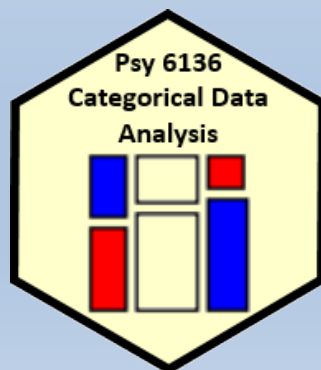


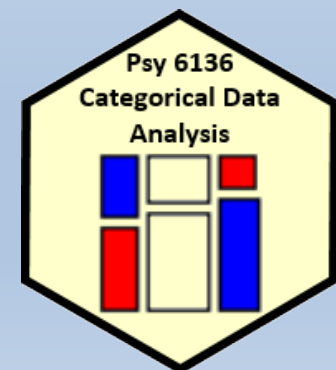
Models & graphs for log odds and log odds ratios



Michael Friendly

Psych 6136

<http://friendly.github.io/6136>



Today's topics

- We've come a long way, but there is more...
- Logit models → general models for **log odds**
 - Two-way tables
 - Three-way + tables
 - Log odds plots
- Models for generalized **odds ratios**
 - Log odds ratios
 - Bivariate response models

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 + ($I \times J \times K_1 \dots$) tables:
 - Log odds as **contrasts** in $\log(n)$
 - 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) \implies more sensitive comparisons

Logit models → Log odds models

- In an $I \times 2$ table for variables $[A \ B]$, where B is a binary response, the logit model expresses the log odds that $B=1$ vs. $B=2$

$$\psi_i^A = \log \left(\frac{m_{i1}}{m_{i2}} \right)$$

- Models pertain to the **one-way** log odds
- This generalizes to $I \times J$ tables, where we consider $(J-1)$ log odds for each level of A , e.g.,
 - Adjacent categories

$$\psi_{ij}^{A\bar{B}} = \log \left(\frac{m_{ij}}{m_{i(j+1)}} \right) \quad j = 1, 2, \dots, J-1$$

- In general, $I \times J \rightarrow (J-1)$ log odds **contrasts** of the B categories for each level of A
- Similar to how **polytomous responses** treated in logistic regression
- Can also use comparisons with a baseline category

J responses → J-1 contrasts/logits

Adjacent-category logits

b1	b2	b3	b4
1	-1		
	1	-1	
		1	-1

$$\text{For } A_i : \psi_j^{\bar{B}} = \log(m_{ij}) - \log(m_{i(j+1)})$$

Reference-level logits

b1	b2	b3	b4
1			-1
	1		-1
		1	-1

$$\text{For } A_i : \psi_j^{\bar{B}} = \log(m_{ij}) - \log(m_{i(J)})$$

Generalized logit models extend the advantages of the standard one to a polytomous response

2-way 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
```

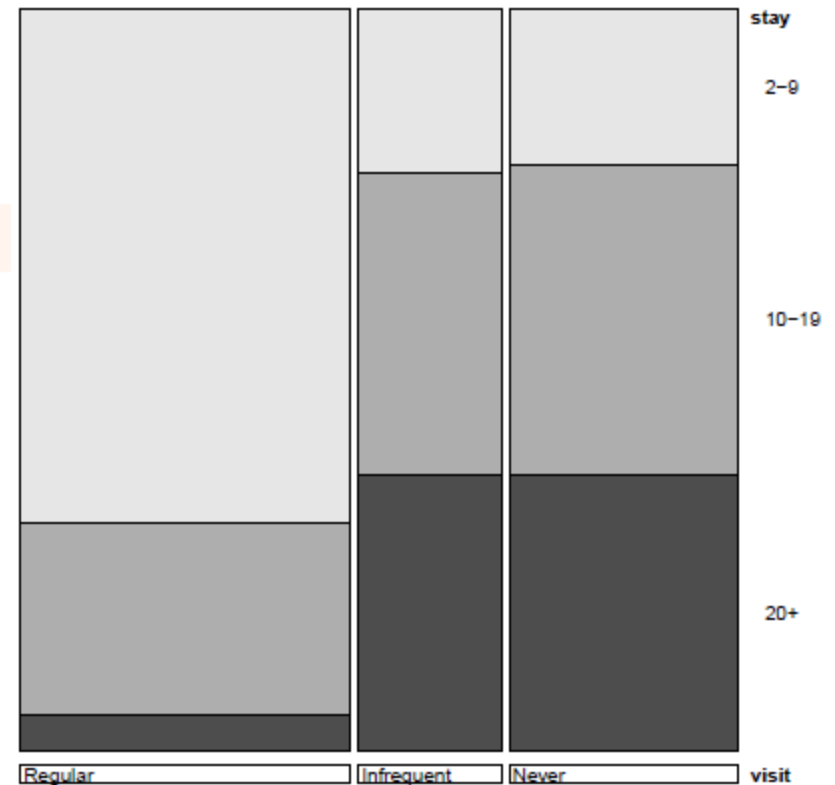
- Length of stay is the response, and it is **ordered**
- Can model the **adjacent** odds or log odds that stay is category j vs $(j+1)$
 - E.g., stay= 2-9 vs. 10-19; stay= 10-19 vs. 20+
- In general, $I \times J \rightarrow I \times (J-1)$ adjacent comparisons
- visit is also **ordered**. Can consider simpler (e.g., linear) models for the log odds

Exploratory plots: Doubledecker

Doubledecker plot

```
doubledecker(HospVisits)
```

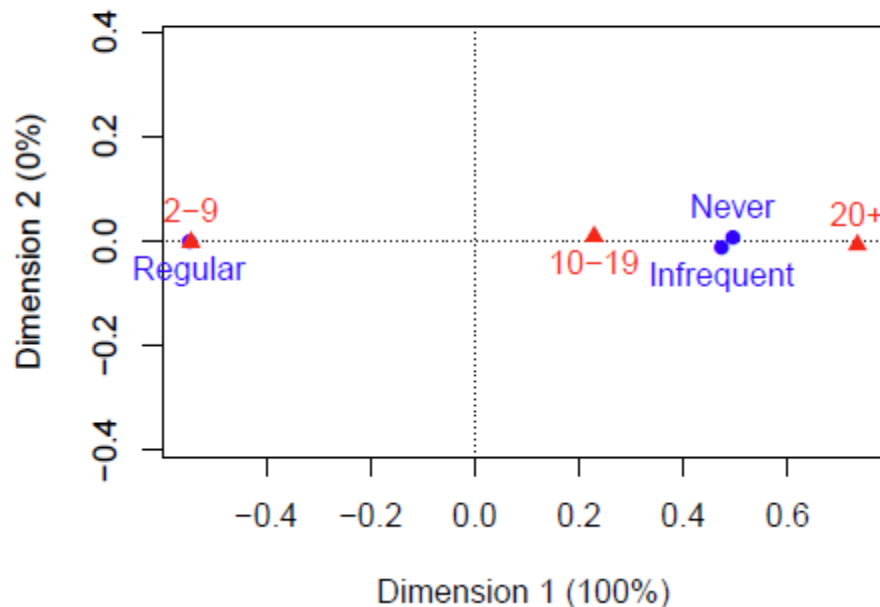
- Shows directly the conditional distributions of `stay` given `visit`
- Length of stay is shorter with frequent visits
- Infrequent and Never don't differ very much



Exploratory plots: 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 **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:

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

- For the hospital visits data, this gives:

```
> t(lodds(HospVisits, response = "stay"))  
log odds for stay by visit
```

visit	stay	
	2-9:10-19	10-19:20+
Regular	0.989	1.6740
Infrequent	-0.606	0.0953
Never	-0.693	0.1178

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	$\psi_{ij}^{AB} = 0$	$I(J - 1)$
constant log odds	$\psi_{ij}^{AB} = \psi$	$I(J - 1) - 1$
uniform B log odds	$\psi_{ij}^{AB} = \psi_i^A$	$I(J - 2)$
parallel log odds	$\psi_{ij}^{AB} = \psi_i^A + \psi_j^B$	$(I - 1)(J - 2)$
saturated	ψ_{ij}^{AB} unspecified	

- The log odds, ψ_{ij}^{AB} 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

I'm simply using `lm()` here. Should use WLS: $\text{weights} = 1/\text{ASE}^2$

```
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.

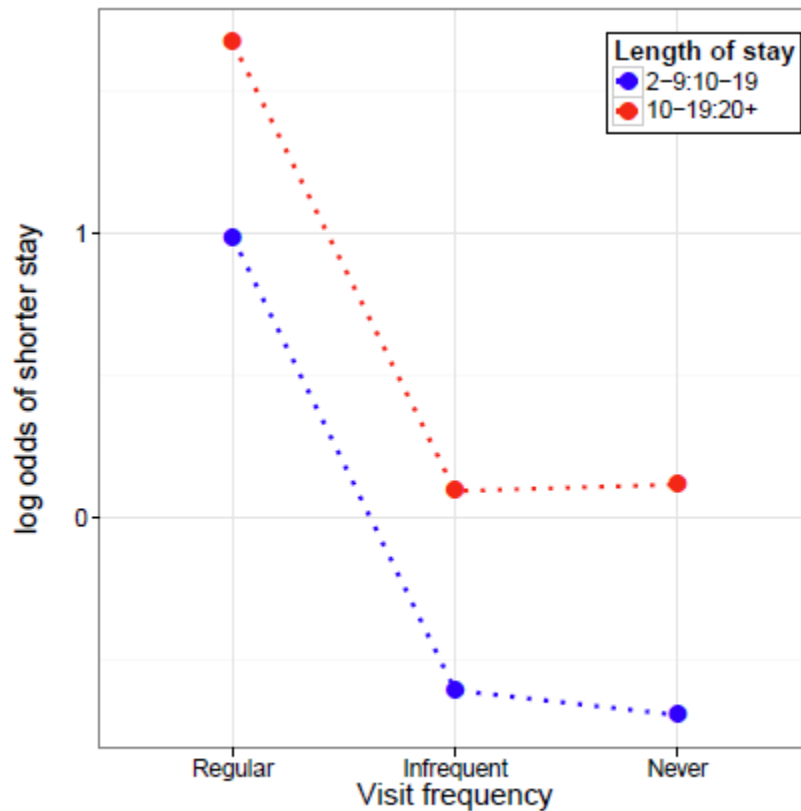
```
mod1a <- lm(logodds ~ as.numeric(visit), data=hosp.lodds)
mod2a <- lm(logodds ~ as.numeric(visit) + stay, data=hosp.lodds)
# compare parallel log odds models
anova(mod.const, mod2a, mod.par)

## 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
```

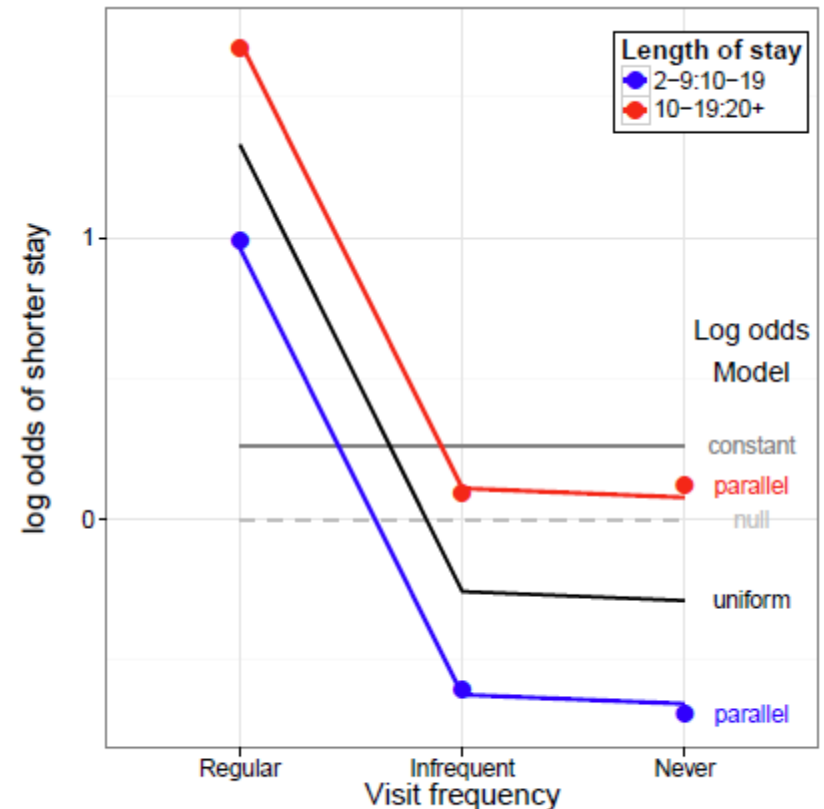
Effects of **visit** are certainly not linear.

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)

Visualizing log odds and models

Basic plot:

```
gg <- ggplot(hosp.lodds, aes(x=visit, y=logodds,  
                             group=stay, color=stay)) +  
  geom_point(size=5) +  
  geom_line(size=1.2, linetype="dotted")  
  ylab("log odds of shorter stay\n") +  
  xlab("Visit frequency") + theme_bw() + ...
```

Add lines for predicted values from the models

```
grid <- hosp.lodds[,1:2]  
gg_lines <- function(grid, mod, size=1.2, color=NULL, ...) {  
  grid$logodds <- stats::predict(mod, grid)  
  if(is.null(color)) geom_line(data=grid, size=size, ...)  
  else geom_line(data=grid, size=size, color=color, ...)  
}  
  
gg + gg_lines(grid, mod.null, color="gray", size=1, linetype="dashed") +  
  gg_lines(grid, mod.const, color=gray(.5), size=1) +  
  gg_lines(grid, mod.unif, color="black", size=1) +  
  gg_lines(grid, mod.par)
```

Three-way+ tables: Log odds

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

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

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

Three-way+ tables: Log odds

- The **parallel log odds** model is

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

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

3-way 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
## 1
## 2+
```

		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

→ Adjacent categories:

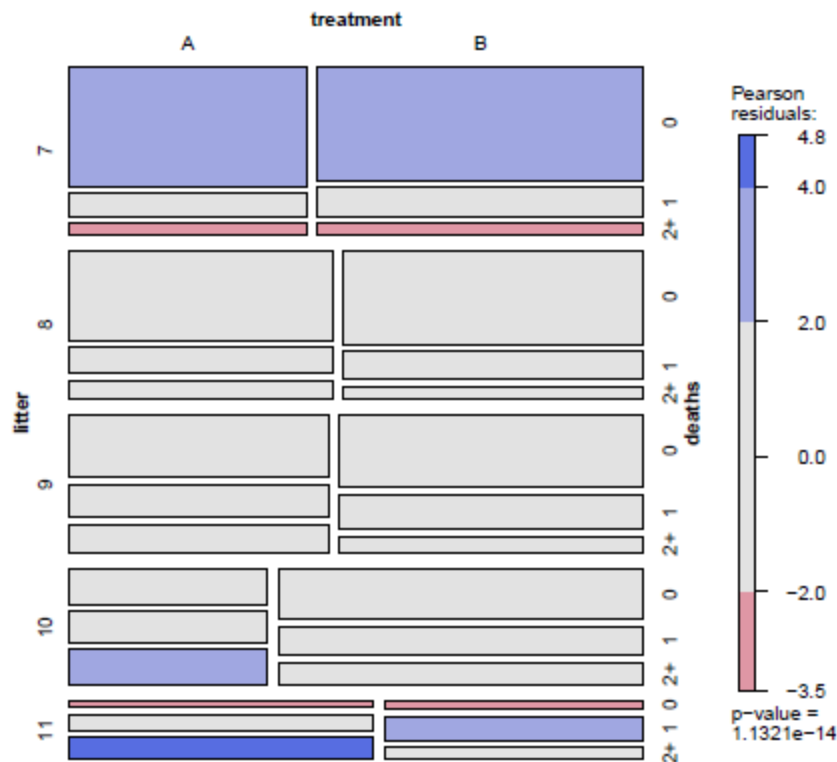
- Odds or log odds of 0 vs. 1 deaths
- Odds or log odds of 1 vs. 2+ deaths

How do these differ with litter size & treatment?

Mice data: mosaic plot

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

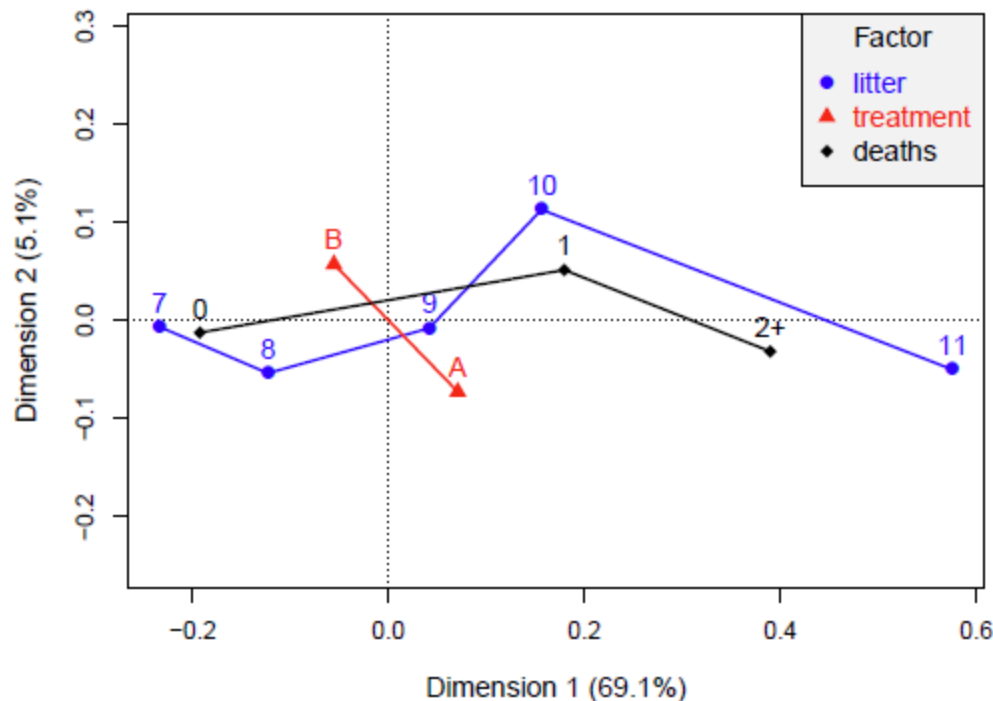
```
mosaic(mice.tab, expected= ~ litter * treatment + deaths)
```



- 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 <- mja(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, \mathbf{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 $\log(\mathbf{T})\mathbf{C}$, where \mathbf{C} is the $K \times K - 1$ matrix of contrasts,

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

Adjacent categories

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

Reference level = 0

In general, any set of $K-1$ $\{1, 0, -1\}$ contrasts can be used

Calculating log odds

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

```
# reshape table to matrix
```

```
T <- matrix(mice.tab,  
            nrow=prod(dim(mice.tab)[1:2]),  
            ncol=dim(mice.tab)[3])  
colnames(T) <- dimnames(mice.tab)[[3]]  
rn <- expand.grid(dimnames(mice.tab)[1:2])  
rownames(T) <- apply(rn, 1, paste, collapse=":")
```

```
C <- matrix(c(1, -1, 0,  
             0, 1, -1), nrow=3)  
lodds <- log(T) %*% C  
colnames(lodds) <- c("0:1", "1:2+")
```

```
> lodds
```

	0:1	1:2+
7:A	1.663	0.788
8:A	1.253	0.336
9:A	0.606	0.182
10:A	0.143	-0.143
11:A	-1.099	-0.348
7:B	1.373	0.999
8:B	1.227	0.754
9:B	0.716	0.788
10:B	0.573	0.201
11:B	-1.099	0.629

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 $\mathbf{n} = \text{vec}(n_{ij\dots})$ be the $N \times 1$ vectorization of the table.
- Then, all log odds and their asymptotic covariance matrix \mathbf{S} can be calculated as:

- $\hat{\psi} = \mathbf{C} \log(\mathbf{n})$
- $\mathbf{S} = \text{Var}[\psi] = \mathbf{C} \text{diag } \mathbf{n}^{-1} \mathbf{C}^T$

where \mathbf{C} is an N -column matrix containing all zeros, except for one $+1$ elements and one -1 elements in each row.

- With strata, \mathbf{C} can be calculated as the Kronecker product

$$\mathbf{C} = \mathbf{C}_R \otimes \mathbf{I}_{K_1} \otimes \mathbf{I}_{K_2} \otimes \dots$$

- Linear models for log odds: $\psi = \mathbf{X}\beta$

Mice data: Log odds

The vcd package contains a general implementation of these ideas:

- `odds()` and `lodds()` : calculate odds or 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")))  
  deaths litter treatment logodds   ASE  
1     0:1      7         A   1.663 0.329  
2     1:2+      7         A   0.788 0.539  
3     0:1      8         A   1.253 0.303  
4     1:2+      8         A   0.336 0.414  
5     0:1      9         A   0.606 0.293  
6     1:2+      9         A   0.182 0.350  
7     0:1     10         A   0.143 0.379  
8     1:2+     10         A  -0.143 0.379  
9     0:1     11         A  -1.099 0.577  
10    1:2+     11         A  -0.348 0.377
```

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
```

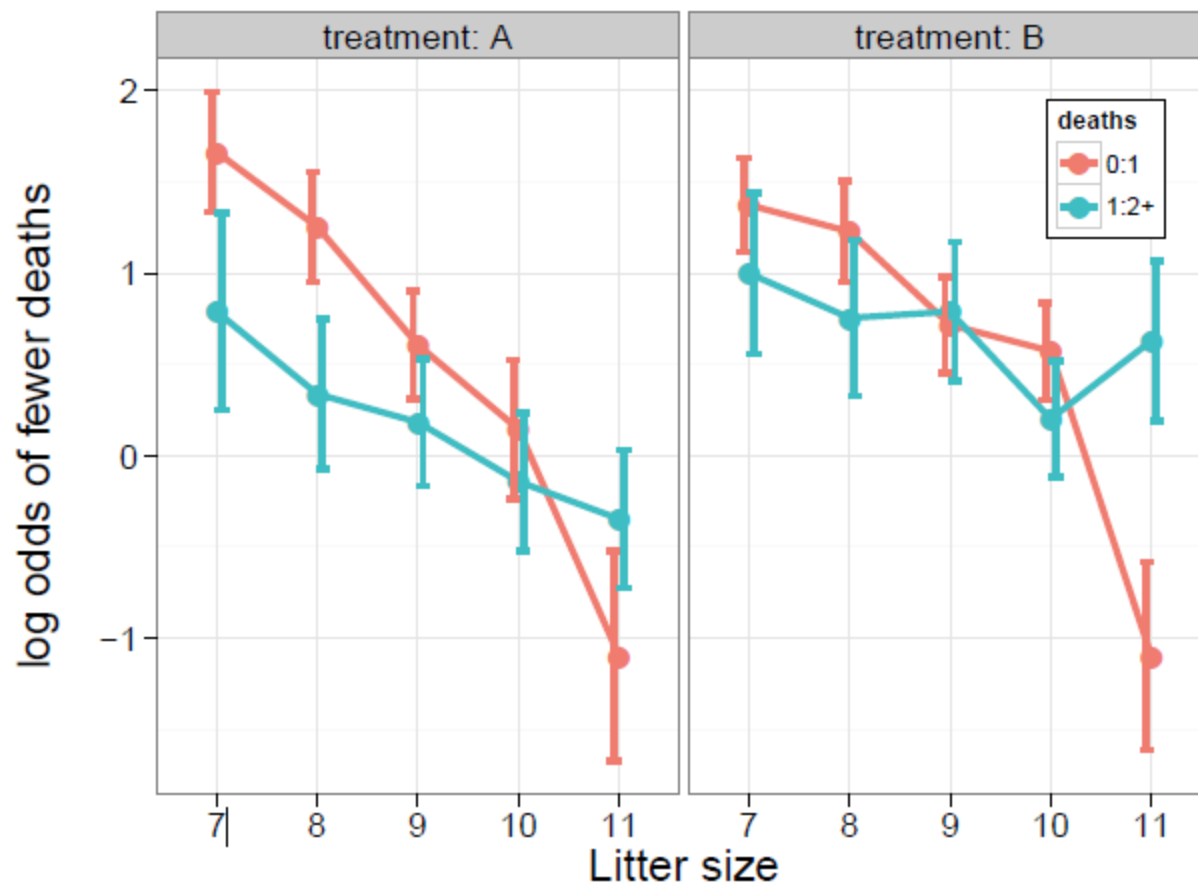
Compare models:

```
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 & models: Data plot

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



Basic plot:

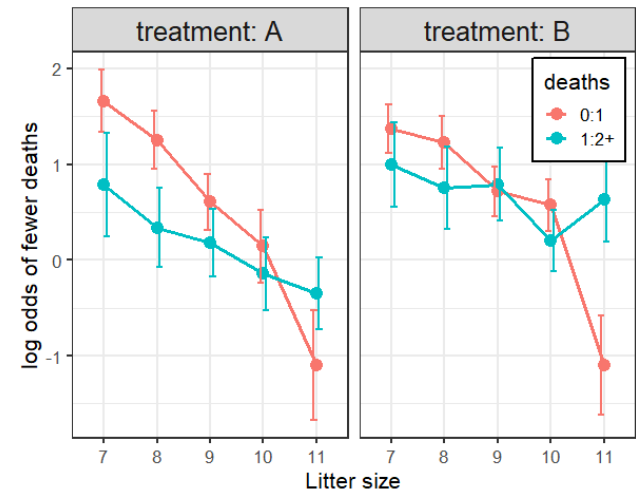
```
gg <- ggplot(mice.lodds, aes(x=litter, y=logodds,  
                             color=deaths, group=deaths)) +  
  geom_point(size=4) +  
  ylab("log odds of fewer deaths") +  
  xlab("Litter size") +  
  theme_bw(base_size = 16) +  
  theme(legend.position = c(.9, .85),  
        legend.background = element_rect(colour = "black")) +  
  facet_grid(. ~ treatment, labeller=label_both) +  
  theme(strip.text = element_text(size = rel(1.2)))
```

Add error bars, dodged

```
bars <- aes(ymin=logodds-ASE,  
            ymax=logodds+ASE)  
gg + geom_line(size=1.2) +  
  geom_errorbar(bars,  
               width=0.25, size=1,  
               position=position_dodge(width=.2))
```

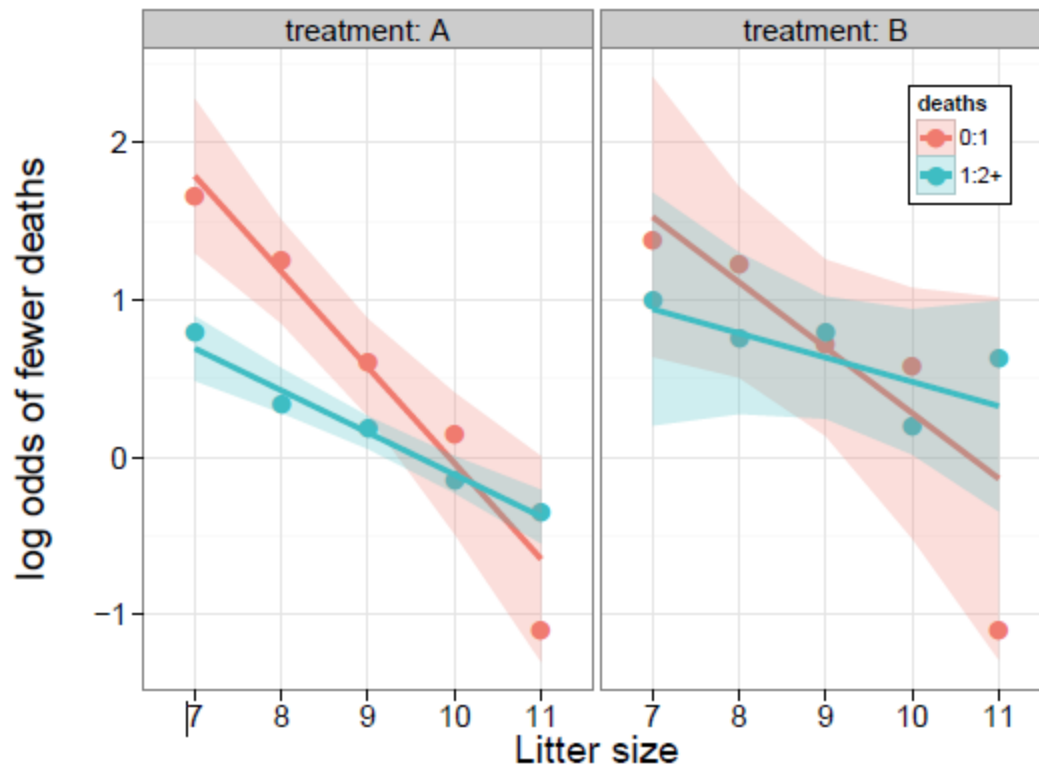
ggplot thinking:

- gg is my basic plot of points
- I can add other layers to it



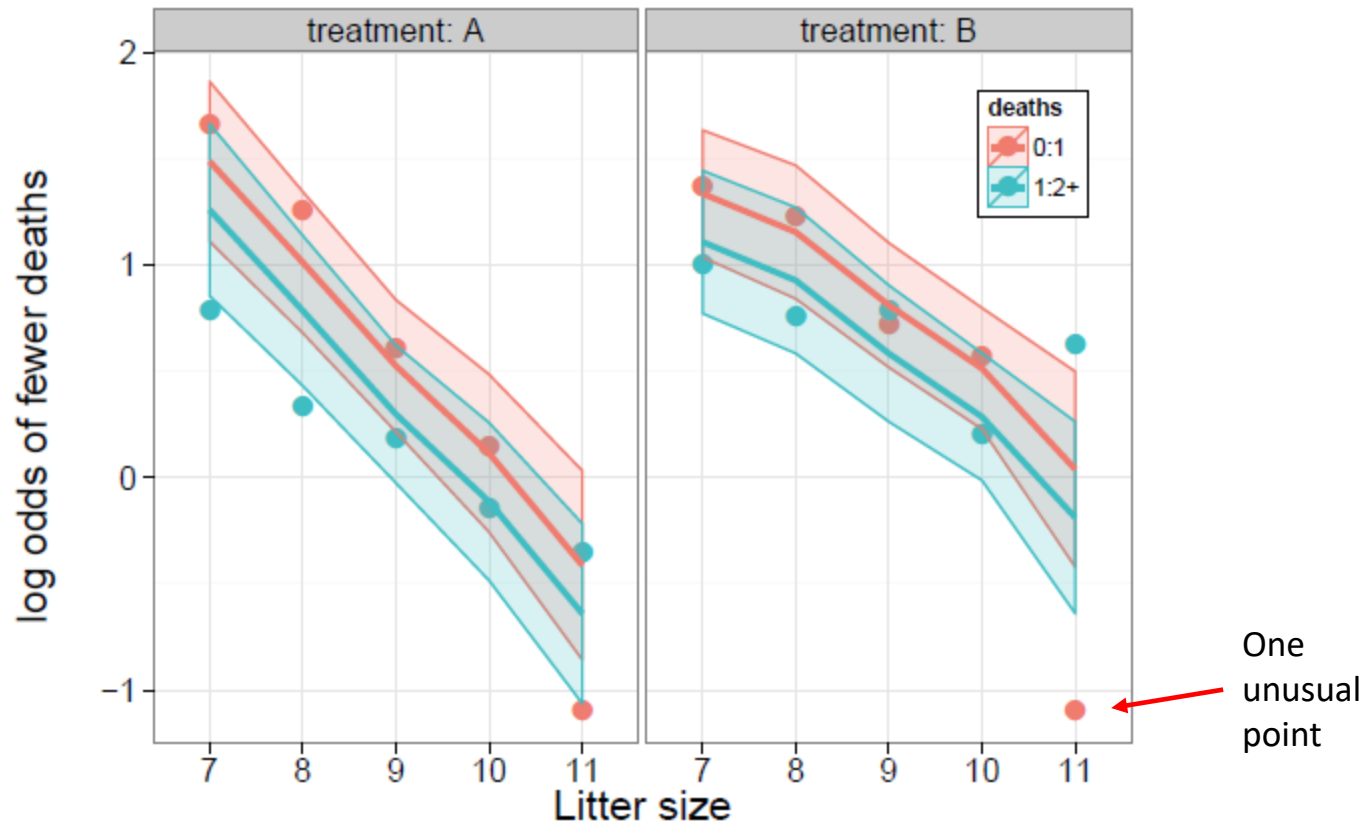
Visualize log odds & models: Smoothing

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



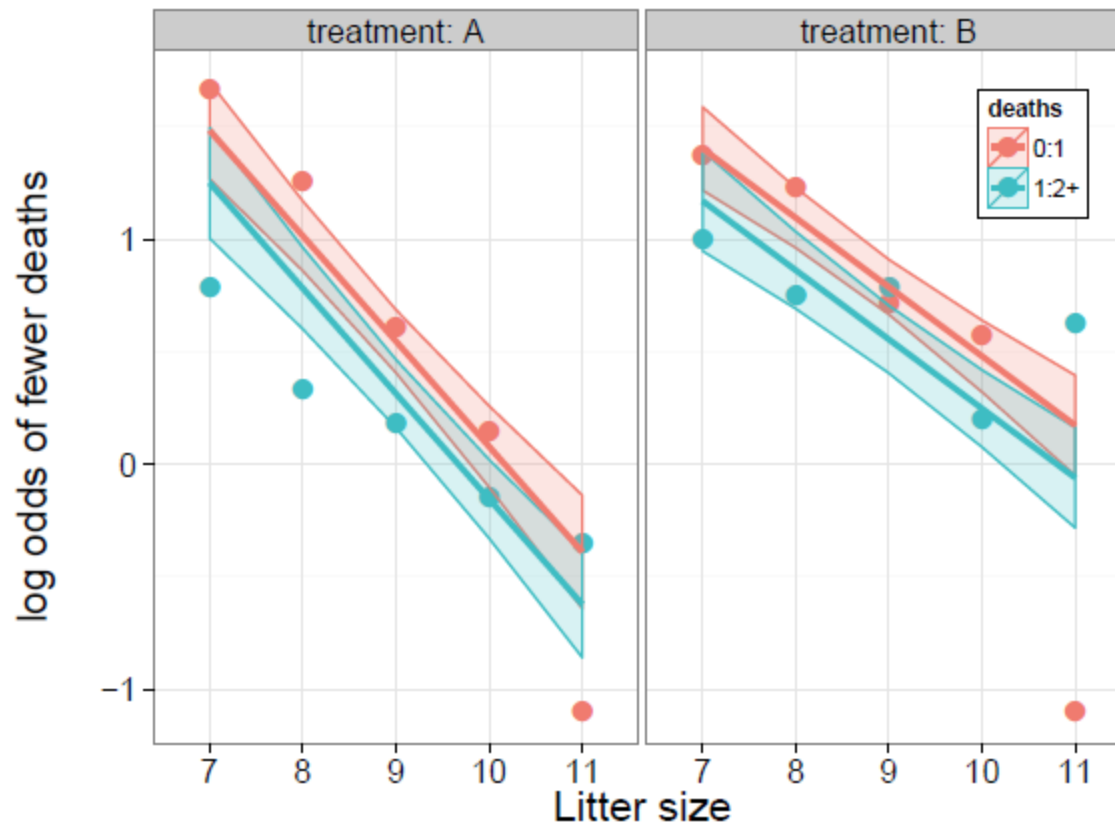
Visualize log odds & models: Data + Model

- Display the fit of the parallel log odds model, $\psi_{ijk}^{AB\bar{C}} = \psi_{ij}^{AB} + \psi_k^C$



Visualize log odds & models: Data + Model

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



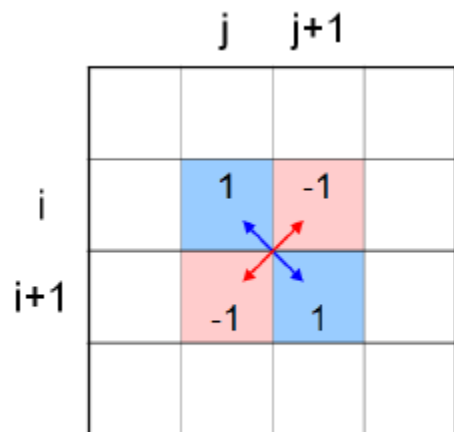
Generalized log odds ratios

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

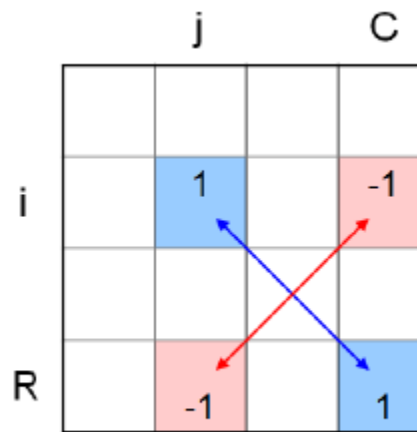
$$\theta_{ij} = \frac{n_{ij}/n_{i+1,j}}{n_{i,j+1}/n_{i+1,j+1}} = \frac{n_{ij} \times n_{i+1,j+1}}{n_{i+1,j} \times n_{i,j+1}}$$

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:

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

- This extends naturally to $\theta_{ij|k}$ 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

$$\log(m_{ij}) = \mu + \lambda_i^A + \lambda_j^B + \gamma \mathbf{a}_i \mathbf{b}_j \equiv \log(\theta_{ij}) = \gamma$$

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

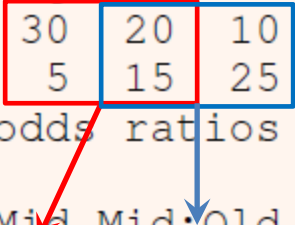
Models for log odds ratios: Computation

- Consider an $R \times C \times K_1 \times K_2 \times \dots$ frequency table $n_{ij\dots}$, with factors $K_1, K_2 \dots$ considered as **strata**.
 - Let $\mathbf{n} = \text{vec}(n_{ij\dots})$ be the $N \times 1$ vectorization of the table.
 - Then, all log odds ratios and their asymptotic covariance matrix \mathbf{S} can be calculated as:
 - $\log(\hat{\theta}) = \mathbf{C} \log(\mathbf{n})$
 - $\mathbf{S} = \text{Var}[\log(\theta)] = \mathbf{C} \text{diag } \mathbf{n}^{-1} \mathbf{C}^T$
- where \mathbf{C} is an N -column matrix containing all zeros, except for two $+1$ elements and two -1 elements in each row.
- With strata, \mathbf{C} can be calculated as $\mathbf{C} = \mathbf{C}_{RC} \otimes \mathbf{I}_{K_1} \otimes \mathbf{I}_{K_2} \otimes \dots$
 - `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:

$$\log(\theta) = \mathbf{C} \log(\mathbf{n}) = \begin{bmatrix} 1 & -1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 1 \end{bmatrix} \log \begin{pmatrix} n_{11} \\ n_{21} \\ n_{12} \\ n_{22} \\ n_{13} \\ n_{23} \end{pmatrix}$$

Models for log odds ratios: Estimation

- A **log odds ratio linear model** for the $\log(\theta)$ is

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

where \mathbf{X} is the design matrix of covariates

- The (asymptotic) ML estimates $\hat{\beta}$ are obtained by GLS via

$$\hat{\beta} = \left(\mathbf{X}^T \mathbf{S}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{S}^{-1} \log(\hat{\theta})$$

where $\mathbf{S} = \text{Var}[\log(\theta)]$ is the estimated covariance matrix

- \implies Standard graphical and diagnostic methods can be adapted to this case.
 - visualization: full-model plots, effect plots, ...
 - diagnostics: influence plots, added-variable plots, ...

Technical note: for simplicity, I use `lm()` for WLS, with $\mathbf{S}^{-1} = \text{diag}(1/\text{ASE}^2)$
Should probably use `nlme::gls()` instead

Example: Breathlessness & wheeze in coal miners

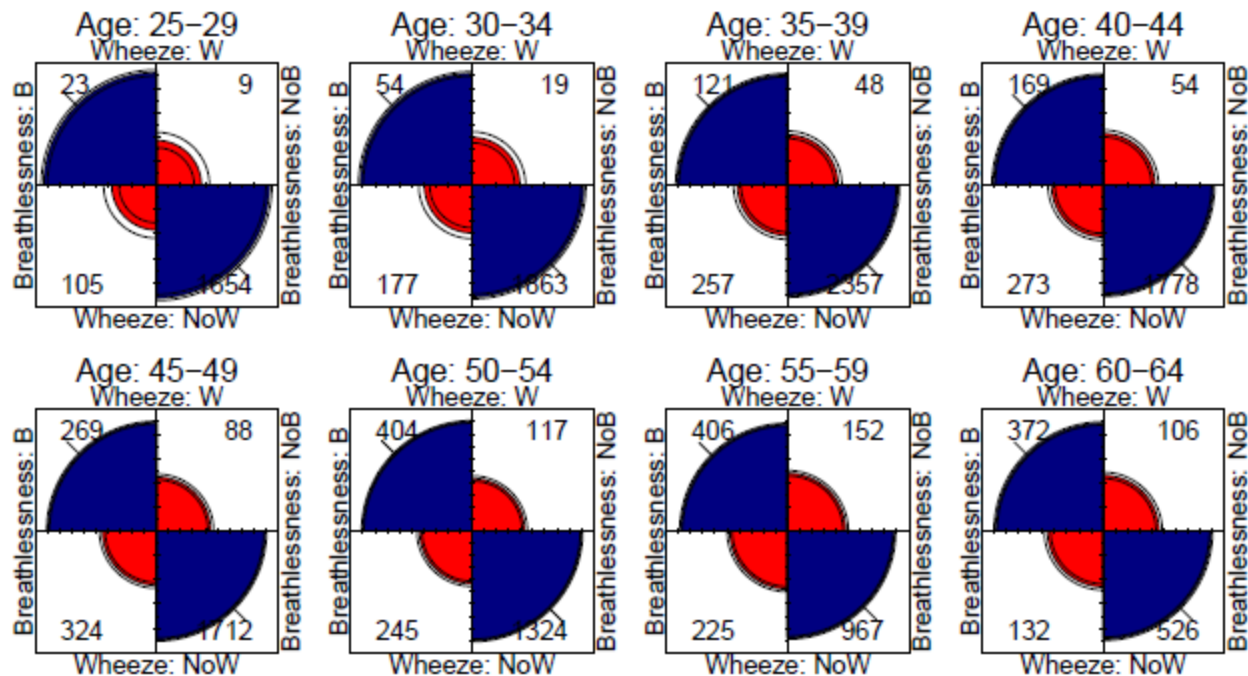
- Ashford & Sowden (1970) gave data on the association between two pulmonary conditions: breathlessness and wheeze, in a large sample of coal miners
- Age is the primary covariate
- How does the association between breathlessness and wheeze vary with age?

```
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, mfc=c(2,4), fontsize=18)
```



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

Coal miners: Log odds & 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?

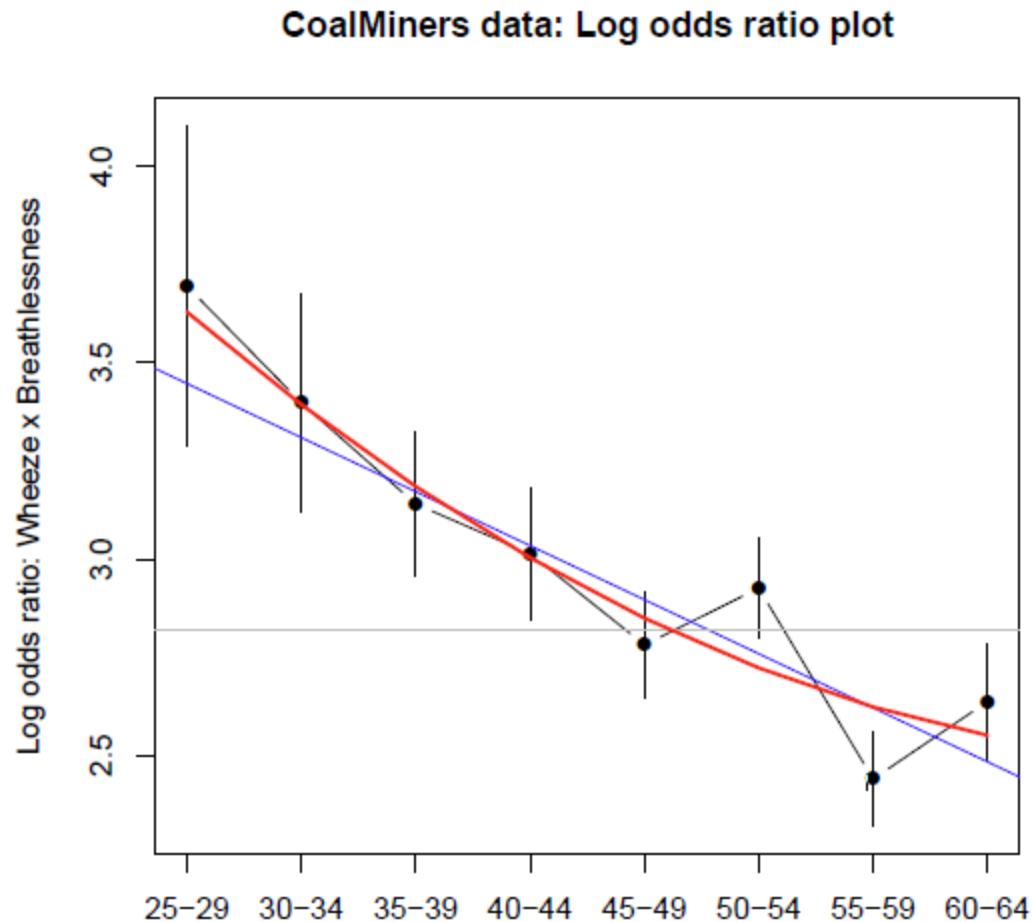
- Uniform association: $\ln(\theta) = \beta_0$
- Linear association: $\ln(\theta) = \beta_0 + \beta_1 \text{ Age}$
- Quadratic association: $\ln(\theta) = \beta_0 + \beta_1 \text{ Age} + \beta_2 \text{ Age}^2$

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, $\ln(\theta) = \beta_0 + \beta_1 \text{ Age}$, 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 (\mathbf{x})
 - the **joint** log odds ratio, ϕ_{12} , vs. \mathbf{x}
- This model has the form

$$\eta(\mathbf{x}) = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_{12} \end{pmatrix} \equiv \begin{pmatrix} \log \text{odds}_1(\mathbf{x}) \\ \log \text{odds}_2(\mathbf{x}) \\ \log \text{OR}_{12}(\mathbf{x}) \end{pmatrix} \equiv \begin{pmatrix} \psi_1 \\ \psi_2 \\ \log \theta_{12} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^\top \beta_1 \\ \mathbf{x}_2^\top \beta_2 \\ \mathbf{x}_{12}^\top \beta_{12} \end{pmatrix}$$

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

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

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

Calculating...

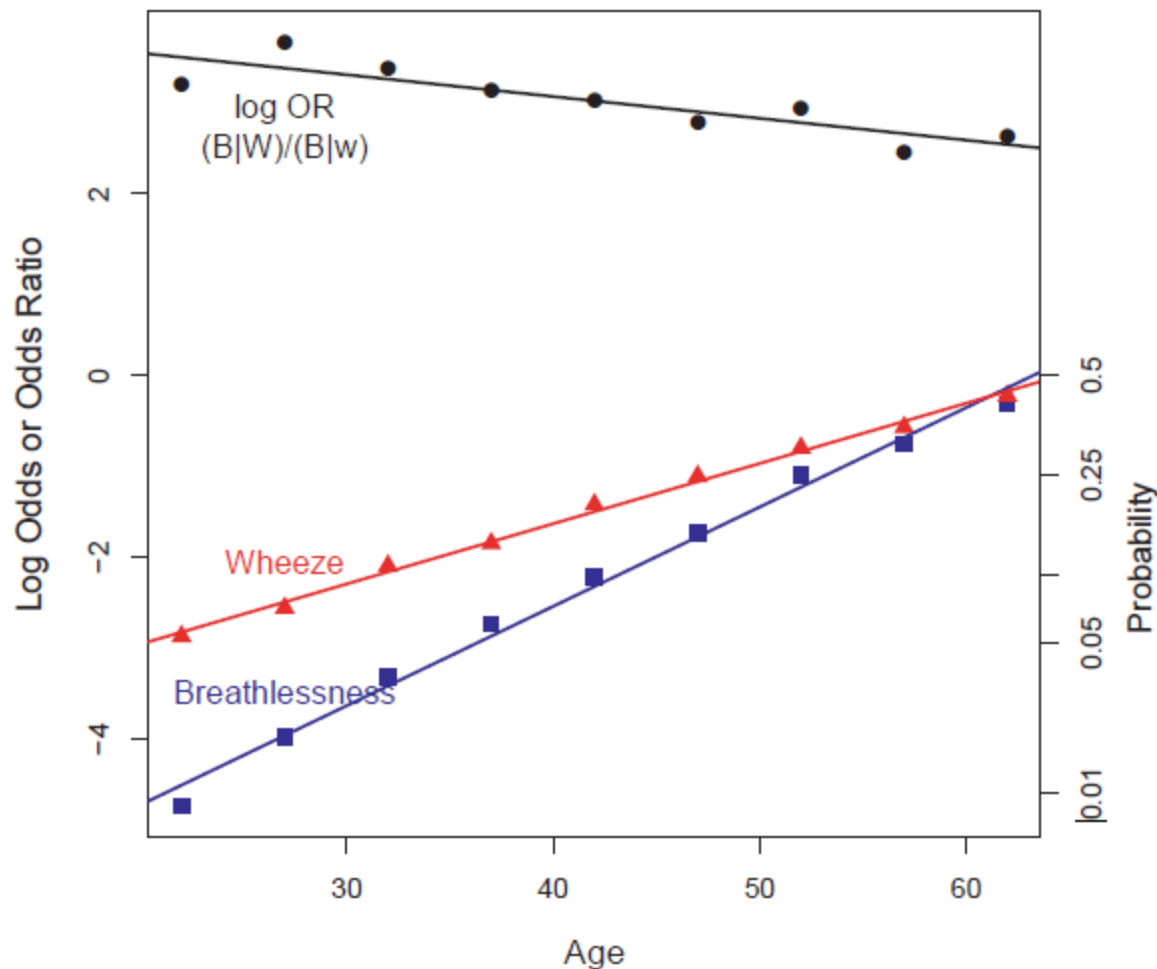
Logits and log odds for a bivariate response can be calculated with `vcdExtra::blogits()`

```
data(coalminers, package = "VGAM")
coalminers <- transform(coalminers, Age = (age - 42) / 5)

logitsCM <- vcdExtra::blogits(coalminers[, 1:4], add = 0.5)
colnames(logitsCM)[1:2] <- c("logitB", "logitW")
logitsCM <- cbind(logitsCM,
                  coalminers[, c("age", "Age")])
logitsCM
```

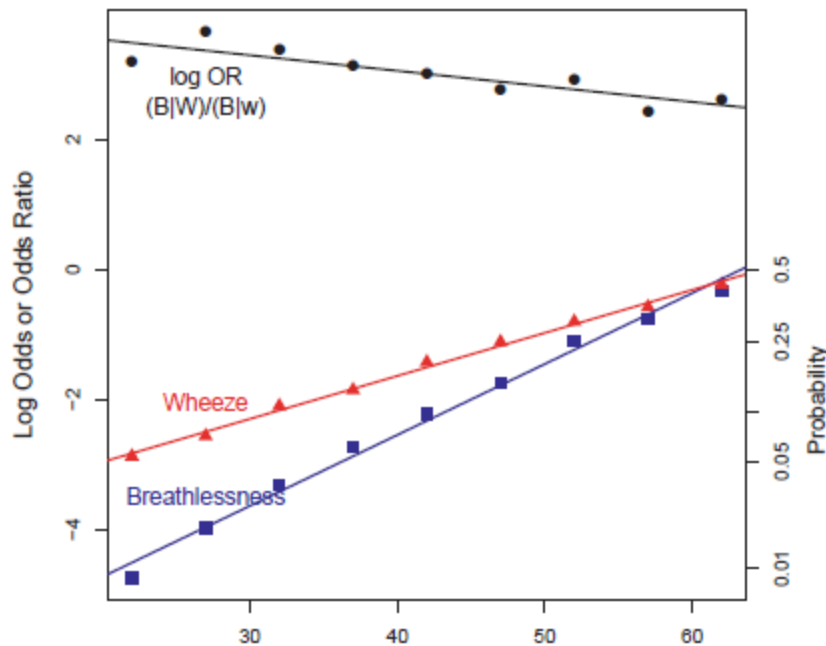
	logitB	logitW	logOR	age	Age
1	-4.736	-2.868	3.20	22	-4
2	-3.977	-2.557	3.66	27	-3
3	-3.317	-2.094	3.38	32	-2
4	-2.733	-1.848	3.13	37	-1
5	-2.215	-1.420	3.01	42	0
6	-1.739	-1.109	2.78	47	1
7	-1.101	-0.797	2.92	52	2
8	-0.758	-0.572	2.44	57	3
9	-0.319	-0.226	2.63	62	4

Linear model for log odds and log odds ratios



Log odds & LORs have similar scales, so it is not terrible to plot them together

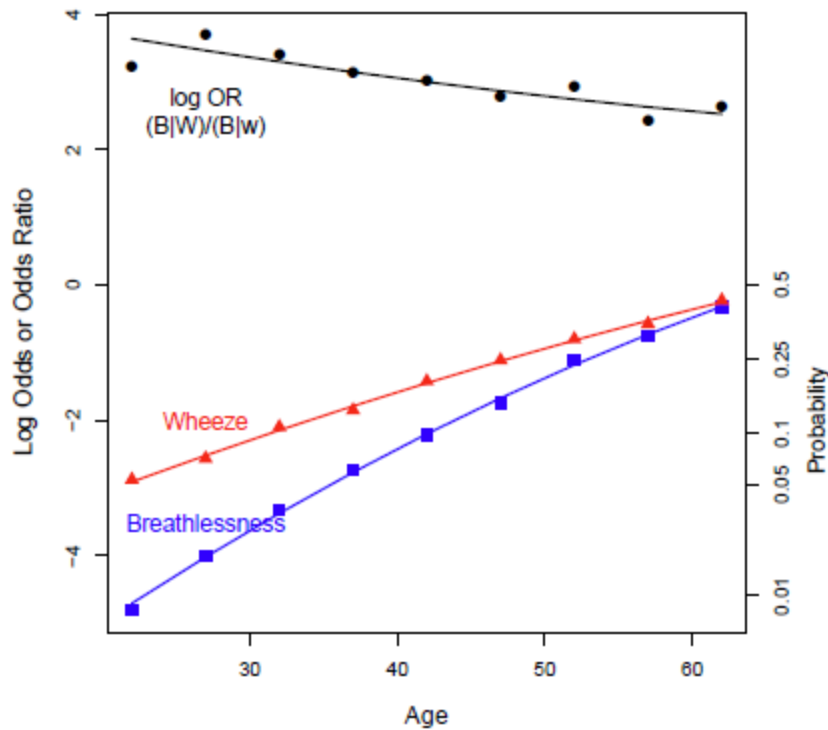
Linear model for log odds and log odds ratios



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

Plotting ...

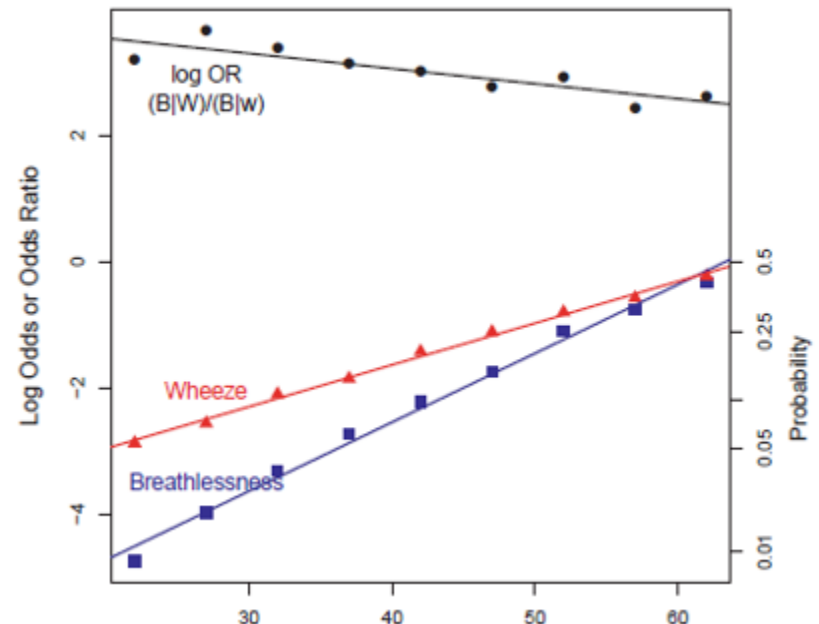
With the data in this form, we can use `matplot()` to plot each column against age

```
matplot(age, logitsCM[, 1:3], type = "p",
        col = col, pch = pch, cex = 1.2, cex.lab = 1.25,
        xlab = "Age", ylab = "Log Odds or Odds Ratio")
abline(lm(logitsCM[,1] ~ age), col = col[1], lwd = 2)
abline(lm(logitsCM[,2] ~ age), col = col[2], lwd = 2)
abline(lm(logitsCM[,3] ~ age), col = col[3], lwd = 2)
```

To plot the quadratic fit, simply use

```
lm(logitsCM[,1] ~ poly(age,2))
```

But: this is NOT a model. It simply fits each set of odds separately



Fitting: VGAM::vglm()

VGAM::vglm() can fit a wide class of models for a vector of multivariate responses

- The family `binom2()` is used for bivariate logistic models
- An argument `zero=` allows the logit or odds ratio submodels to be constrained to intercept-only

```
cm.vglm <- vglm(cbind(nBnW, nBW, BnW, BW) ~ Age,  
               binom2.or(zero = NULL), data = coalminers)  
  
exp(coef(cm.vglm, matrix = TRUE))
```

	logitlink(mu1)	logitlink(mu2)	loglink(oratio)	
(Intercept)	0.104	0.226	20.530	
Age	1.673	1.385	0.877	

Each 5 years of age:

- Multiplies odds of breathlessness by 1.67, a 67% increase
- Multiplies odds of wheeze by 1.38, a 38% increase
- Multiplies the OR for association by 0.88, a 12 % decrease

Plotting the model fit

VGAM::fitted() returns the **fitted** values on the probability scale

VGAM::depvar() returns the **observed** values on the probability scale

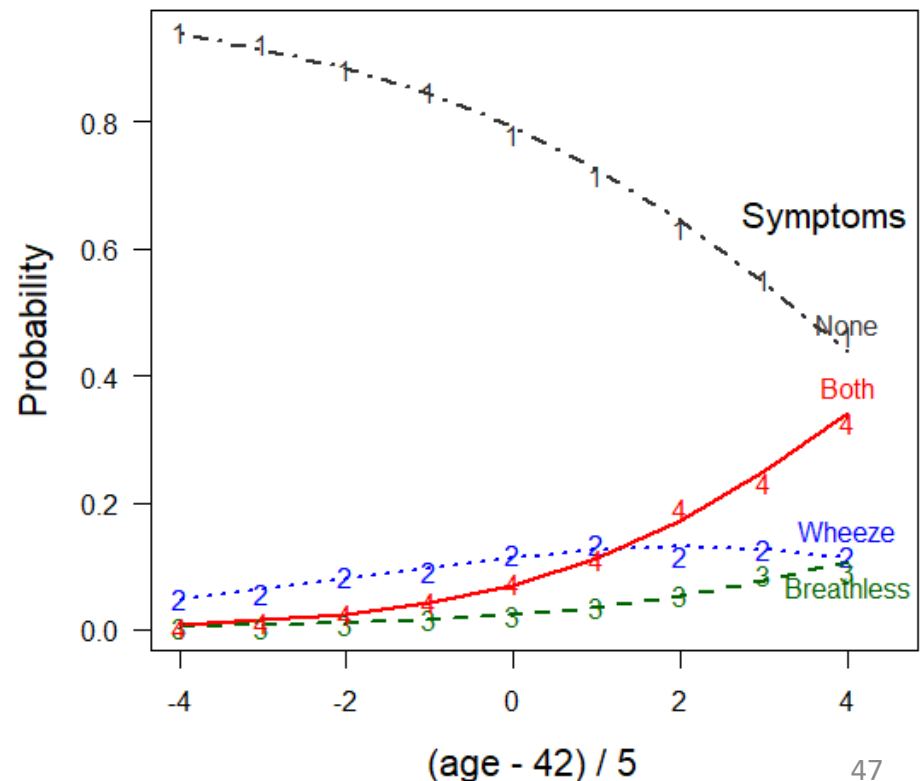
```
> P <- fitted(cm.vglm)
> colnames(P) <- c("bw", "bW",
                  "Bw", "BW")

> P
  bw    bW    Bw    BW
1 0.94 0.049 0.0046 0.0085
2 0.91 0.064 0.0070 0.0148
3 0.88 0.080 0.0105 0.0254
4 0.84 0.097 0.0158 0.0428
5 0.79 0.114 0.0239 0.0704
...
```

You can get these on the logit scale using the inverse logit function, `qlogis()`

```
LP <- qlogis(P)
LY <- qlogis(Y)
```

The plot is made using `matplot()`



Other possibilities

We can also model the relations with age as a quadratic, cubic, ...

```
cm.vglm2 <- vglm(cbind(nBnW, nBW, BnW, BW) ~ poly(Age,2),  
  binom2.or(zero = NULL), data = coalminers)
```

VGAM also implements vector **generalized additive** models, fit using `vgam()`

```
cm.vgam <- vgam(cbind(nBnW, nBW, BnW, BW) ~ s(Age, df = 2),  
  binom2.or(zero = NULL), data = coalminers)
```


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+)

Education	Attitude	Age Memory	15–24		25–39		40+	
			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

Attitudes: 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)
```

- 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)

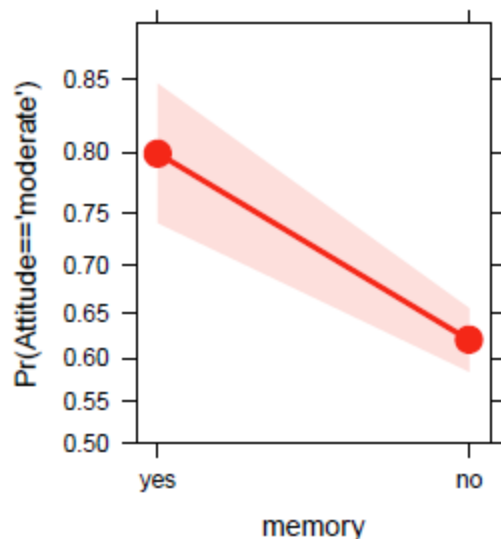
## Analysis of Deviance Table (Type II tests)
##
## Response: attitude
##          LR Chisq Df Pr(>Chisq)
## memory      29.5  1  5.6e-08 ***
## education    50.3  2  1.2e-11 ***
## age          0.6  2    0.73
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

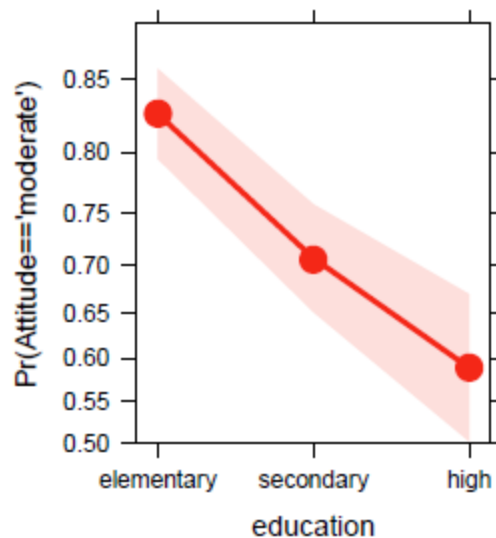
Attitude: Effect plots

- **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

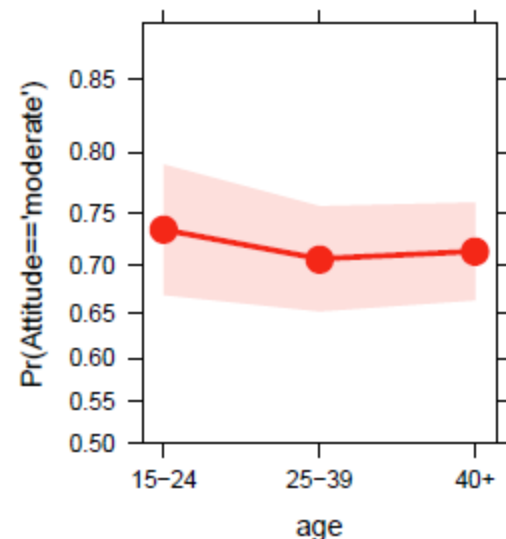
memory effect plot



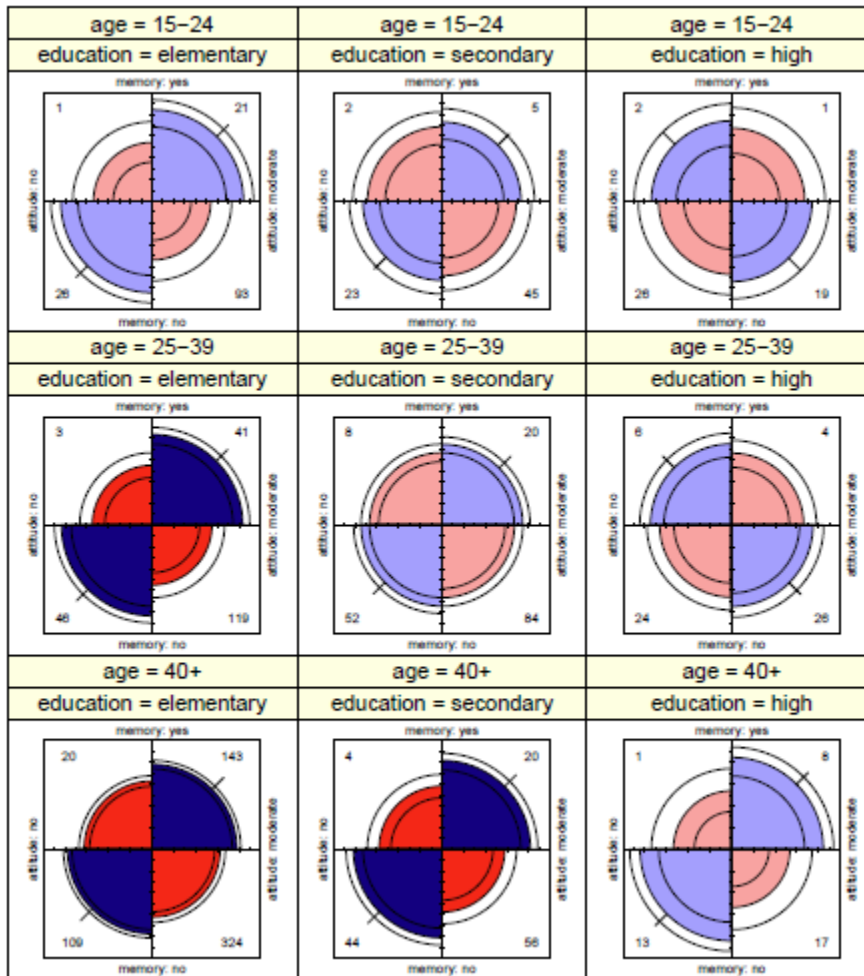
education effect plot



age effect plot



Association of attitude with memory: Fourfold plots



How does the association of attitude and memory vary with education and age?

Each fourfold plot visualizes the log odds ratio between them

What's going on here?

Log odds ratio plot

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

```
## log odds ratios for memory and attitude by age, education
```

```
##
```

```
##          education
```

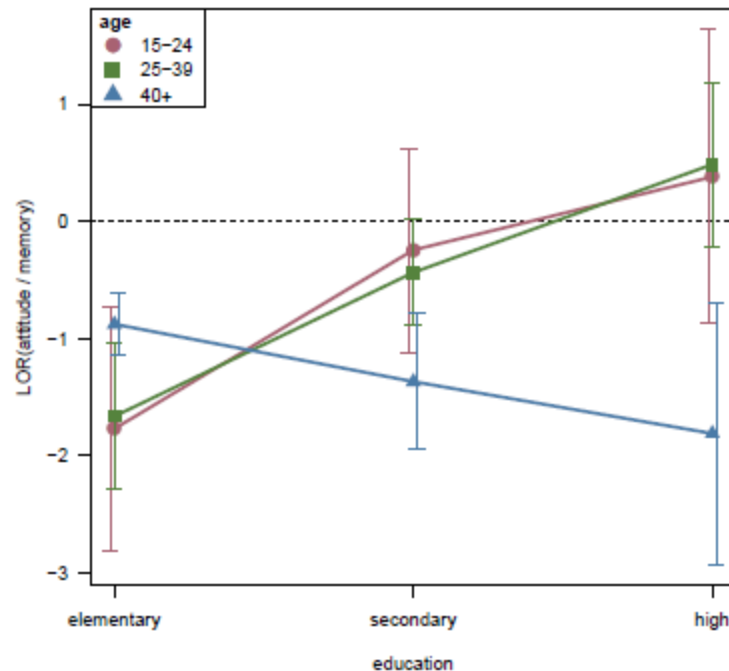
```
## age      elementary secondary    high
```

```
## 15-24    -1.7700    -0.2451    0.3795
```

```
## 25-39    -1.6645    -0.4367    0.4855
```

```
## 40+      -0.8777    -1.3683   -1.8112
```

log odds ratios for attitude and memory by education, age



- Structure now completely clear
- Little difference between younger groups
- Opposite pattern for the 40+
- Fit an LOR model to confirm appearances (SEs large)!

Summary

- Logit models for a binary response generalize readily to a polytomous response
 - → Models for log odds, familiar interpretation
 - Handles 3+ way table, ordinal variables
 - Simple plots for interpretation
- Generalized odds ratios handle bivariate responses
 - Simple linear models for LOR
 - Easy to model log odds for each response and the LOR simultaneously
 - Easy to visualize results