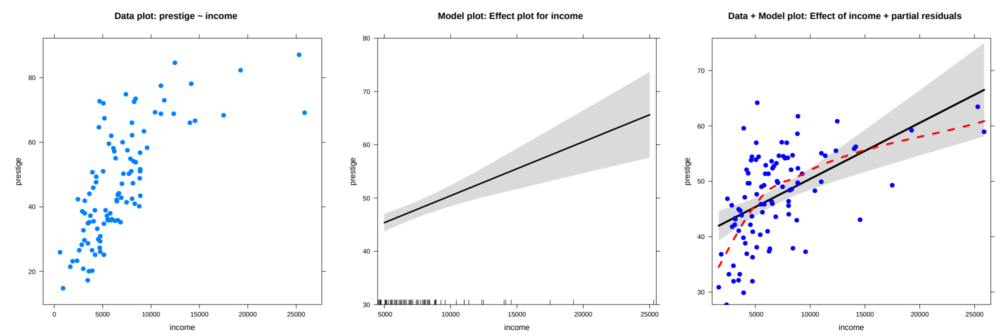
Plots: Data, Model, Data + Model

- **Data plots**: well-known. They help answer different kinds of questions:
  - What do the data look like?
  - Are there unusual features?
  - What kinds of summaries would be useful?
- **Model plots**: less well-known, but also help answer important questions:
  - What does the model look like? (plot predicted values)
  - How does the model change when its parameters change? (plot competing models)
  - How does the model change when the data is changed? (e.g., influence plots)
- **Data + Model plots** combine these features, and lead to other questions:
  - How well does a model fit the data?
  - Does a model fit uniformly good or bad, or just good/bad in some regions?
  - How can a model be improved?
  - (Model uncertainty: show confidence/prediction intervals or regions)
  - (Data support: where is data too “thin” to make a difference?)

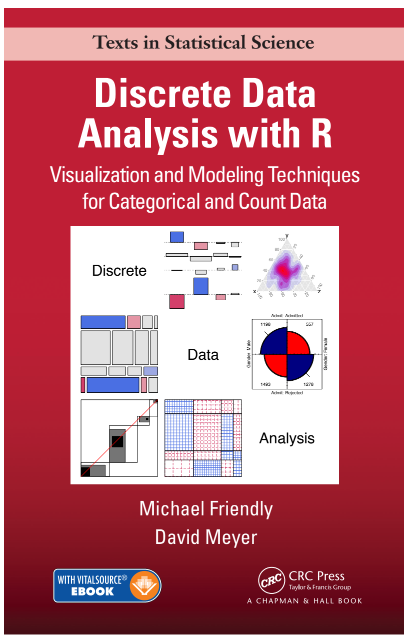
Plots: Data, Model, Data + Model

Example: Linear model— Prestige $\sim$ Income + Education + Type

- Data plot: marginal relation of Income on Prestige
- Model (effect) plot: conditional fitted values, controlling for other variables
- Data + Model plot: Effect of Income (model) + partial residuals (data)



Shameless plug



- This talk draws on material from our new book, out $\sim$ Jan., 2016.
- The successor to my earlier book, *Visualizing Categorical Data*
- Supported by many enhancements in the vcd, vcdExtra and ca packages for R
- Large collection of real data sets used in Examples (170) and Exercises (88)
- All Examples contain reproducible R code

Talk plan: Main ideas

- Familiar case— Binary responses:
  - Every loglinear model for a binary response has an equivalent form in terms of log odds [“logit” models]
  - Log odds models have simple interpretations
  - Data + model plots give simple descriptions of data and models
- Extend to two-way ($I \times J$) and three-way + ($I \times J \times K_1 \ldots$) tables:
  - Log odds as contrasts in $\log(\boldsymbol{n})$
  - Variety of simple models for log odds (ANOVA-like)
  - Easily incorporate ordinal variables
  - Data + model plots give simple descriptions of data and models
- Generalized log odds ratios capture associations between two focal variables
  - Simple linear models for LOR
  - Direct visualization (Data + model plots) $\Longrightarrow$ more sensitive comparisons

Simple example: UCB Admissions

Data on admission to graduate programs at UC Berkeley, by Dept, Gender and Admission

```
structable(Dept ~ Gender+Admit, UCBAdmissions)

##                 Dept   A   B   C   D   E   F
## Gender Admit
## Male   Admitted      512 353 120 138  53  22
##        Rejected      313 207 205 279 138 351
## Female Admitted       89  17 202 131  94  24
##        Rejected       19   8 391 244 299 317
```

or, as a two-way table (collapsed over Dept),

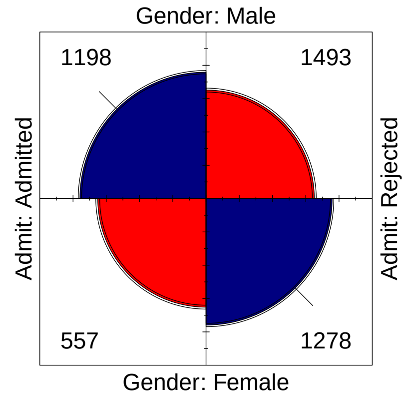
```
structable(~ Gender + Admit, UCBAdmissions)

##        Admit Admitted Rejected
## Gender
## Male             1198     1493
## Female            557     1278
```


Fourfold displays for 2 × 2 tables

General ideas:

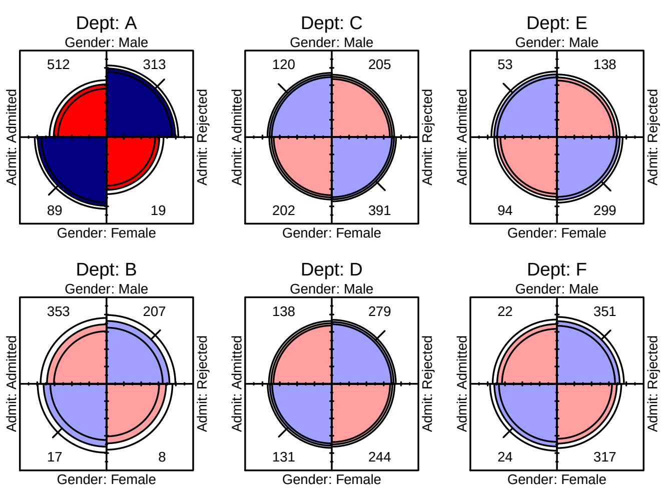
- Model-based graphs can show *both* data and model tests (or other statistical features)
- Visual attributes tuned to support perception of relevant statistical comparisons



- Quarter circles: radius $\sim \sqrt{n_{ij}} \Rightarrow$ **area** $\sim$ **frequency**
- Independence: Adjoining quadrants $\approx$ align
- Odds ratio: ratio of areas of diagonally opposite cells
- Confidence rings: Visual test of $H_0 : \theta = 1 \leftrightarrow$ adjoining rings overlap

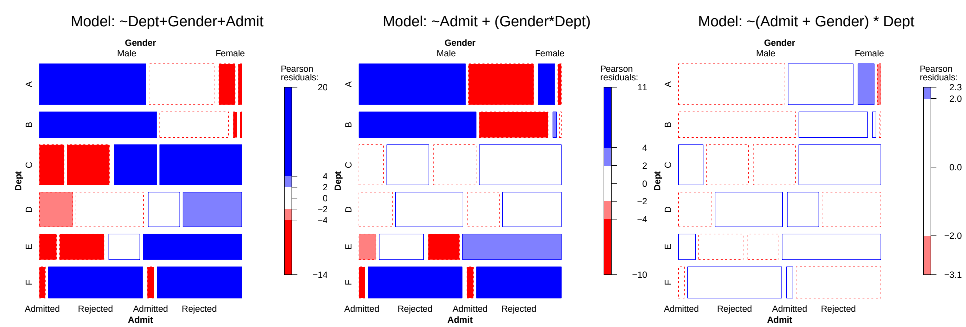
Fourfold displays for 2 × 2 × k tables

- Stratified analysis: one fourfold display for each department
- Each 2 × 2 table standardized to equate marginal frequencies
- Shading: highlight departments for which $H_a : \theta_i \neq 1$



Mosaic displays

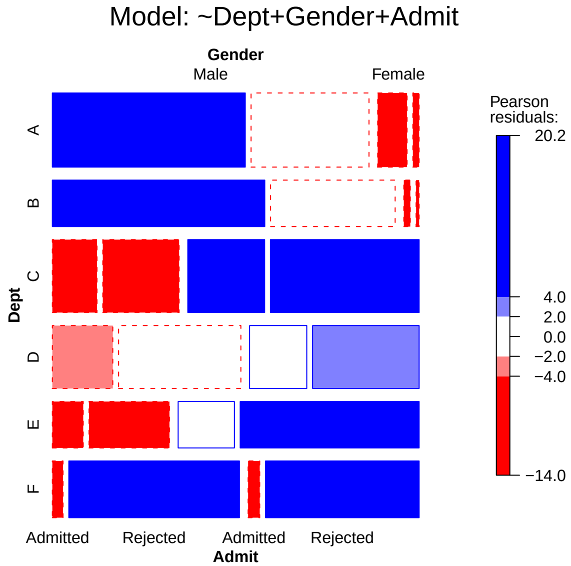
- Tiles: Area $\sim$ observed frequencies, n_{ijk}
- Friendly shading (highlight association *pattern*):
  - Residuals: $r_{ijk} = (n_{ijk} - \hat{m}_{ijk}) / \sqrt(\hat{m}_{ijk})$
  - Color— blue: $r > 0$, red: $r < 0$
  - Saturation: $|r| < 2$ (none), > 4 (max), else (middle)
- (Other shadings highlight *significance*)
- (Other color schemes: HSV, HCL, . . .)



Mosaic displays: Fitting & visualizing models

Mutual independence model: Dept ⊥ Gender ⊥ Admit

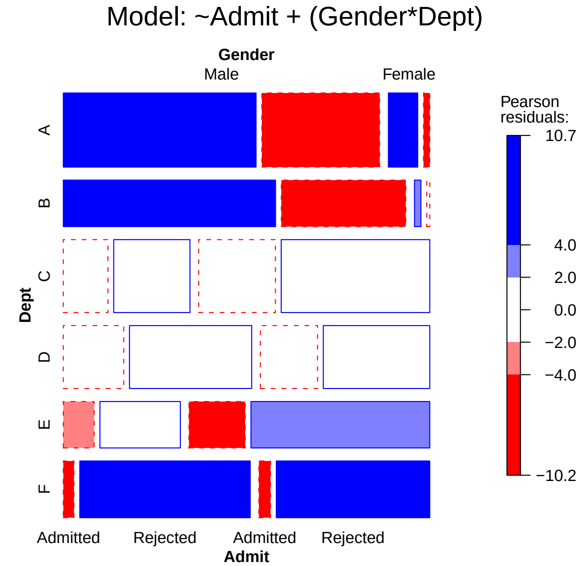
```r
berk.mod0 <- loglm(~ Dept + Gender + Admit, data=UCB)
mosaic(berk.mod0, gp=shading_Friendly, ...)
```



Mosaic displays: Fitting & visualizing models

Joint independence model: Admit ⊥ (Gender, Dept)

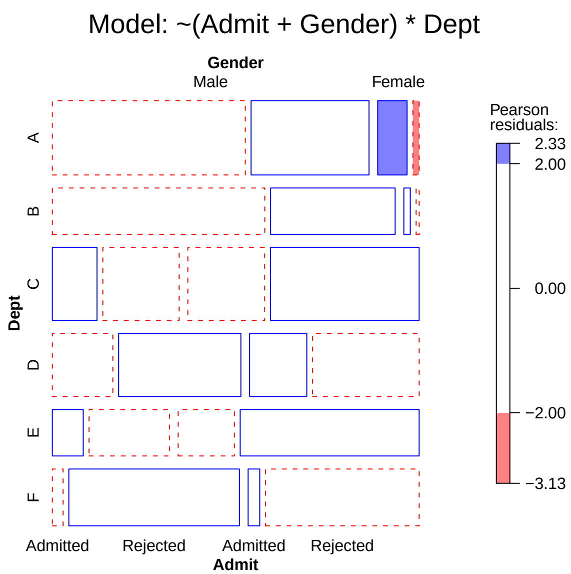
```r
berk.mod1 <- loglm(~ Admit + (Gender * Dept), data=UCB)
mosaic(berk.mod1, gp=shading_Friendly, ...)
```



Mosaic displays: Fitting & visualizing models

Conditional independence model: Admit ⊥ Gender | Dept

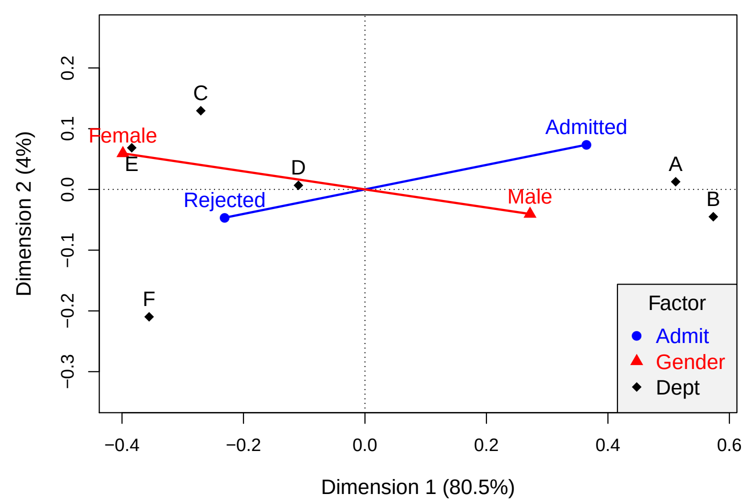
```r
berk.mod2 <- loglm(~ (Admit + Gender) * Dept, data=UCB)
mosaic(berk.mod2, gp=shading_Friendly, ...)
```



MCA

What can we learn from MCA?

```r
ucb.mca <- mjca(UCBAdmissions)
plot(ucb.mca)
```



Logit models and log odds

- For a binary response variable, each loglinear model has an equivalent logit model for log odds
- These provide a simpler way to formulate and test model(s)
- Data + Model plots are simpler to interpret the data and fitted results.

- Consider a three-way table, with variable C as a binary response, with expecected frequencies, m_{ijk}
  - For $A = i$ and $B = j$, the log odds that $C = 1$ versus $C = 2$ is

$$\psi_{ij}^{AB} = \log\left(\frac{m_{ij1}}{m_{ij2}}\right) = \log(m_{ij1}) - \log(m_{ij2}) \ .$$

  - Models now pertain to a two-way table of log odds, ψ_{ij}^{AB}
  - Plots can show observed values as points, fitted models as lines, uncertainty as error bars

Logit models and log odds

- Equivalent log odds forms:
  - the model of joint independence, [AB][C] , asserts constant log odds, $\psi_{ij}^{AB} = \alpha$
  - the model of conditional independence, [AB][AC] , allows log odds to vary with A, $\psi_{ij}^{AB} = \alpha + \beta_i^A$
  - the model of homogeneous association, [AB][AC][BC] , allows log odds to vary with A & B, $\psi_{ij}^{AB} = \alpha + \beta_i^A + \beta_j^B$

Table: Equivalent loglinear and logit models for a three-way table, with C as a binary response variable.

Loglinear model	Logit model	Logit formula
$[AB][C]$	α	`C ~ 1`
$[AB][AC]$	$\alpha + \beta_i^A$	`C ~ A`
$[AB][BC]$	$\alpha + \beta_j^B$	`C ~ B`
$[AB][AC][BC]$	$\alpha + \beta_i^A + \beta_j^B$	`C ~ A + B`
$[ABC]$	$\alpha + \beta_i^A + \beta_j^B + \beta_{ij}^{AB}$	`C ~ A * B`

Berkeley data: log odds models

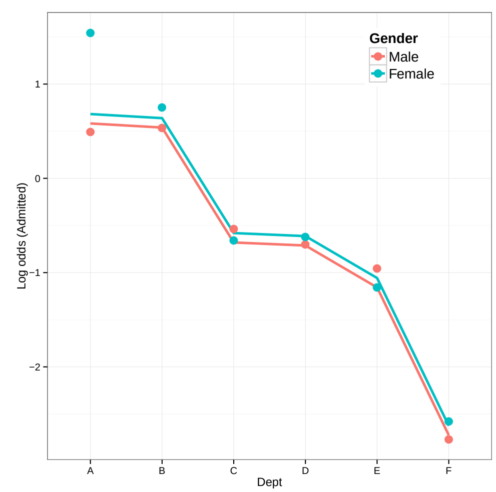
- For the `UCBAdmissions` data, the loglinear model of homogeneous association is $[AD][AG][DG]$.
- This model doesn't fit very well: $G^2(5) = 20.2$. Why?
- The equivalent log odds model is:

$$\psi_{ij} = \log\left(\frac{m_{\text{Admit}(ij)}}{m_{\text{Reject}(ij)}}\right) = \alpha + \beta_i^{\text{Dept}} + \beta_j^{\text{Gender}} \ .$$

- This is the parallel odds model, $\sim$ a main-effects ANOVA model.
- Fit this using `glm()`:

```
berkeley <- as.data.frame(UCBAdmissions)
berk.logit2 <- glm(Admit == "Admitted" ~ Dept + Gender,
                   data = berkeley, weights = Freq,
                   family = "binomial")
```


Berkeley data: log odds models

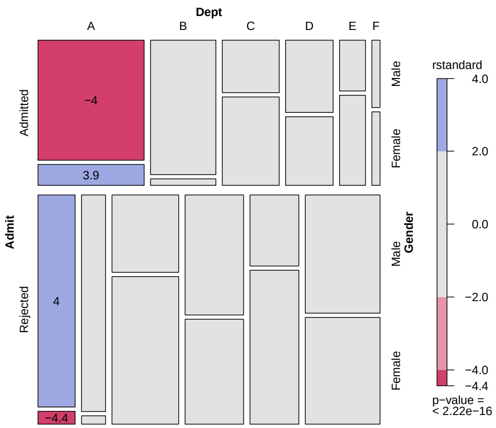


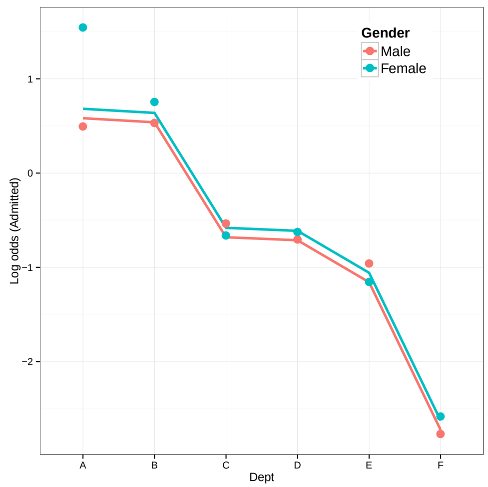
- Data + Model plot
- The effect of gender is extremely small (NS)
- Main lack of fit is for Dept A
- Fitted values for departments have a sensible interpretation
- i.e., reflect overall rate of admission

Berkeley data: log odds models

Compare with mosaic display: log odds plot is much clearer

Model: [DeptGender][DeptAdmit][AdmitGender]





Berkeley data: log odds models

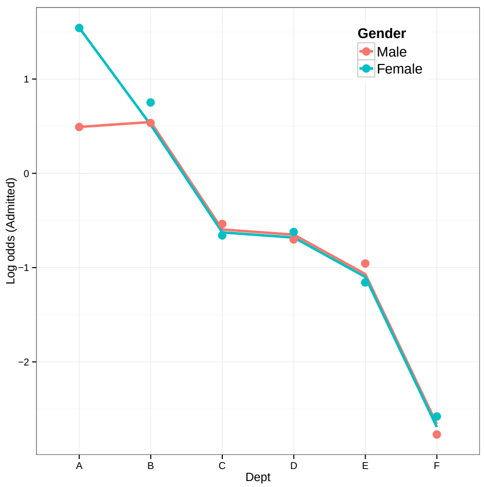
Fit a simpler, more adequate model for log odds:

- Drop the general 1 df term for Gender ([AG] in the loglinear model)
- Replace with a specific 1 df term for Gender, only in Dept. A

$$\psi_{ij} = \alpha + \beta_i^{\text{Dept}} + I(j = 1)\beta^{\text{Gender}} \ .$$

- This model now fits very well: $G^2(5) = 2.68$

Berkeley data: log odds models



- Data + Model plot
- Excellent fit is now evident
- Simple interpretation:
  - Admission depends *only* on department
  - ... except in Dept A

Two-way Tables: Log odds

The log odds approach extends directly to general $I \times J$ tables:

- Consider a two way $I \times J$ table of variables A and B, where B is the response and A is explanatory.
- Questions:
  - How does the distribution of categories of B vary over the levels of A?
  - How to visualize associations?
  - How to test precise hypotheses?
- Log odds approach:[1]
  - $I \times J \rightarrow (J-1)$ log odds contrasts for the categories of B for each level of A
  - What models summarise these values?
  - (Similar to polytomous response models in logistic regression)

[1] These ideas stem from Goodman (1983), *Biometrics* and related papers.

Example: Hospital Visits

How does the length of stay in hospital differ among schizophrenic patients, classified by the frequency of visiting by friends and relatives?

```
data(HospVisits, package="vcdExtra")
HospVisits

##              stay
## visit         2-9 10-19 20+
##   Regular      43    16   3
##   Infrequent    6    11  10
##   Never         9    18  16
```

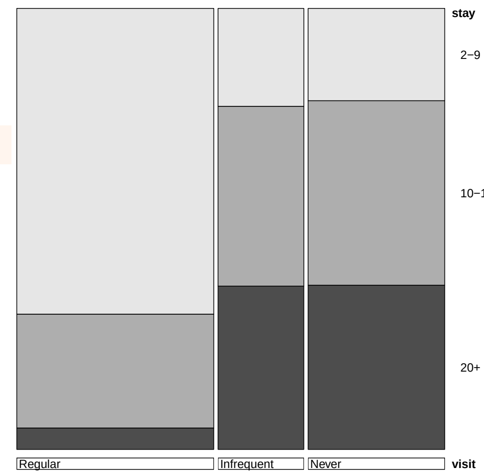
- Both frequency of `visit` (explanatory) and length of `stay` are ordinal variables
- Standard methods (loglinear models) treat these as nominal ("factors")
- Specialized models can take ordinality into account, e.g., with linear or quadratic effects

Exploratory visualizations: Doubledecker plot

Doubledecker plot

```
doubledecker(HospVisits)
```

- Shows directly the conditional distributions of `stay` given `visit`
- Length of stay is shorter with frequent visits
- Infrequent and Never don't differ very much

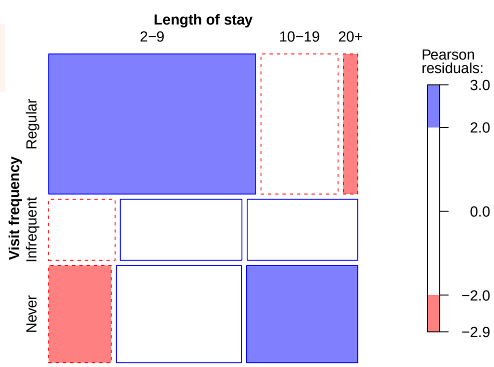


Exploratory visualizations: Mosaic plot

Mosaic plot

```
mosaic(HospVisits,
   gp=shading_Friendly)
```

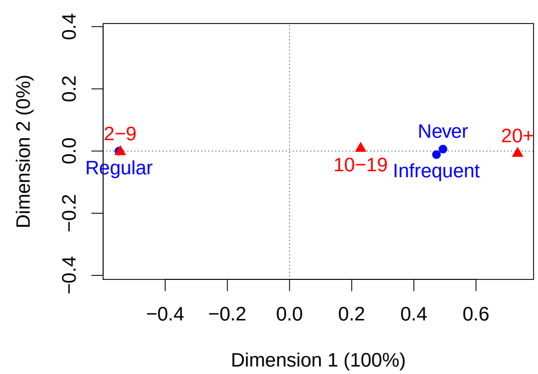
- Also shows the conditional distributions as area-proportional tiles
- Color shows departure (residuals) from the independence model
- The "opposite corner" pattern signals a possibly unidimensional relationship between visit and stay



Exploratory visualizations: CA

What does CA tell us?

```
plot(ca(HospVisits))
```



- Association is entirely 1D!
- Infrequent and Never category points don't differ much
- Greater visit frequency associated with shorter stay

But, how can we test and and visualize these ideas with models?

Models for log odds

- Start with the saturated loglinear model for the two-way table

$$\log m_{ij} = \mu + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$$

- For adjacent categories of the response variable B, the odds, $\omega_{ij}^{A\bar{B}}$ and log odds, $\psi_{ij}^{A\bar{B}}$, that the response is in category j rather than $j+1$ are:

$$\text{odds: } \omega_{ij}^{A\bar{B}} = \frac{m_{ij}}{m_{i,j+1}} \qquad \text{log odds: } \psi_{ij}^{A\bar{B}} = \log\left(\frac{m_{ij}}{m_{i,j+1}}\right) \quad , j = 1, \ldots, J-1$$

- For the hospital visits data, this gives:

```
t(lodds(HospVisits, response=2))

## log odds for visit by stay
##
##              visit
## stay          2-9:10-19 10-19:20+
##   Regular        0.9886   1.67398
##   Infrequent    -0.6061   0.09531
##   Never         -0.6931   0.11778
```


Models for log odds

A variety of simple models can be specified in terms of log odds:

Table: Models for adjacent log odds in an $I \times J$ table with B as the response

Model	log odds parameters	degrees of freedom
null log odds	$\psi_{ij}^{A\bar{B}} = 0$	$I(J-1)$
constant log odds	$\psi_{ij}^{A\bar{B}} = \psi$	$I(J-1)-1$
uniform B log odds	$\psi_{ij}^{A\bar{B}} = \psi_i^A$	$I(J-2)$
parallel log odds	$\psi_{ij}^{A\bar{B}} = \psi_i^A + \psi_j^B$	$(I-1)(J-2)$
saturated	$\psi_{ij}^{A\bar{B}}$ unspecified	

- The log odds, $\psi_{ij}^{A\bar{B}}$ can be viewed as entries in an $I \times (J-1)$ table
- These models are analogous to ANOVA tests of the A, B and $A * B$ effects in this table.

Fit some models

```
mod.null  <- lm(logodds ~ -1,     data=hosp.lodds)      # null
mod.const <- lm(logodds ~ 1,      data=hosp.lodds)      # constant
mod.unif  <- lm(logodds ~ visit, data=hosp.lodds)       # uniform
mod.par   <- lm(logodds ~ visit + stay, data=hosp.lodds)  # parallel
```

Compare models:

```
anova(mod.null, mod.const, mod.unif, mod.par)

## Analysis of Variance Table
##
## Model 1: logodds ~ -1
## Model 2: logodds ~ 1
## Model 3: logodds ~ visit
## Model 4: logodds ~ visit + stay
##   Res.Df  RSS Df Sum of Sq   F Pr(>F)
## 1      6 4.65
## 2      5 4.24  1      0.41 177 0.0056 **
## 3      4 3.43  1      0.81 345 0.0029 **
## 4      2 0.00  2      3.43 734 0.0014 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


Ordinal variables

When the levels of A are ordinal, we can also test for linear effects.

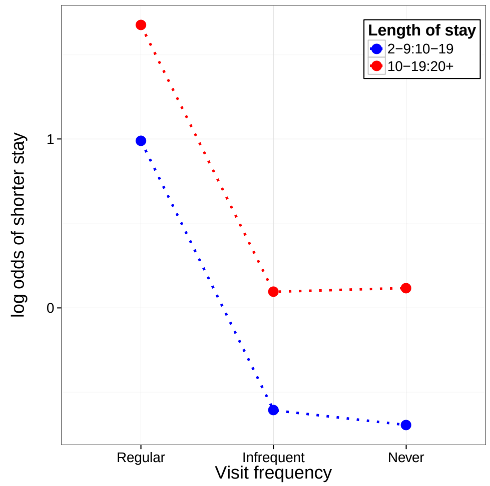
```
mod1a <- lm(logodds ~ as.numeric(visit), data=hosp.lodds)
mod2a <- lm(logodds ~ as.numeric(visit) + stay, data=hosp.lodds)
# compare parallel log odds models
anova(mod.const, mod2a, mod.par)

## Analysis of Variance Table
##
## Model 1: logodds ~ 1
## Model 2: logodds ~ as.numeric(visit) + stay
## Model 3: logodds ~ visit + stay
##   Res.Df  RSS Df Sum of Sq   F Pr(>F)
## 1      5 4.24
## 2      2 0.00  3      4.23 604 0.0017 **
## 3      2 0.00  0      0.00
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

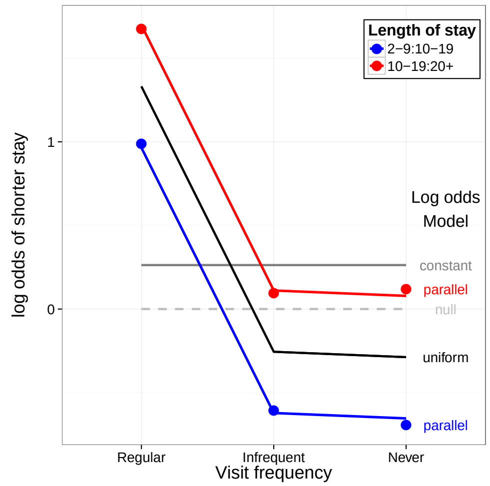
Effects of `visit` are certainly not linear.

Visualizing log odds and models

Plots of observed and fitted log odds: easy interpretation of data and models



Data plot: Observed log odds



Data + Model plot (fitted log odds)

Three-way+ Tables: Log odds I

These methods naturally extend to three- and higher-way tables:

- Consider a three-way $I \times J \times K$ table of variables A, B and C, where C is the response (or focal variable)
- The standard loglinear model is:

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

- For categories k and $k+1$ the adjacent log odds for C are

$$\text{log odds: } \psi_{ijk}^{AB\overline{C}} = \log\left(\frac{m_{ijk}}{m_{i,j+1}}\right) \quad , \quad k = 1, \ldots, K-1$$

- These log odds can be viewed as entries in a two-way, $IJ \times (K-1)$ table.

Three-way+ Tables: Log odds II

- The parallel log odds model is

$$\begin{aligned} \psi_{ijk}^{AB\overline{C}} &= \Psi_{ij}^{AB} + \psi_k^C \\ &= \psi + \psi_i^A + \psi_j^B + \psi_{ij}^{AB} + \psi_k^C \end{aligned}$$

  where the Ψ_{ij}^{AB} are unspecified and the ψ parameters obey standard (sum-to-zero) constraints.
- Simpler models:

$$\begin{aligned} \text{uniform log odds:} \quad & \psi_k^C = 0 \\ \text{joint independence:} \quad & \Psi_{ij}^{AB} = \psi \end{aligned}$$

- Even simpler models: null effects of A ($\psi_i^A = 0$) or B ($\psi_j^B = 0$)
- Linear effects models: An ordinal A can use $\psi_i^A = i \times \beta_A$ to test for linearity

Example: Mice Depletion Data

- Kastenbaum and Lamphiear (1959) gave a $3 \times 5 \times 2$ table of the number of deaths (0, 1, 2+) in 657 litters of mice, classified by litter size (7–11) and treatment (“A”, “B”)
- How does number of deaths depend on litter size and treatment?

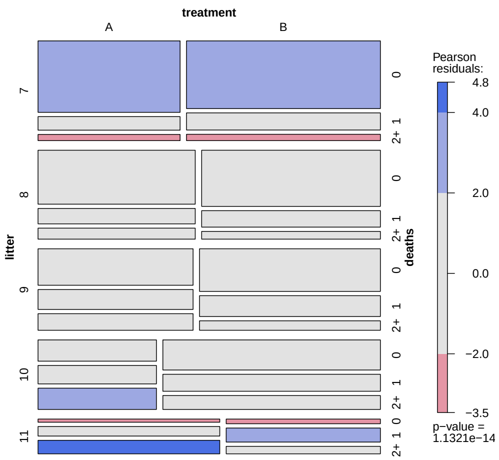
```
data(Mice, package="vcdExtra")
mice.tab <- xtabs(Freq ~ litter + treatment + deaths, data=Mice)
ftable(litter + treatment ~ deaths, data=mice.tab)

##        litter     7     8     9    10    11
##        treatment  A  B  A  B  A  B  A  B  A  B
## deaths
## 0                58 75 49 58 33 45 15 39  4  5
## 1                11 19 14 17 18 22 13 22 12 15
## 2+                5  7 10  8 15 10 15 18 17  8
```

Mice data: Mosaic plot

Fit and display the model of joint independence, [litter, treatment] [deaths]

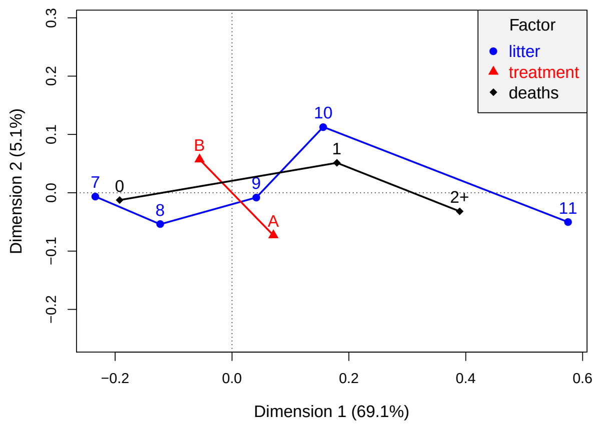
```
mosaic(mice.tab, expected= ~ litter * treatment + deaths)
```



- What can we see?
- Small litters more likely to have 0 deaths
- Large litters more likely to have 2+ deaths
- More deaths with treatment A than B

Mice data: MCA

```
mice.mca <- mjca(mice.tab)
plot(mice.mca)
```



What can we see?

- Larger litter size associated with more deaths
- More deaths with treatment A than B
- What model? How to simplify?

Calculating log odds

For a three-way table, a simple way to calculate all (log) odds is to reshape the data as a two-way matrix, $\boldsymbol{T}$, with $I \times J$ rows and K columns.

```
##          0  1 2+
## 7:A     58 11  5
## 8:A     49 14 10
## 9:A     33 18 15
## 10:A    15 13 15
## 11:A     4 12 17
...
```

The $IJ \times (K-1)$ table of adjacent log odds can then be calculated as $\log(\boldsymbol{T})\boldsymbol{C}$, where $\boldsymbol{C}$ is the $K \times K-1$ matrix of contrasts,

$$\boldsymbol{C} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}$$

Calculating log odds

More generally,

- Consider an $R \times K_1 \times K_2 \times \dots$ frequency table $n_{ij\dots}$, with factors $K_1, K_2 \dots$ considered as strata.
- Let $\boldsymbol{n} = \text{vec}(n_{ij\dots})$ be the $N \times 1$ vectorization of the table.
- Then, all log odds and their asymptotic covariance matrix $\boldsymbol{S}$ can be calculated as:
  - $\widehat{\psi} = \boldsymbol{C}\log(\boldsymbol{n})$
  - $\boldsymbol{S} = \text{Var}[\psi] = \boldsymbol{C}\,\text{diag}\,\boldsymbol{n}^{-1}\,\boldsymbol{C}^{\mathsf{T}}$

  where $\boldsymbol{C}$ is an N-column matrix containing all zeros, except for one $+1$ elements and one -1 elements in each row.
- With strata, $\boldsymbol{C}$ can be calculated as the Kronecker product $\boldsymbol{C} = \boldsymbol{C}_R \otimes \boldsymbol{I}_{K_1} \otimes \boldsymbol{I}_{K_2} \otimes \cdots$
- Linear models for log odds: $\psi = \boldsymbol{X}\beta$

Mice data: Log odds

The vcd package now contains a general implementation of these ideas:

- `odds()` and `lodds()`: calculate odds and log odds for 1 variable in an *n*-way table
- provides methods (`coef()`, `vcov()`, `confint()` ...) for `"lodds"` objects

```
(mice.lodds <- as.data.frame(lodds(mice.tab, response="deaths")))

##    litter treatment deaths logodds    ASE
## 1     0:1         7      A  1.6625 0.3289
## 2    1:2+         7      A  0.7885 0.5394
## 3     0:1         8      A  1.2528 0.3030
## 4    1:2+         8      A  0.3365 0.4140
## 5     0:1         9      A  0.6061 0.2930
## 6    1:2+         9      A  0.1823 0.3496
## 7     0:1        10      A  0.1431 0.3789
## 8    1:2+        10      A -0.1431 0.3789
## 9     0:1        11      A -1.0986 0.5774
...
```


Mice data: Fit models

Use WLS, with weights $\sim ASE^{-2}$

```
mod0 <- lm(logodds ~ 1, weights=1/ASE^2, data=mice.lodds)
mod1 <- lm(logodds ~ litter + treatment, weights=1/ASE^2, data=mice.lodds)
mod2 <- lm(logodds ~ litter * treatment, weights=1/ASE^2, data=mice.lodds)
mod3 <- lm(logodds ~ litter * treatment + deaths, weights=1/ASE^2, data=mi
```

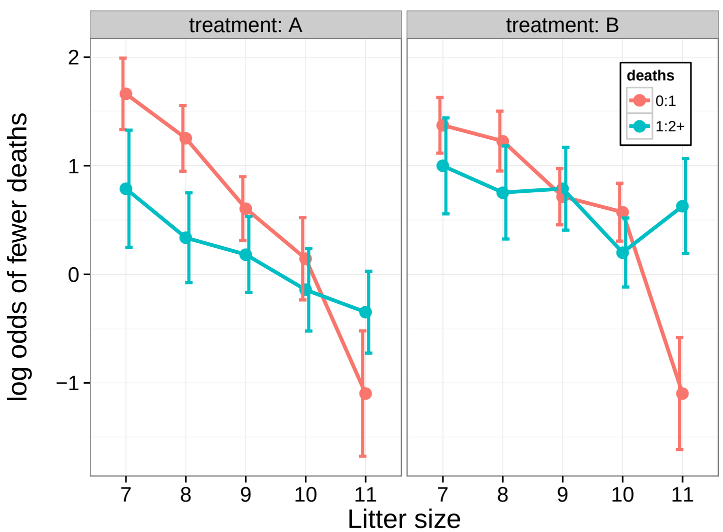
Compare models:

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

## Analysis of Variance Table
##
## Model 1: logodds ~ 1
## Model 2: logodds ~ litter + treatment
## Model 3: logodds ~ litter * treatment
## Model 4: logodds ~ litter * treatment + deaths
##   Res.Df  RSS Df Sum of Sq     F  Pr(>F)
## 1     19 65.0
## 2     14 17.8  5      47.2 18.22 0.00018 ***
## 3     10  6.7  4      11.1  5.36 0.01737 *
## 4      9  4.7  1       2.1  3.98 0.07723 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

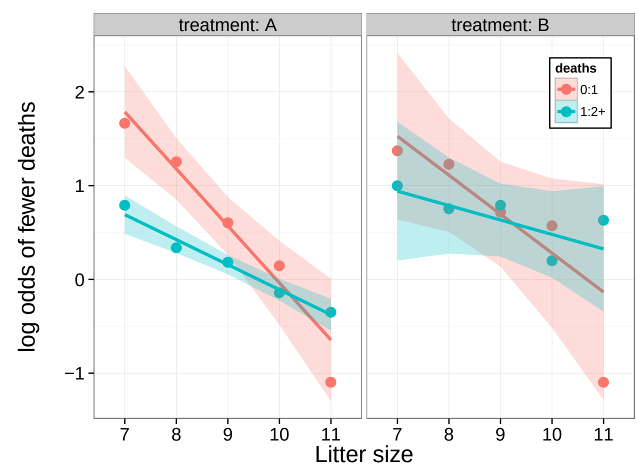

Visualize log odds and models: Data plot

- Data plot: log odds with error bars: $\psi_{ijk}^{AB\overline{C}} \pm 1 ASE_{\psi}$
- This is equivalent to the saturated model for log odds



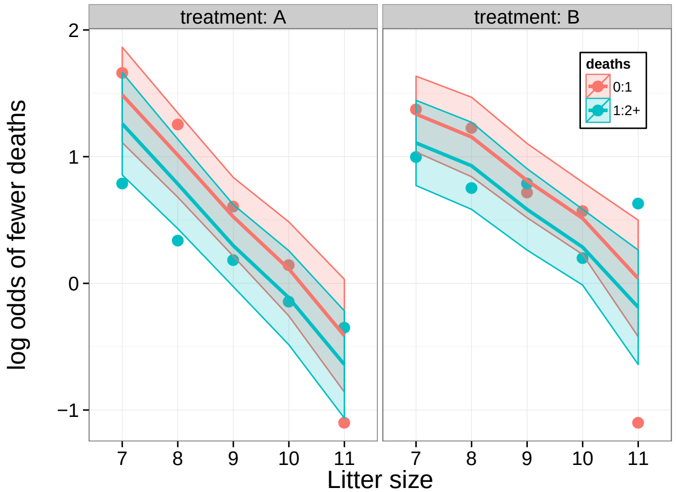
Visualize log odds and models: Smoothing

- Apply a linear smoother (weighed linear regression) to each
- This is equalvalent to a model with a three-way term, `as.numeric(litter)*treatment*deaths`
- Error bands show model uncertainty



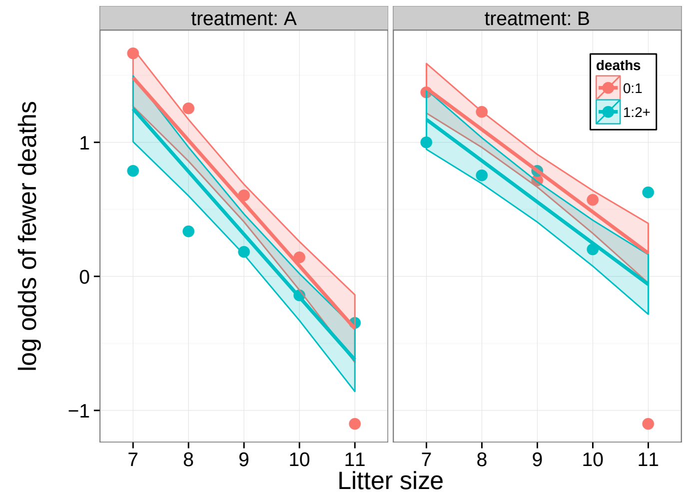
Visualize log odds and models: Model + Data plots

- Display the fit of the parallel log odds model, $\psi^{AB\bar{C}}_{ijk} = \Psi^{AB}_{ij} + \psi^{C}_{k}$



Visualize log odds and models: Model + Data plots

- Simplify the model: fit only linear effects of `litter`
- `lm(logodds   as.numeric(litter)*treatment + deaths)`
- Error bands show smaller model uncertainty



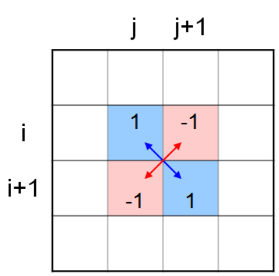
Generalized log odds ratios

- In any two-way, $R \times C$ table, *all* associations can be represented by a set of $(R-1) \times (C-1)$ odds ratios,

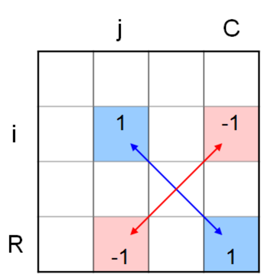
$$\theta_{ij} = \frac{n_{ij}/n_{i+1,j}}{n_{i,j+1}/n_{i+1,j+1}} = \frac{n_{ij} \times n_{i+1,j+1}}{n_{i+1,j} \times n_{i,j+1}}$$

Simpler in terms of log odds ratios:

$$\log(\theta_{ij}) = \begin{pmatrix} 1 & -1 & -1 & 1 \end{pmatrix} \log \begin{pmatrix} n_{ij} & n_{i+1,j} & n_{i,j+1} & n_{i+1,j+1} \end{pmatrix}^{T}$$



local odds ratios



ref='last' odds ratios

Generalized log odds ratios

- $\log \theta_{ij} \sim \mathcal{N}(0, \sigma^2)$, with estimated asymptotic standard error:

$$\widehat{\sigma}(\log \theta_{ij}) = (n_{ij}^{-1} + n_{i+1,j}^{-1} + n_{i,j+1}^{-1} + n_{i+1,j+1}^{-1})^{1/2}$$

- This extends naturally to $\theta_{ij \mid k}$ in higher-way tables, stratified by one or more "control" variables.
- Many models have a simpler form expressed in terms of $\log(\theta_{ij})$.
  - e.g., Uniform association model

$$\log(m_{ij}) = \mu + \lambda^A_i + \lambda^B_j + \gamma \boldsymbol{a}_i \boldsymbol{b}_j \equiv \log(\theta_{ij}) = \gamma$$

- Direct visualization of log odds ratios permits more sensitive comparisons than area-based displays.

Models for log odds ratios: Computation

- Consider an $R \times C \times K_1 \times K_2 \times \ldots$ frequency table $n_{ij\ldots}$, with factors $K_1, K_2 \ldots$ considered as strata.
- Let $\boldsymbol{n} = \text{vec}(n_{ij\ldots})$ be the $N \times 1$ vectorization of the table.
- Then, all log odds ratios and their asymptotic covariance matrix $\boldsymbol{S}$ can be calculated as:
  - $\log(\widehat{\boldsymbol{\theta}}) = \boldsymbol{C} \log(\boldsymbol{n})$
  - $\boldsymbol{S} = \text{Var}[\log(\boldsymbol{\theta})] = \boldsymbol{C} \,\text{diag}\, \boldsymbol{n}^{-1} \, \boldsymbol{C}^{\mathsf{T}}$

  where $\boldsymbol{C}$ is an N-column matrix containing all zeros, except for two $+1$ elements and two -1 elements in each row.
- With strata, $\boldsymbol{C}$ can be calculated as $\boldsymbol{C} = \boldsymbol{C}_{RC} \otimes \boldsymbol{I}_{K_1} \otimes \boldsymbol{I}_{K_2} \otimes \cdots$
- `loddsratio()` in vcd provides generic methods (`coef()`, `vcov()`, `confint()`, ...)
- `plot()` method gives reasonable data and model plots.

Models for log odds ratios: Computation

For example, for a 2×3 table, there are two adjacent odds ratios

```
##    Age
## Sex Yng Mid Old
##   M  30  20  10
##   F   5  15  25
## log odds ratios for Sex and Age
##
## Yng:Mid Mid:Old
##   1.504   1.204
```

These are calculated as:

$$\log(\boldsymbol{\theta}) = \boldsymbol{C} \log(\boldsymbol{n}) = \begin{bmatrix} 1 & -1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 1 \end{bmatrix} \log \begin{pmatrix} n_{11} \\ n_{21} \\ n_{12} \\ n_{21} \\ n_{13} \\ n_{23} \end{pmatrix}$$

Models for log odds ratios: Estimation

- A log odds ratio linear model for the $\log(\boldsymbol{\theta})$ is

$$\log(\boldsymbol{\theta}) = \boldsymbol{X}\beta$$

  where $\boldsymbol{X}$ is the design matrix of covariates
- The (asymptotic) ML estimates $\widehat{\beta}$ are obtained by GLS via

$$\widehat{\beta} = \left(\boldsymbol{X}^{\mathsf{T}} \boldsymbol{S}^{-1} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^{\mathsf{T}} \boldsymbol{S}^{-1} \log\left(\widehat{\boldsymbol{\theta}}\right)$$

  where $\boldsymbol{S} = \text{Var}[\log(\boldsymbol{\theta})]$ is the estimated covariance matrix
- $\Longrightarrow$ Standard graphical and diagnostic methods can be adapted to this case.
  - visualization: full-model plots, effect plots, ...
  - diagnostics: influence plots, added-variable plots, ...

Example: Breathlessness & Wheeze in Coal Miners

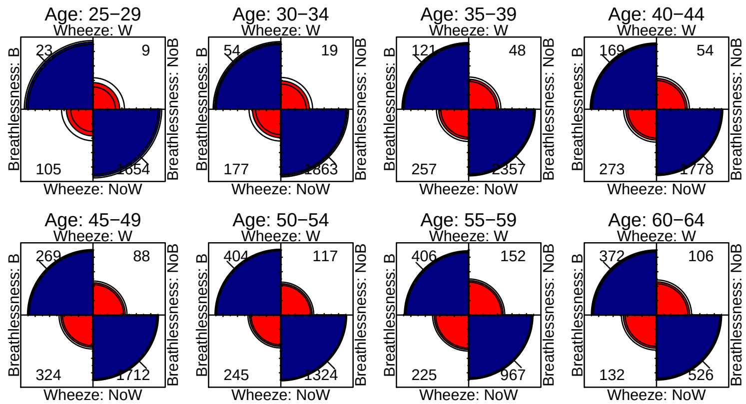
- Ashford & Sowden (1970) gave data on the association between two pulmonary conditions: breathlessness and wheeze, in a large sample of coal miners
- Age is the primary covariate
- How does the association between breathlessness and wheeze vary with age?

```
ftable(CoalMiners)

##                       Age 25-29 30-34 35-39 40-44 45-49 50-54 55-59 60-
## Breathlessness Wheeze
## B              W             23    54   121   169   269   404   406   3
##                NoW            9    19    48    54    88   117   152   1
## NoB            W            105   177   257   273   324   245   225   1
##                NoW         1654  1863  2357  1778  1712  1324   967   5
```

Example: Breathlessness & Wheeze in Coal Miners

```
fourfold(CoalMiners, mfcol=c(2,4), fontsize=18)
```



- There is a strong + association at all ages
- But can you see the trend?

Coal Miners: Models

```
(lor.CM <- loddsratio(CoalMiners))

## log odds ratios for Breathlessness and Wheeze by Age
##
## 25-29 30-34 35-39 40-44 45-49 50-54 55-59 60-64
## 3.695 3.398 3.141 3.015 2.782 2.926 2.441 2.638
```

How does LOR vary with Age?

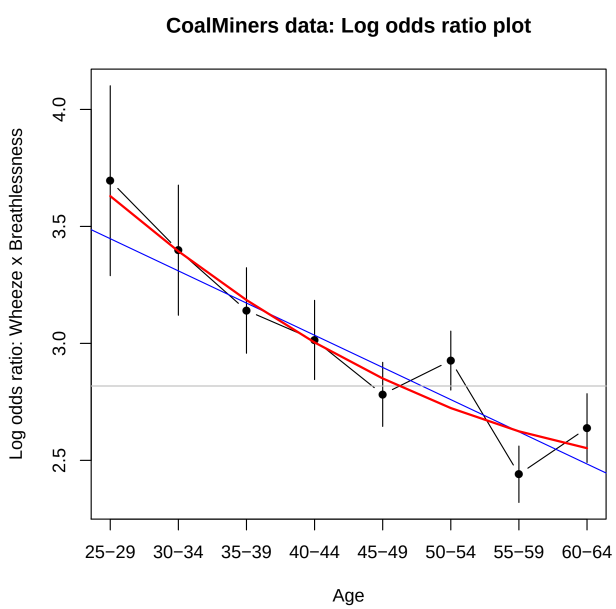
- Uniform association: $\ln(\theta) = \beta_0$
- Linear association: $\ln(\theta) = \beta_0 + \beta_1$ Age
- Quadratic association: $\ln(\theta) = \beta_0 + \beta_1$ Age $+ \beta_2$ Age^2

Fit models using WLS:

```
lor.CM.df <- as.data.frame(lor.CM)
age <- seq(25, 60, by = 5)
CM.mod0 <- lm(LOR ~ 1,            weights=1/ASE^2, data=lor.CM.df)
CM.mod1 <- lm(LOR ~ age,          weights=1/ASE^2, data=lor.CM.df)
CM.mod2 <- lm(LOR ~ poly(age,2), weights=1/ASE^2, data=lor.CM.df)
```


Coal Miners: LOR plot

Plot log odds ratios and fitted regressions: The trend is now clear!



Coal Miners: Model comparisons

Standard ANOVA procedures allow tests of nested competing models:

```
anova(CM.mod0, CM.mod1, CM.mod2)

## Analysis of Variance Table
##
## Model 1: LOR ~ 1
## Model 2: LOR ~ age
## Model 3: LOR ~ poly(age, 2)
##   Res.Df   RSS Df Sum of Sq     F Pr(>F)
## 1      7 25.61
## 2      6  6.34  1     19.28 17.23 0.0089 **
## 3      5  5.60  1      0.74  0.66 0.4525
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(vcdExtra::`LRstats()` gives direct tests of each model, and AIC, BIC)
The linear model, $\ln(\theta) = \beta_0 + \beta_1$ Age, gives the best fit.

Going further: Bivariate response models

- In this example, breathlessness and wheeze are two binary responses
- A bivariate logistic response model fits simultaneously
  - the marginal log odds of each response, ψ_1, ψ_2 vs. predictors ($\boldsymbol{x}$)
  - the joint log odds ratio, ϕ_{12}, vs. $\boldsymbol{x}$
- This model has the form

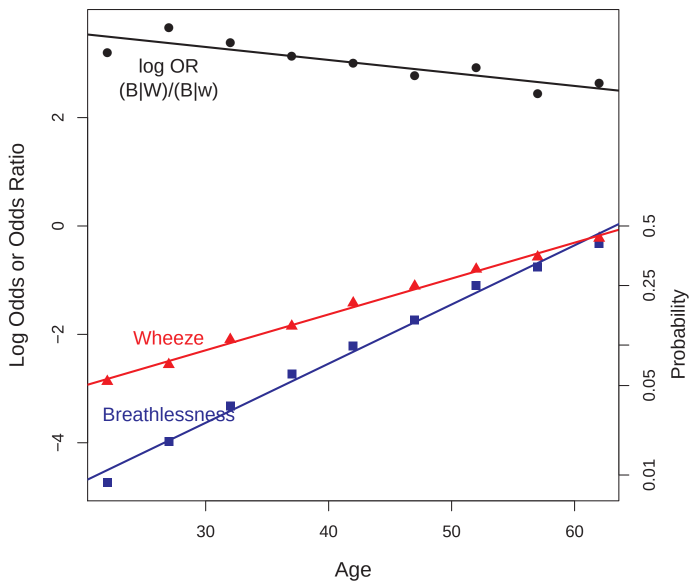
$$\boldsymbol{\eta}(\boldsymbol{x}) = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_{12} \end{pmatrix} \equiv \begin{pmatrix} \log \text{odds}_1(\boldsymbol{x}) \\ \log \text{odds}_2(\boldsymbol{x}) \\ \log \text{OR}_{12}(\boldsymbol{x}) \end{pmatrix} \equiv \begin{pmatrix} \psi_1 \\ \psi_2 \\ \log \theta_{12} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1^{\mathsf{T}} \boldsymbol{\beta}_1 \\ \boldsymbol{x}_2^{\mathsf{T}} \boldsymbol{\beta}_2 \\ \boldsymbol{x}_{12}^{\mathsf{T}} \boldsymbol{\beta}_{12} \end{pmatrix}$$

where $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_{12} \subset \boldsymbol{x}$

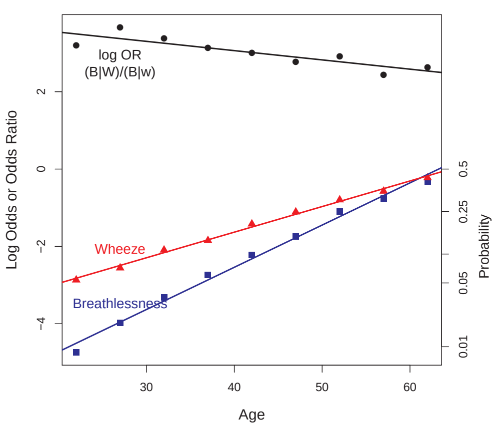
- For example, with one x, the following model allows linear effects on log odds, with a constant log odds ratio

$$\begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_{12} \end{pmatrix} = \begin{pmatrix} \alpha_1 + \beta_1 x \\ \alpha_2 + \beta_2 x \\ \log(\theta) \end{pmatrix} \qquad (1)$$

Linear model for log odds and log odds ratios



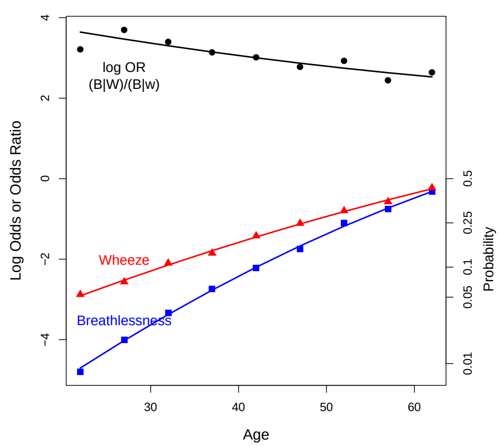
Linear model for log odds and log odds ratios



This data + model plot has a simple interpretation:

- Prevalence of breathlessness and wheeze both increase with age
- Breathlessness is less prevalent at young age, but increases faster
- Their association decreases approx. linearly, but is still strong

Quadratic model for log odds and log odds ratios



- Allowing quadratic fits in age serves as a sensitivity check
- The story is pretty much the same

Example: Attitudes toward corporal punishment

A four-way table, classifying 1,456 persons in Denmark (`Punishment` data in vcd).

- Attitude: approves moderate punishment of children ("moderate"), or refuses any punishment ("no")
- Memory: Person recalls having been punished as a child?
- Education: highest level (elementary, secondary, high)
- Age group: (15–24, 25–39, 40+)

		Age	15–24		25–39		40+	
Education	**Attitude**	**Memory**	Yes	No	Yes	No	Yes	No
Elementary	No		1	26	3	46	20	109
	Moderate		21	93	41	119	143	324
Secondary	No		2	23	8	52	4	44
	Moderate		5	45	20	84	20	56
High	No		2	26	6	24	1	13
	Moderate		1	19	4	26	8	17

Questions

Interest focuses on several questions:

- How does Attitude toward punishment depend on Memory, Education and Age?
  - Model log odds approve of moderate corporal punishment
  - Standard logit model:

```
glm(attitude   memory + education + age, data=Punishment,
weight=Freq, family=binomial)
```

  - Visualize: Effect plots for model terms
- How does association between Attitude and Memory vary with Education and Age?
  - Model log odds ratio (Attitude, Memory)
  - Visualize: LOR plots

Log odds model for Attitude

Fit the main-effects model for Attitude on other predictors:

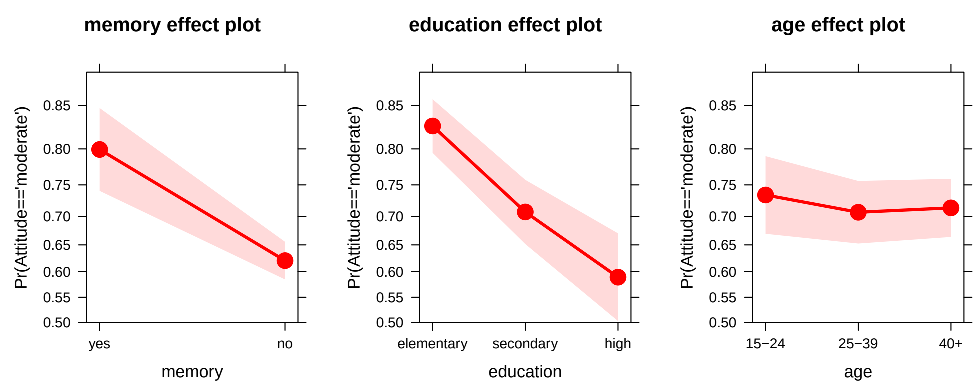
```
pun.logit <- glm(attitude ~ memory + education + age,
                 data=Punishment, weight=Freq, family=binomial)
Anova(pun.logit)

## Analysis of Deviance Table (Type II tests)
##
## Response: attitude
##           LR Chisq Df Pr(>Chisq)
## memory        29.5  1    5.6e-08 ***
## education     50.3  2    1.2e-11 ***
## age            0.6  2       0.73
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

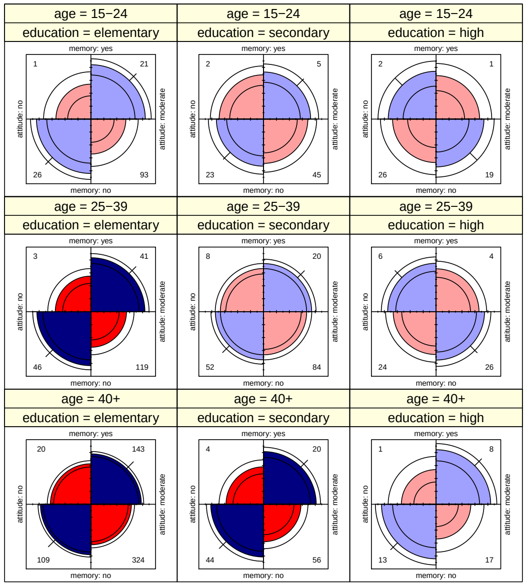
- Only Memory and Education have significant effects
- A more complex model with all two-way interactions showed no improvement

Effect plots

- **Model plots**, showing fitted values for high-order terms in any model
- Other predictors averaged over in each plot
- Simple interpretation:
  - Those who remembered punishment as children more likely to approve
  - Approval decreases with education
  - No effect of age



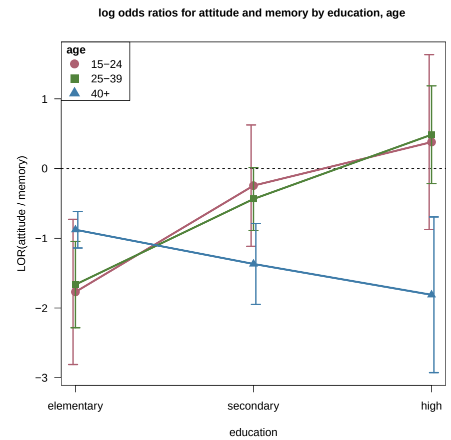
Association of attitude with memory: Fourfold plots



Log odds ratio plot

```
(lor.pun <- loddsratio(punish))

## log odds ratios for memory and attitude by age, education
##
##         education
## age      elementary secondary    high
##    15-24    -1.7700   -0.2451  0.3795
##    25-39    -1.6645   -0.4367  0.4855
##    40+      -0.8777   -1.3683 -1.8112
```



- Structure now completely clear
- Little diff[ce] between younger groups
- Opposite pattern for the 40+
- Fit an LOR model to confirm appearences (SEs large)!

Summary & conclusions

- Data exploration and model building are two parts of data analysis
  - Goal of data analysis: tell a useful, credible story
  - Different kinds of plots are useful: data plots model plots, data + model plots
- Plots in the mosaic family are useful, but may be complex for large tables
- Plots in the CA/MCA family are useful, but often don't go far enough
- log odds: Simple models and plots for one focal (response) variable
  - Simple extension of logit models for a binary response
  - Easy calculation: contrasts of log frequency
  - Easy estimation: weighted linear models for log odds: $\psi = \boldsymbol{X}\beta$
- log odds ratios: Simple models and plots for two focal variables
  - Express all associations in terms of $\log(\theta_{ij})$
  - Simple weighted linear models: $\log(\boldsymbol{\theta}) = \boldsymbol{X}\beta$
  - Simple data + model plots
- Now available in the vcd: `lodds()` and `loddsratio()`.

Further information

DDAR Friendly, M. & Meyer, D. (2016). *Discrete Data Analysis with R: Visualization and Modeling Techniques for Categorical and Count Data* Chapman & Hall/CRC, Jan. 2016. https://www.crcpress.com/9781498725835

vcd Zeileis A, Meyer D & Hornik K (2006). The Strucplot Framework: Visualizing Multi-Way Contingency Tables with **vcd**. *Journal of Statistical Software*, **17**(3), 1–48. http://www.jstatsoft.org/v17/i03/ `vignette("strucplot", package="vcd")`.

vcdExtra Friendly M & others (2010). **vcdExtra**: vcd additions. http://CRAN.R-project.org/package=vcdExtra. `vignette("vcd-tutorial")`.