Multinomial Regression Models

Objectives:

- Multinomial distribution and likelihood
- Ordinal data: Cumulative link models (POM).
- Ordinal data: Continuation models (CRM).

Models for Multinomial Data

Example Data:

- Wisconsin Study of Diabetic Retinopathy (WESDR).

- Diabetic retinopathy is one of the leading causes of blindness in people aged 20-75 years in the US.

- Disease characterized by appearance of small hemorrhages in the retina which progress and lead to severe visual loss.

- **Disease severity**:
  - None
  - Mild
  - Moderate
  - Proliferative

- **Covariates**:
  - duration of diabetes
  - glycosolated hemoglobin
  - diastolic blood pressure
  - age at diagnosis

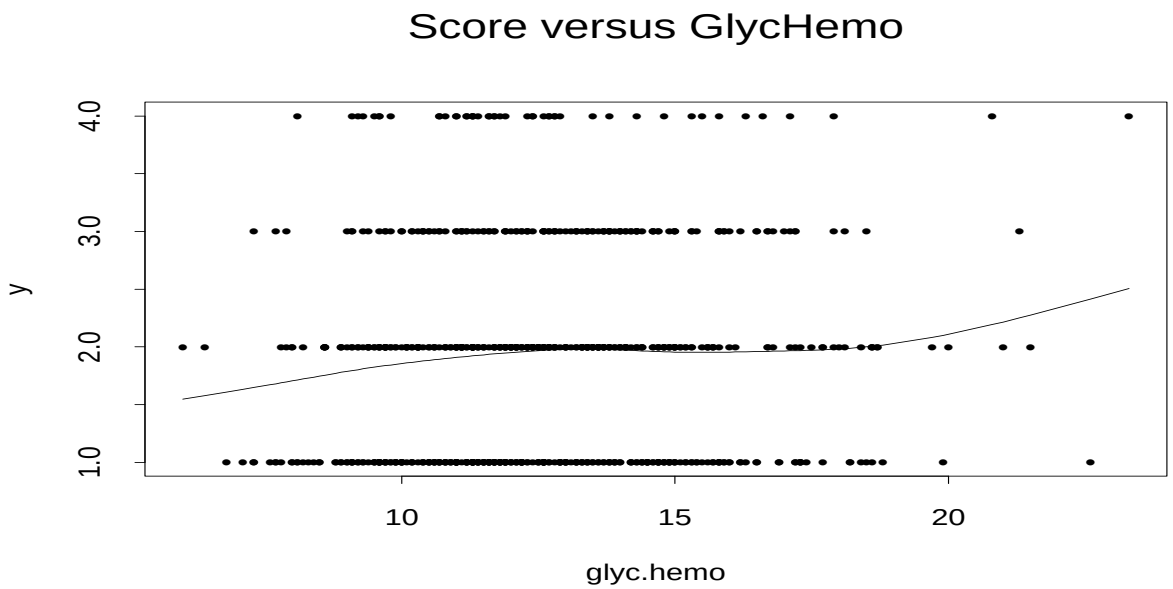
Models for Ordinal Data

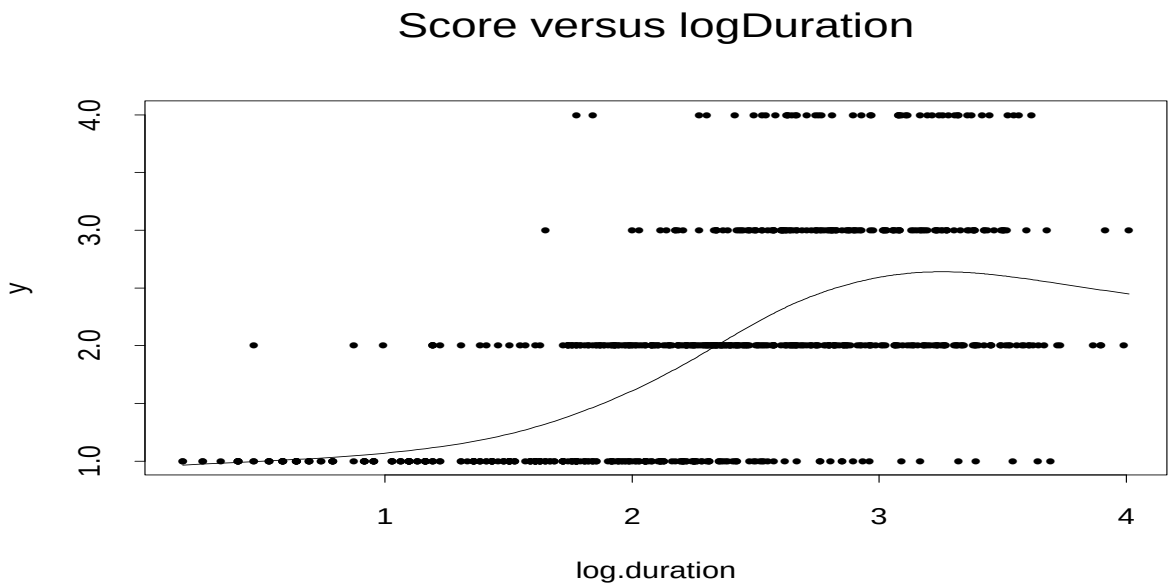
Example Data:

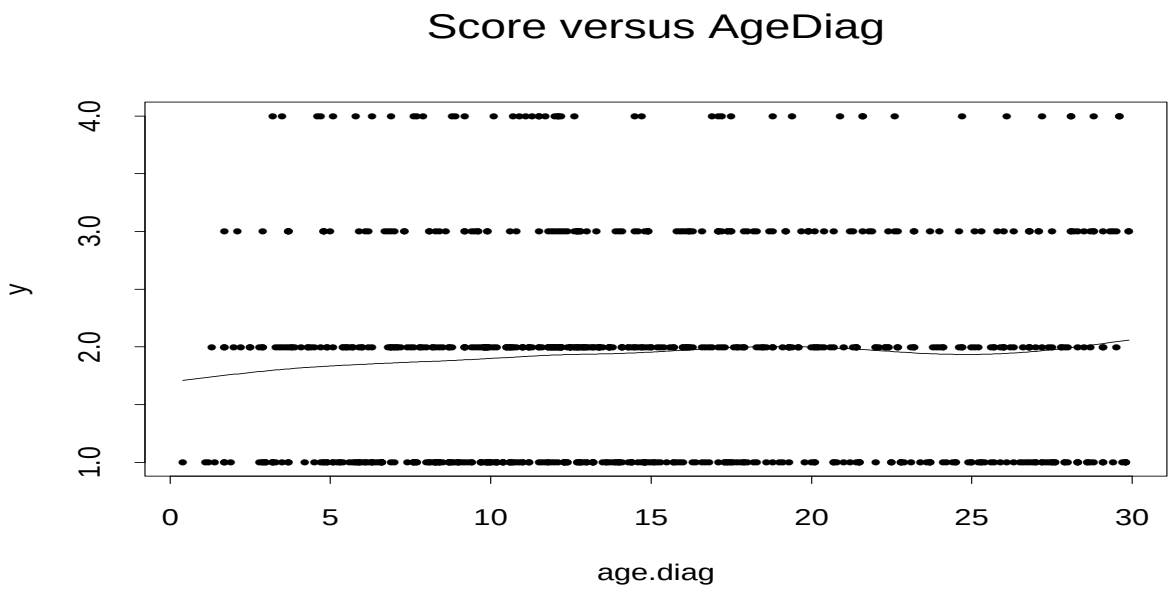
Q: How does the distribution of disease severity vary as a function of covariates?

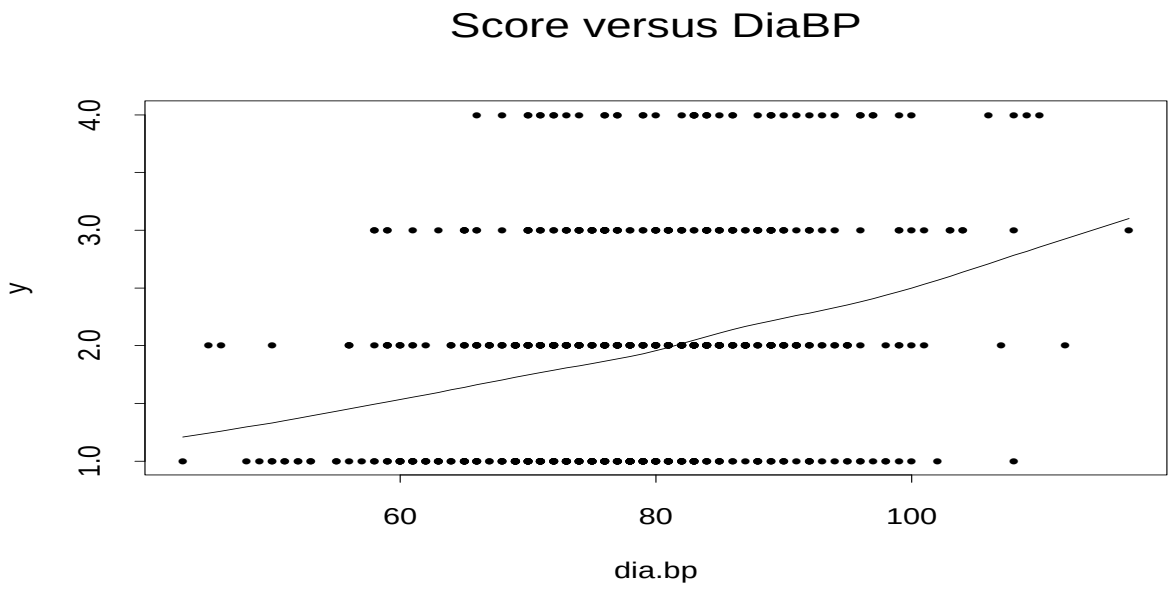
Q: Possible approaches to analysis?

Ideas?


Score versus GlycHemo
y
1.0
2.0
3.0
4.0
10
15
20
glyc.hemo


Score versus logDuration
y
1.0
2.0
3.0
4.0
1
2
3
4
log.duration


Score versus AgeDiag
y
1.0
2.0
3.0
4.0
0
5
10
15
20
25
30
age.diag


Score versus DiaBP
y
1.0
2.0
3.0
4.0
60
80
100
dia.bp

Multinomial Response Models

◦ Common categorical outcomes take more than two levels:

- Pain severity = low, medium, high
- Conception trials = 1, 2 if not 1, 3 if not 1-2

◦ The basic probability model is the multi-category extension of the Bernoulli (Binomial) distribution – multinomial.

◦ Univariate outcome with multivariate representation:

- Let $O_i \in [1, 2, 3, \ldots C]$ be a categorical outcome.
- Let Y_{ij} be the indicator $Y_{ij} = \mathbf{1}(O_i = j)$, $j = 1, 2, \ldots (C-1)$.

$$O_i = j \quad \Leftrightarrow \quad Y_{ij} = 1 \quad \text{and} \quad Y_{ik} = 0 \quad \forall k \neq j$$

$$O_i \quad \Leftrightarrow \quad \boldsymbol{Y}_i = \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{iC} \end{pmatrix}$$

◦ Probability given by

$$P(O_i) = P(\boldsymbol{Y}_i) = \pi_{i1}^{Y_{i1}} \pi_{i2}^{Y_{i2}} \ldots (1 - \sum_{k=1}^{C-1} \pi_{ik})^{(1-\sum_{k=1}^{C-1} Y_{ik})}$$

$$E[Y_{ij}] = \pi_{ij}$$

$$\mathsf{cov}[Y_{ij}, Y_{ik}] = \begin{cases} -\pi_{ij}\pi_{ik} & j \neq k \\ \pi_{ij}(1 - \pi_{ij}) & j = k \end{cases}$$

What is Ordinal Data?

- Categories (fixed number) that are ordered much like the *ordinal* numbers:

  first, *second*, ...

- It doesn't make sense to talk about a distance between the categories:
  - ⋆ HIGH, MEDIUM, LOW
  - ⋆ NEVER, RARELY, SOMETIMES, ALWAYS
  - ◦ e.g. "Drachman Class"

Ordinal Data

Q: Is this type of data really that common?

A: Lincoln Moses (Stanford) surveyed vol. 306 of NEJM and found that 32/168 articles contained ordered categorical data!

"We found no indication that any of the authors in vol. 306 used an analysis that takes account of the ordering." (p. 447)

Ordinal? Binary? Continuous?

Q: Do we lose much information by GROUPING (categorizing) a measurement?

Example: rather than analysis of BMI as a continuous measurement we recode into: non-obese; obese; severely obsese.

Q: Do we gain much information by using the ordered categories over just DICHOTOMIZING?

Example: rather than analysis of BMI **categories** we use the indicator of severely obsese.

Ordinal? Binary? Continuous?

Evaluation:

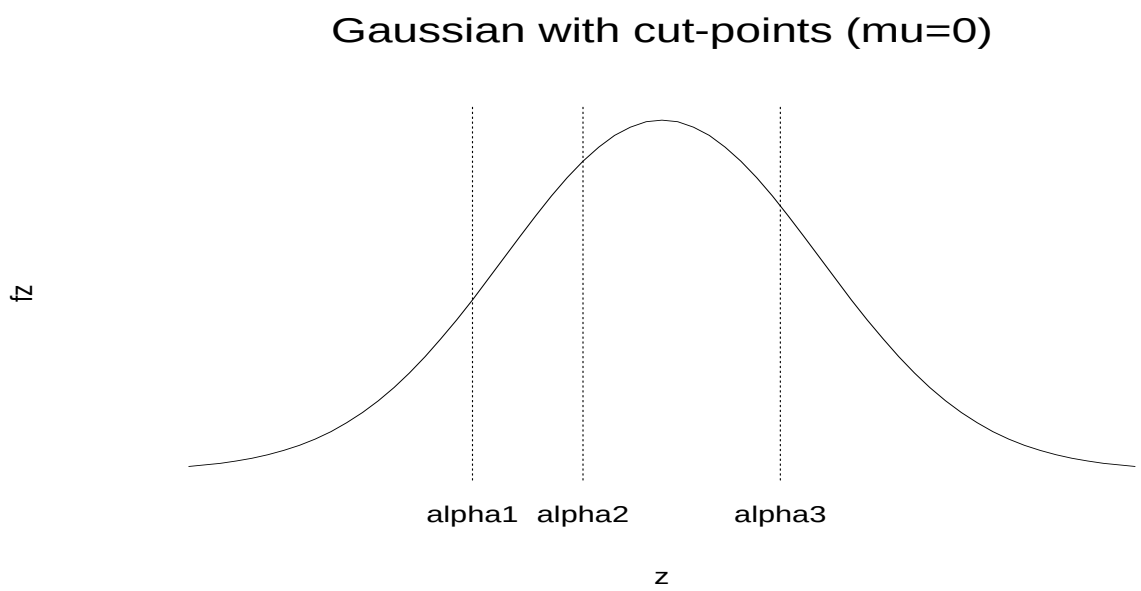
1) Assume an underlying Gaussian: $O_i \sim \mathcal{N}(\mu, 1)$.

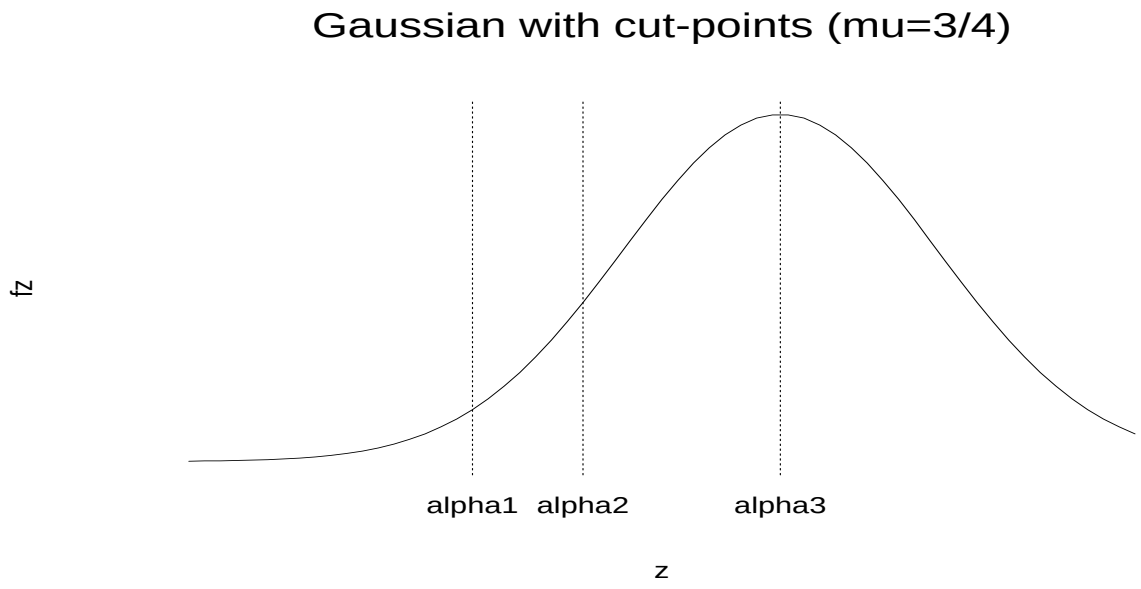
2) Define CUTPOINTS $\boldsymbol{\alpha} \in \mathcal{R}^C$.

Define: $Y_{ij} = \mathbf{1}(O_i \in (\alpha_{j-1}, \alpha_j])$

$$\begin{aligned} E(Y_{ij}) &= P(O_i \in (\alpha_{j-1}, \alpha_j]) \\ &= P(O_i \leq \alpha_j) - P(O_i \leq \alpha_{j-1}) \\ \pi_{ij} &= \Phi(\alpha_j - \mu) - \Phi(\alpha_{j-1} - \mu) \end{aligned}$$

$\boldsymbol{Y}_i$ represents a multinomial random variable with $m = 1$, $\boldsymbol{\pi}_i$.


Gaussian with cut-points (mu=0)
fz
alpha1
alpha2
alpha3
z


Gaussian with cut-points (mu=3/4)
fz
alpha1
alpha2
alpha3
z

Ordinal versus Continuous?

Given a sample of $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n$ we can obtain the MLE for μ (and we'll just assume $\boldsymbol{\alpha}$ is known).

We obtain (derive this yourself!):

$$\mathcal{I}_n = \sum_i \sum_j \frac{\partial \pi_{ij}}{\partial \mu} \frac{1}{\pi_{ij}} \frac{\partial \pi_{ij}}{\partial \mu}$$

If we assume that the $\boldsymbol{Y}_i$'s are i.i.d then we can drop the subscript i and we can compare the information from the categorized, or grouped, data Y_i to the information in the continuous O_i:

Asymptotic relative efficiency:

$$\frac{\mathsf{var}(\widehat{\mu}_O)}{\mathsf{var}(\widehat{\mu}_{\boldsymbol{Y}})} = \frac{1/n}{\left[n \sum_j \frac{\partial \pi_j}{\partial \mu} \frac{1}{\pi_j} \frac{\partial \pi_j}{\partial \mu}\right]^{-1}} = \mathcal{I}_1$$

Q: How does the efficiency depend on C, the number of categories?

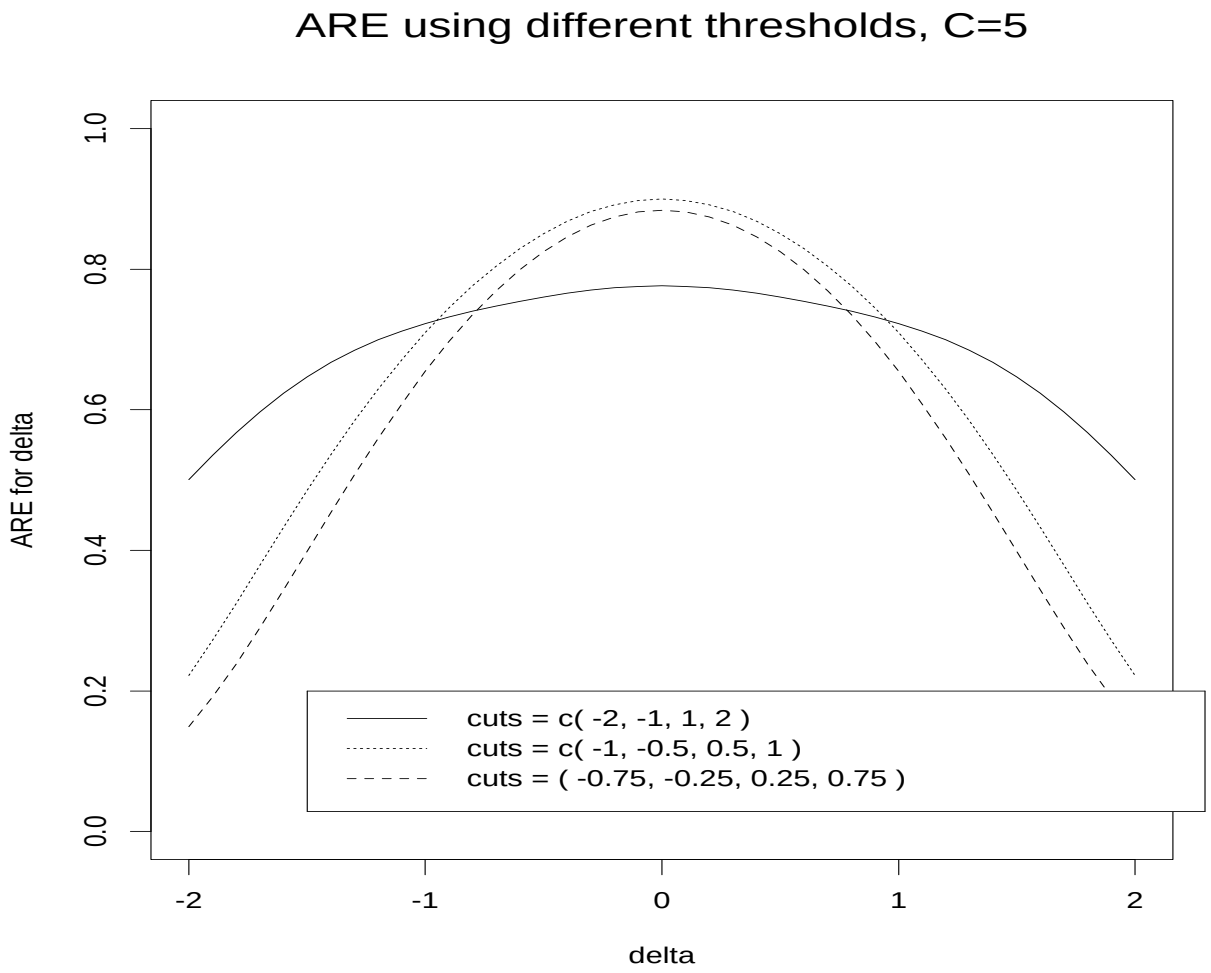
Q: How does the efficiency depend on the locations α_j?

Efficiency versus C

For this I actually used the MLE's variance based on estimating the location difference $\mu_1 - \mu_2 = \Delta$ and the cut-point parameters.

For the cutpoints I used the $1/(C+1)$-tiles
of $\mathcal{N}(0,1)$ shifted by $\Delta/2$ (ie. cut-points are between the two locations).

C	ARE
1	0.565
2	0.727
3	0.800
4	0.841
5	0.867
6	0.884
7	0.897
8	0.907
9	0.914
10	0.920


ARE using different thresholds, C=5
ARE for delta
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
delta
cuts = c(-2, -1, 1, 2)
cuts = c(-1, -0.5, 0.5, 1)
cuts = (-0.75, -0.25, 0.25, 0.75)

Some references on grouping data :

Connor R.J. (1972), "Grouping for testing trends in categorical data", *Journal of the American Statistical Association*, 67: 601-604.

- binary & continuous $\Rightarrow 2 \times k$
- Efficiency from 65% - 97% for testing (k=2,...,6)

Armstrong B.G. and Sloan M. (1989), "Ordinal regression models for epidemiologic data", *American Journal of Epidemiology*, 129: 191-204.

- POM versus logistic regression.
- A smart dichotomization yields 75-80% efficiency.

Sankey S.S and Weissfeld L.A. (1998), "A study of the effect of dichotomizing ordinal data upon modeling", *Communications in Statistics, Part B – Sim & Comp*, 27: 871-887.

Probability Model = Multinomial

- Fixed number of categories (C).
- Fixed sample size (m).

Insect species $1, 2, \ldots, C$. Sites $s_1, s_2, \ldots, s_m$.	
Scheme 1	When first insect enters trap it shuts. All traps left until catch 1.
Scheme 2	If insect enters it can't get out. All traps left for 24 hours.
$\boldsymbol{Y}_{1,i} = (0, 0, 0, 1, 0)$	$\boldsymbol{Y}_{2,i} = (0, 4, 1, 0, 2)$

Probability Model = Multinomial

Given Y_{ij} = number of type j out of m_i; $\sum_j Y_{ij} = m_i$

$$P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, \ldots, Y_{iC} = y_{iC}) = \frac{m_i!}{y_{i1}! y_{i2}! \ldots y_{iC}!} \prod_{j=1}^{C} \pi_{ij}^{y_{ij}}$$

In this representation the parameter π_{ij} can be defined via:

- Given $m_i = 1$:
- Given $m_i > 1$:

Some properties of multinomial

1) $w_1, w_2, \ldots, w_C$ independent $w_i \sim \mathcal{P}(\lambda_i)$.

Then $(w_1, w_2, \ldots, w_C \mid \sum_j w_j = m) \sim$

$$\text{mult}(m, \frac{\lambda_1}{\sum_j \lambda_j}, \frac{\lambda_2}{\sum_j \lambda_j}, \ldots, \frac{\lambda_C}{\sum_j \lambda_j})$$

2) If $\boldsymbol{Y} \sim \text{mult}(m, \boldsymbol{\pi})$

$$\Longrightarrow \quad Y_j \sim \text{binomial}(m, \pi_j)$$

3) $\text{cov}(Y_j, Y_k) \;=\; -m\pi_j\pi_k$ for $j \neq k$.

4) REPRODUCIBLE: Let $\boldsymbol{Y} = (\boldsymbol{Y}_1, \boldsymbol{Y}_2)$ then

$$(\boldsymbol{Y}_1, \mathsf{sum}(\boldsymbol{Y}_2)\ ) \sim \mathsf{mult}(m, \boldsymbol{\pi}_1, \mathsf{sum}(\boldsymbol{\pi}_2)\ )$$

5) REPRODUCIBLE for CONDITIONALS:

Condition on Y_j: $(Y_1, Y_2, \ldots, (-j), \ldots, Y_C \mid Y_j = t) \sim$

$$\mathsf{mult}(m-t, \frac{\pi_1}{1-\pi_j}, \frac{\pi_2}{1-\pi_j}, \ldots, (-j), \ldots, \frac{\pi_C}{1-\pi_j})$$

6) Define cumulative totals $Z_j = \sum_{k=1}^{j} Y_k$, $\gamma_j = \sum_{k=1}^{j} \pi_k$:

$$\begin{aligned} Z_j &\sim \mathsf{binomial}(m, \gamma_j) \\ Z_i \mid Z_j = z_j,\ i < j &\sim \mathsf{binomial}(z_j, \frac{\gamma_i}{\gamma_j}) \end{aligned}$$

Multinomial as Vector Exponential Family

$$
\begin{aligned}
P(\boldsymbol{Y}_i) &= \left[\prod_{j=1}^{C-1} \pi_{ij}^{Y_{ij}}\right] (1 - \sum_{k}^{C-1} \pi_{ik})^{(1-\sum_{k}^{C-1} Y_{ik})} \\
&= \exp\left[\sum_{j}^{C-1} Y_{ij} \cdot \log(\pi_{ij}) + \right. \\
&\qquad \left. (1 - \sum_{k}^{C-1} Y_{ik}) \cdot \log(1 - \sum_{k}^{C-1} \pi_{ik})\right]
\end{aligned}
$$

Multinomial as Vector Exponential Family

$$
\begin{aligned}
P(\boldsymbol{Y}_i) &= \exp\left[\sum_{j}^{C-1} Y_{ij} \cdot \log(\pi_{ij}/\pi_{iC}) + \log(\pi_{iC})\right] \\
&= \exp\left[\boldsymbol{Y}_i^T \boldsymbol{\theta}_i - b(\boldsymbol{\theta}_i)\right]
\end{aligned}
$$

Note: we have "dropped" the last category.

where

$$\theta_{ij} = \log(\pi_{ij}/\pi_{iC})$$

$$b(\boldsymbol{\theta}_i) = -\log(\pi_{iC}) = \log[1 + \sum_{k}^{C-1} \exp(\theta_{ik})] \quad (\textit{verify!})$$

$$\frac{\partial}{\partial \theta_{ij}} b(\boldsymbol{\theta}_i) = \pi_{ij} \quad (\textit{verify!})$$

$$\frac{\partial^2}{\partial \theta_{ij} \partial \theta_{ik}} b(\boldsymbol{\theta}_i) = \begin{cases} -\pi_{ij}\pi_{ik} & k \neq j \\ \pi_{ij}(1 - \pi_{ij}) & k = j \end{cases}$$

Models for Ordinal Data

dia.bp80 | y	1	2	3	4	RowTotl
0	200	144	59	16	419
	0.477	0.344	0.141	0.038	0.58
1	75	126	69	31	301
	0.249	0.419	0.229	0.103	0.42
ColTotl	275	270	128	47	720
	0.382	0.375	0.178	0.065	

Test for independence of all factors

Chi^2 = 45.46906 d.f.= 3 (p=7.354828e-10)

```
Call:  crosstabs( ~ dia.bp80 + (y > 1))
dia.bp80|(y > 1)
        |FALSE  |TRUE   |RowTotl|
--------+-------+-------+-------+
0       |200    |219    |419    |
        |0.48   |0.52   |0.58   |
--------+-------+-------+-------+
1       | 75    |226    |301    |
        |0.25   |0.75   |0.42   |
--------+-------+-------+-------+
ColTotl|275    |445    |720    |
--------+-------+-------+-------+
Test for independence of all factors
        Chi^2 = 38.62691 d.f.= 1 (p=5.130671e-10)

   odds ratio = 2.74 ,  ( 1.983 ,  3.786 )

   log-odds ratio = 1.008 ,  s.e.= 0.165
```

```
Call:  crosstabs( ~ dia.bp80 + (y > 2))
dia.bp80|(y > 2)
        |FALSE  |TRUE   |RowTotl|
--------+-------+-------+-------+
0       |344    | 75    |419    |
        |0.82   |0.18   |0.58   |
--------+-------+-------+-------+
1       |201    |100    |301    |
        |0.67   |0.33   |0.42   |
--------+-------+-------+-------+
ColTotl|545    |175    |720    |
--------+-------+-------+-------+
Test for independence of all factors
        Chi^2 = 22.35406 d.f.= 1 (p=2.267331e-06)
        Yates' correction not used

   odds ratio = 2.276 ,  ( 1.611 ,  3.215 )

   log-odds ratio = 0.822 ,  s.e.= 0.176
```

```
Call:  crosstabs( ~ dia.bp80 + (y > 3))
dia.bp80|(y > 3)
        |FALSE  |TRUE   |RowTotl|
--------+-------+-------+-------+
0       |403    | 16    |419    |
        |0.962  |0.038  |0.58   |
--------+-------+-------+-------+
1       |270    | 31    |301    |
        |0.897  |0.103  |0.42   |
--------+-------+-------+-------+
ColTotl|673    |47     |720    |
-------+-------+-------+-------+
Test for independence of all factors
        Chi^2 = 12.05597 d.f.= 1 (p=0.0005162686)

   odds ratio = 2.848 ,  ( 1.539 ,  5.269 )

   log-odds ratio = 1.047 ,  s.e.= 0.314
```

Proportional Odds Model

- model the cumulative indicators: $Z_{ij} = \mathbf{1}(O_i \leq j)$

$$\begin{aligned} \text{Define} \quad \gamma_{ij} &= P(O_i \leq j) \\ &= E(Z_{ij}) \end{aligned}$$

$$\boxed{\textbf{GLM}} \quad g(\gamma_{ij}) = \beta_{(0,j)} - \boldsymbol{X}_i\boldsymbol{\beta}$$

- For

a single j this is equivalent to logistic regression when we use a logit link.

- The model with the logit link is called the **Proportional odds model**.

- Another common link is the complementary log-log link, $g(\gamma) = \log(-\log(1-\gamma))$, which is the discrete time proportional hazards model. (more on this later)

- There is an ordering constraint:

$$\beta_{(0,1)} \leq \beta_{(0,2)} \cdots \leq \beta_{(0,C-1)}$$

Score equations

The underlying model is the multinomial which yields

$$\frac{\partial \log \mathcal{L}}{\partial \boldsymbol{\beta}} = \sum_i \left(\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\beta}} \right)^T \boldsymbol{\Sigma}_i^{-1} \left(\boldsymbol{Y}_i - \boldsymbol{\pi}_i \right)$$

Note: $\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 0 \\ & \vdots & & & \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix}$

Score equations

Properties of $\mathbf{L}$:

$$\begin{aligned} \mathbf{L}\boldsymbol{Y}_i &= \boldsymbol{Z}_i \\ \mathbf{L}\boldsymbol{\pi}_i &= \boldsymbol{\gamma}_i \end{aligned}$$

$$\frac{\partial \log \mathcal{L}}{\partial \boldsymbol{\beta}} = \sum_i \left(\frac{\partial \mathbf{L}\boldsymbol{\pi}_i}{\partial \boldsymbol{\beta}} \right)^T \left(\mathbf{L}\boldsymbol{\Sigma}_i \mathbf{L}^T \right)^{-1} \mathbf{L} \left(\boldsymbol{Y}_i - \boldsymbol{\pi}_i \right)$$

$$\frac{\partial \log \mathcal{L}}{\partial \boldsymbol{\beta}} = \sum_i \left(\frac{\partial \boldsymbol{\gamma}_i}{\partial \boldsymbol{\beta}} \right)^T \left[\mathsf{cov}(\boldsymbol{Z}_i) \right]^{-1} \left(\boldsymbol{Z}_i - \boldsymbol{\gamma}_i \right)$$

Reparameterization of POM

- There are several ways this model can be parameterized.
- We have used indicators for $O_{ij} \leq j$ and:

$$\text{logit}(\gamma_{ij}) = \beta_{(0,j)} - \boldsymbol{X}_i\boldsymbol{\beta}$$

- Alternatively we may model $Z^*_{ij} = 1(O_{ij} > j)$ and:

$$E(Z^*_{ij}) = 1 - \gamma_{ij} = \mu^*_{ij}$$

$$\begin{aligned} \text{logit}(\mu^*_{ij}) &= -\beta_{(0,j)} + \boldsymbol{X}_i\boldsymbol{\beta} \\ &= \beta^*_{(0,j)} + \boldsymbol{X}_i\boldsymbol{\beta} \end{aligned}$$

```
 Ordinal Regression
 Cumulative Link GLM

     link = logit

 -2*maxL  = 1712.95

 Regression Coefficients and S.E.'s
           estimate       S.E.          z
      c1  0.1102717 0.09516786   1.158707
      c2 -1.5880006 0.11420678 -13.904609
      c3 -3.1492179 0.17159607 -18.352506
dia.bp80  0.9425179 0.14279883   6.600319
```

```
 Ordinal Regression
 Cumulative Link GLM

     link = logit

 -2*maxL  = 1694.77

 Regression Coefficients and S.E.'s
          estimate        S.E.          z
    c1  0.64781282 0.081852364   7.914406
    c2 -1.07600308 0.088720335 -12.128032
    c3 -2.66667489 0.152165220 -17.524865
dia.bp  0.05209068 0.006700984   7.773586
```

```
Ordinal Regression
Cumulative Link GLM

    link = logit

-2*maxL  = 1337.79

Regression Coefficients and S.E.'s
                 estimate        S.E.            z
          c1  0.179575318 0.106663795   1.6835639
          c2 -2.374525515 0.142756804 -16.6333614
          c3 -4.245398609 0.204729189 -20.7366552
   glyc.hemo  0.095083429 0.029752798   3.1957810
log.duration  2.126978406 0.135675906  15.6769058
    age.diag  0.009541847 0.010055159   0.9489504
      dia.bp  0.045025747 0.007348497   6.1272053
```

Ordinal Data Regression

- Models for ordinal data can be viewed as "simultaneous regressions" for the cumulative indicators $Z_{ij} = \mathbf{1}(O_i \leq j)$.

- Probability model is the multinomial.

- Models for ordinal data use the order information but do not assign numbers to the outcome levels. Models are also invariant to collapsing the categories (ie. combine 4th and 5th levels).

Ordinal Data Regression

- Models for ordinal data are useful as something intermediate to just dichotomizing the data, or to assigning scores and using linear regression.

- **Fact** For a single binary covariate $X = 0/1$ the score test using the POM model is the Wilcoxon! (see MN exercise 5.10). Thus, these models can be thought of as models for the rank of the data.

Models for Multinomial Data

- Another approach to the analysis of ordered categorical data is to model **conditional** expectations.

- Infection in 0-6 months.
  Infection in 6-12 months **given** uninfected at 6.
  Infection in 12-18 months **given** uninfected at 12.

- Salmon dies before point #1.
  Salmon dies before point #2 **given** past #1.
  Salmon dies before point #3 **given** past #2.

Multinomial Framework

Define:

- $Y_{ij} = \mathbf{1}(O_i = j)$
- $\boldsymbol{Y}_i = \mathsf{vec}(Y_{ij})$
- $H_{ij} = 1 - \sum_{k=1}^{j-1} Y_{ik}$ $\qquad\qquad$ $H_{i1} = 1$

Multinomial Framework

Model:

$$
\begin{aligned}
E[Y_{ij} \mid H_{ij} = 1] &= \mu_{ij} \\
&= \frac{\pi_{ij}}{1 - \gamma_{i(j-1)}}
\end{aligned}
$$

$$
\begin{aligned}
g(\mu_{ij}) &= \alpha_j + \boldsymbol{X}_{ij}\boldsymbol{\beta}_j \\
&\qquad j = 1, 2, \ldots, C-1
\end{aligned}
$$

Note : $E[Y_{ij} \mid H_{ij}] = \mu_{ij} \cdot H_{ij}$

Example:

	$j = 1, 2, \ldots, 5$				
	1	2	3	4	5
Y_{ij}	0	0	1	0	0
H_{ij}	1	1	1	0	0
μ_{ij}	μ_{i1}	μ_{i2}	μ_{i3}	μ_{i4}	μ_{i5}
$E[Y_{ij} \mid H_{ij}]$	μ_{i1}	μ_{i2}	μ_{i3}	0	0

Multinomial – CRM Models

- Models for $E(Y_{ij} \mid H_{ij})$ are sometimes called "continuation ratio models" (Agresti 1990).

- These conditional models are also known as discrete-time survival models, in particular, if we assume that the ordered categories are time intervals, $\mathcal{G}_j = (t_{j-1}, t_j]$, and O_i records the interval that the time, T_i falls in:

$$O_i = j \quad \Leftrightarrow \quad T_i \in \mathcal{G}_j$$

Multinomial – CRM Models

PH Assumption: $P(T_i > t) = \exp[-\Lambda(t)]$

$$= \exp[-\Lambda_0(t)\exp(X_i\boldsymbol{\beta})]$$

then

$$E(Y_{ij} \mid H_{ij} = 1) = \mu_{ij}$$

$$\log(-\log(1-\mu_{ij})) = \alpha_j + \boldsymbol{X}_i\boldsymbol{\beta}$$

$$\alpha_j = \log[\Lambda_0(t_j) - \Lambda_0(t_{j-1})]$$

Note: The PH model uses a common $\boldsymbol{\beta}$ rather than a coefficient specific to each time interval, $\boldsymbol{\beta}_j$.

CRM Model

- The multinomial probability factors:

$$
\begin{aligned}
P(\boldsymbol{Y}_i = \boldsymbol{y}_i) &= \prod_{j=1}^{C} \pi_{ij}^{y_{ij}} \\
&= P(y_{i1})P(y_{i2} \mid y_{i1})P(y_{i3} \mid y_{i1}, y_{i2}) \ldots \\
&\quad \ldots P(y_{iC} \mid y_{ik}\ k < C)
\end{aligned}
$$

CRM Model

$$P(\boldsymbol{Y}_i = \boldsymbol{y}_i) = \prod_{j=1}^{C-1} \left(\frac{\pi_{ij}}{1-\gamma_{i(j-1)}} \right)^{y_{ij}} \left(1 - \frac{\pi_{ij}}{1-\gamma_{i(j-1)}} \right)^{h_{ij}-y_{ij}}$$

$$= \prod_{j=1}^{C-1} \left(\frac{\pi_{ij}}{1-\gamma_{i(j-1)}} \right)^{y_{ij}} \left(\frac{1-\gamma_{ij}}{1-\gamma_{i(j-1)}} \right)^{h_{ij}-y_{ij}}$$

Note:

$$\text{logit}\left(\frac{\pi_{ij}}{1-\gamma_{i(j-1)}}\right) = \log\left(\frac{\pi_{ij}}{1-\gamma_{ij}}\right)$$

Comments:

- This likelihood factorization implies that we obtain the MLE for this model using the pairs $(Y_{ij}, H_{ij} = 1)$ and treating them as independent Bernoulli random variables.

- For any (j, k)

$$E\left[(Y_{ij} - \mu_{ij} H_{ij})(Y_{ik} - \mu_{ik} H_{ik})\right] \quad = \quad 0$$

Q: justification for this?

WESDR Example:

```
#
nsubjects <- nrow( data )
#
wisc.data <- data.frame(
        y = wisc.all.data[,"r.level"],
        glyc.hemo = wisc.all.data[,"glyc.hemo"],
        log.duration = log( wisc.all.data[,"duration"] ),
        age.diag = wisc.all.data[,"age.diag"],
        dia.bp = wisc.all.data[,"dia.bp"]
  )
print( wisc.data[1:5,] )
#
###
###  Construct CRM data...
###
#
##### (1) construct the pairs (Y,H)
#
```

```
ncuts <- max(wisc.data$y) - 1
print( paste("ncuts =", ncuts, "\n\n") )
y.crm <- NULL
h.crm <- NULL
id <- NULL
for( j in 1:nsubjects ){
  yj <- rep( 0, ncuts )
  if( wisc.data$y[j] <= ncuts ) yj[ wisc.data$y[j] ] <- 1
  hj <- 1 - c(0,cumsum(yj)[1:(ncuts-1)])
  y.crm <- c( y.crm, yj )
  h.crm <- c( h.crm, hj )
  id <- c( id, rep(j,ncuts) )
}
```

```
##### (2) construct the intercepts
#
level <- factor( rep(1:ncuts, nsubjects ) )
int.mat <- NULL
for( j in 1:ncuts ){
  intj <- rep( 0, ncuts )
  intj[ j ] <- 1
  int.mat <- cbind( int.mat, rep( intj, nsubjects) )
}
dimnames(int.mat) <- list( NULL, paste("Int",c(1:ncuts),sep="") )
#
##### (3) expand the X's
#
glyc.hemo <- rep( wisc.data$glyc.hemo , rep(ncuts,nsubjects) )
log.duration <- rep( wisc.data$log.duration , rep(ncuts,nsubjects) )
age.diag <- rep( wisc.data$age.diag , rep(ncuts,nsubjects) )
dia.bp <- rep( wisc.data$dia.bp , rep(ncuts,nsubjects) )
#
print( cbind( y.crm, h.crm, level, int.mat, glyc.hemo, log.duration,
                age.diag, dia.bp)[1:15,] )
#
```

```
##### (4) drop the H=0 and build dataframe
#
keep <- h.crm==1
wisc.crm.data <- data.frame(
                          id = id[keep],
                          y = y.crm[keep],
                          level = level[keep],
                          glyc.hemo = glyc.hemo[keep],
                          log.duration = log.duration[keep],
                          age.diag = age.diag[keep],
                          dia.bp = dia.bp[keep] )
#
print( wisc.crm.data[1:15,] )
```

Original Data:

	y	glyc.hemo	log.duration	age.diag	dia.bp
1	3	13.7	2.332144	29.9	89
2	2	13.5	2.292535	18.6	90
3	3	13.8	2.747271	28.7	85
4	1	8.4	3.540959	20.1	99
5	3	12.8	3.131137	29.3	84

Expanded Data:

	id	y.crm	h.crm	level	Int1	Int2	Int3	glyc.hemo
[1,]	1	0	1	1	1	0	0	13.7
[2,]	1	0	1	2	0	1	0	13.7
[3,]	1	1	1	3	0	0	1	13.7
[4,]	2	0	1	1	1	0	0	13.5
[5,]	2	1	1	2	0	1	0	13.5
[6,]	2	0	0	3	0	0	1	13.5
[7,]	3	0	1	1	1	0	0	13.8
[8,]	3	0	1	2	0	1	0	13.8
[9,]	3	1	1	3	0	0	1	13.8
[10,]	4	1	1	1	1	0	0	8.4
[11,]	4	0	0	2	0	1	0	8.4
[12,]	4	0	0	3	0	0	1	8.4
[13,]	5	0	1	1	1	0	0	12.8
[14,]	5	0	1	2	0	1	0	12.8
[15,]	5	1	1	3	0	0	1	12.8

	id	y	level	glyc.hemo	log.duration	age.diag	dia.bp
1	1	0	1	13.7	2.332144	29.9	89
2	1	0	2	13.7	2.332144	29.9	89
3	1	1	3	13.7	2.332144	29.9	89
4	2	0	1	13.5	2.292535	18.6	90
5	2	1	2	13.5	2.292535	18.6	90
6	3	0	1	13.8	2.747271	28.7	85
7	3	0	2	13.8	2.747271	28.7	85
8	3	1	3	13.8	2.747271	28.7	85
9	4	1	1	8.4	3.540959	20.1	99
10	5	0	1	12.8	3.131137	29.3	84
11	5	0	2	12.8	3.131137	29.3	84
12	5	1	3	12.8	3.131137	29.3	84
13	6	0	1	13.0	3.258097	21.8	70
14	6	0	2	13.0	3.258097	21.8	70
15	6	1	3	13.0	3.258097	21.8	70

WESDR Models

```
#####
##### Model fitting...
#####
#
options( contasts="contr.treatment" )
#
table( wisc.crm.data$level )
#
fit1 <- glm( y ~ glyc.hemo, family=binomial,
             subset=(as.integer(level)==1), data = wisc.crm.data )
summary( fit1, cor=F )
fit2 <- glm( y ~ glyc.hemo, family=binomial,
             subset=(as.integer(level)==2), data = wisc.crm.data )
summary(fit2, cor=F )
fit3 <- glm( y ~ glyc.hemo, family=binomial,
```

```
            subset=(as.integer(level)==3), data = wisc.crm.data )
summary(fit3, cor=F )
#
fit4 <- glm( y ~ level + glyc.hemo, family=binomial,
               data = wisc.crm.data )
summary( fit4, cor=F )
fit5 <- glm( y ~ level * glyc.hemo, family=binomial,
               data = wisc.crm.data )
#
anova( fit4, fit5 )
#
```

Separate fits:

```
table(level)
 :  1 :  2 :  3
  720  445  175

***LEVEL=1
Coefficients:
                  Value Std. Error    t value
(Intercept)  0.22510208 0.38001813  0.5923456
  glyc.hemo -0.05627671 0.02974703 -1.8918429
-------------------------------------------------
```

```
***LEVEL=2
Coefficients:
                  Value Std. Error   t value
(Intercept)  1.01503739 0.48744691  2.082355
  glyc.hemo -0.04551101 0.03731596 -1.219613
-------------------------------------------------

***LEVEL=3
                 Value Std. Error   t value
(Intercept) 0.27807196 0.87839180 0.3165694
  glyc.hemo 0.05636293 0.06755021 0.8343856
-------------------------------------------------
```

CRM fits:

```
                  Value Std. Error     t value
(Intercept)  0.02245317 0.28434528  0.07896444
  level:  2  0.92319834 0.12399549  7.44541891
  level:  3  1.50031056 0.18751941  8.00082817
  glyc.hemo -0.04008980 0.02184728 -1.83500205


(Dispersion Parameter for Binomial family taken to be 1 )

    Null Deviance: 1857.608 on 1339 degrees of freedom

Residual Deviance: 1754.334 on 1336 degrees of freedom

Number of Fisher Scoring Iterations: 3
```

CRM fits:

```
Analysis of Deviance Table

Response: y

               Terms Resid. Df Resid. Dev             Test Df Deviance
1 level + glyc.hemo       1336   1754.334
2 level * glyc.hemo       1334   1751.906 +level:glyc.hemo  2 2.428885
```

Summary

- Multinomial models use a multivariate response vector to represent the outcome.

- For these models the **mean** ($\boldsymbol{\pi}_i$) determines the **covariance**.

- Proportional odds models consider all possible dichotomizations.

- Continuation ratio models consider an ordered sequence of **conditional** expectations.

References :

McCullagh & Nelder, Chapter 5 *Polytomous Regression*

Ananth C.V. and Kleinbaum D.G. (1997), "Regression models for ordinal responses: A review of methods and applications", *Int J of Epi*, 26: 1323-1333.

Fahrmeir & Tutz, Chapter 3 *Multicategorical responses*

Agresti A. (1999), "Modelling ordered categorical data: recent advances and future challenges", *Statistics in Medicine*, 18: 2191-2207.