The partial proportional odds model in the analysis of longitudinal ordinal data

Anne-Francoise DONNEAU

Medical Informatics and Biostatistics
School of Public Health
University of Liège
Promotor: Pr. A. Albert

18 May 2010

Content of the presentation

Notation

Problem: Analysis of ordinal longitudinal data

Notation

Problem: Analysis of ordinal longitudinal data

Units: Subjects, objects, $(i = 1, \cdots, N)$

Notation

Problem: Analysis of ordinal longitudinal data

Units: Subjects, objects, $(i = 1, \cdots, N)$

Outcome: Ordinal variable Y with K levels

Notation

Problem: Analysis of ordinal longitudinal data

Units: Subjects, objects, $(i = 1, \cdots, N)$

Outcome: Ordinal variable Y with K levels

Measurement: Measurements at T occasions, $\mathbf{Y_i} = (Y_{i1}, \cdots, Y_{iT})'$

Notation

Problem: Analysis of ordinal longitudinal data

Units: Subjects, objects, $(i = 1, \cdots, N)$

Outcome: Ordinal variable Y with K levels

Measurement: Measurements at T occasions, $\mathbf{Y_i} = (Y_{i1}, \cdots, Y_{iT})'$

Covariates: $T \times p$ covariates matrix $\mathbf{X_i} = (\mathbf{x_{i1}}, \cdots, \mathbf{x_{iT}})'$
Time, gender, age ...

Notation

Problem: Analysis of ordinal longitudinal data

Units: Subjects, objects, $(i = 1, \cdots, N)$

Outcome: Ordinal variable Y with K levels

Measurement: Measurements at T occasions, $\mathbf{Y_i} = (Y_{i1}, \cdots, Y_{iT})'$

Covariates: $T \times p$ covariates matrix $\mathbf{X_i} = (\mathbf{x_{i1}}, \cdots, \mathbf{x_{iT}})'$
Time, gender, age ...

Domains: Medicine, psychology, social science,...

Motivating example - Quality of life

Dataset

- 247 patients with malignant brain cancer treated by RT+CT or RT
- Assessment of the quality of life at 8 occasions
  Baseline, End RT, End RT + (3,6,9)months, End RT + (1, 1.5, 2)years
- EORTC QLQC30 questionnaire - Appetite loss scale
  Have you lacked appetite? ('Not at all', 'A little', 'Quite a bit', 'Very much')
- Covariates : Time, Treatment (RT+CT vs RT), Tumor cell (pure vs mixed)

Motivating example - Quality of life

Dataset

- 247 patients with malignant brain cancer treated by RT+CT or RT
- Assessment of the quality of life at 8 occasions
  Baseline, End RT, End RT + (3,6,9)months, End RT + (1, 1.5, 2)years
- EORTC QLQC30 questionnaire - Appetite loss scale
  Have you lacked appetite? ('Not at all', 'A little', 'Quite a bit', 'Very much')
- Covariates : Time, Treatment (RT+CT vs RT), Tumor cell (pure vs mixed)

Summary of the data

N=247, T=8, K=4

Motivating example - Quality of life

Dataset

- 247 patients with malignant brain cancer treated by RT+CT or RT
- Assessment of the quality of life at 8 occasions
  Baseline, End RT, End RT + (3,6,9)months, End RT + (1, 1.5, 2)years
- EORTC QLQC30 questionnaire - Appetite loss scale
  Have you lacked appetite? ('Not at all', 'A little', 'Quite a bit', 'Very much')
- Covariates : Time, Treatment (RT+CT vs RT), Tumor cell (pure vs mixed)

Summary of the data

N=247, T=8, K=4

Questions of interest

- Treatment effect
- Tumor cell effect

Proportional odds model

Aim is to find a model that takes into account

- the ordinal nature of the outcome under study
- the correlation between repeated observations
- the unavoidable presence of missing data

Proportional odds model

Aim is to find a model that takes into account

- the ordinal nature of the outcome under study
- the correlation between repeated observations
- the unavoidable presence of missing data

Proportional odds model

$$logit[Pr(Y_{ij} \leq k)] = \theta_k + \mathbf{x'_{ij}}\beta \quad , \quad i = 1, \cdots, N; \quad j = 1, \cdots, T \\ , \quad k = 1, \cdots, K-1$$

Proportional odds model

Aim is to find a model that takes into account

- the ordinal nature of the outcome under study
- the correlation between repeated observations
- the unavoidable presence of missing data

Proportional odds model

$$logit[Pr(Y_{ij} \leq k)] = \theta_k + \mathbf{x'_{ij}}\beta \quad , \quad i = 1, \cdots, N; \quad j = 1, \cdots, T$$
$$, \quad k = 1, \cdots, K-1$$

Properties: invariant when reversing the order of categories
deleting/collapsing some categories

Proportional odds model

Aim is to find a model that takes into account

- the ordinal nature of the outcome under study
- the correlation between repeated observations
- the unavoidable presence of missing data

Proportional odds model

$$logit[Pr(Y_{ij} \leq k)] = \theta_k + \mathbf{x}'_{\mathbf{ij}}\beta \quad , \quad i = 1, \cdots, N; \quad j = 1, \cdots, T$$
$$, \quad k = 1, \cdots, K-1$$

Properties: invariant when reversing the order of categories
deleting/collapsing some categories

Assumption : relationship between Y and X is the same for all categories of Y

Testing the proportional odds model

Tests for assessing proportionality when the outcomes are uncorrelated were extended to longitudinal data (Stiger, 1999).

What if the proportional odds assumption is violated?

- Fitting a more general model
- Dichotomize the ordinal variable and fit separate binary logistic regression models (Bender, 1998).

Our solution

- Fitting a model that allows relaxing the proportional odds assumption when necessary

The partial proportional odds model

The partial proportional odds model (Peterson and Harrel, 1990) allows non-proportional odds for all or a subset q of the p explanatory covariates.

In univariate case,

$$logit[Pr(Y \leq k)] = \theta_k + \mathbf{x'}\boldsymbol{\beta} + \mathbf{z'}\boldsymbol{\gamma_k} \quad , k = 1, \cdots, K-1$$

where $\mathbf{z}$ is a q-dimensional vector ($q \leq p$) of the explanatory variables for which the proportional odds assumption does not hold and $\boldsymbol{\gamma_k}$ is the (q x 1) corresponding vector of coefficients and $\boldsymbol{\gamma_1} = \mathbf{0}$. When $\boldsymbol{\gamma_k} = \mathbf{0}$ for all k, the model reduces to the proportional odds model

Extension of the partial proportional odds model to longitudinal data (Donneau et al., 2010)

In a longitudinal setting,

$$logit[Pr(Y_{ij} \leq k)] = \theta_k + \mathbf{x'_{ij}}\beta + \mathbf{z'_{ij}}\gamma_\mathbf{k} \quad , \quad i = 1, \cdots, N; \quad j = 1, \cdots, T \quad , \quad k = 1, \cdots, K-1$$

where $(\mathbf{z_{i1}}, \cdots, \mathbf{z_{iT}})'$ is a $(T \times q)$ matrix, $q \leq p$, of a subset of q-explanatory variables for which the proportional odds assumption does not apply and $\boldsymbol{\gamma}_\mathbf{k}$ is the (q x 1) corresponding vector of regression parameters with $\boldsymbol{\gamma}_\mathbf{1} = \mathbf{0}$.

As an example (p=2 and q=1), assume that the proportional odds assumption holds for X_1 and not for X_2, then

$$logit[Pr(Y_{ij} \leq k)] = \theta_k + \beta_1 X_1 + (\beta_2 + \gamma_{k,2}) X_2$$

Estimation

Estimation of the regression parameters

- GEE - extension of GLM to longitudinal data (Liang and Zegger, 1986)
- Define of a $(K-1)$ expanded vector of binary responses $\mathbf{Y_{ij}} = (Y_{ij,1}, ..., Y_{ij,(K-1)})'$ where $Y_{ijk} = 1$ if $Y_{ij} \leq k$ and 0 otherwise
- $logit[Pr(Y_{ij} \leq k)] = logit[Pr(Y_{ijk} = 1)] \rightarrow$ member of GLM family

$$\sum_{i=1}^{N} \frac{\partial \boldsymbol{\pi}_\mathbf{i}'}{\partial \beta} \mathbf{W}_\mathbf{i}^{-1} (\mathbf{Y_i} - \boldsymbol{\pi}_\mathbf{i}) = 0$$

where $\mathbf{Y_i} = (\mathbf{Y_{i1}}, ..., \mathbf{Y_{iT}})'$, $\boldsymbol{\pi}_\mathbf{i} = E(\mathbf{Y_i})$ and $\mathbf{W_i = V_i^{1/2} R_i V_i^{1/2}}$ with $\mathbf{V_i}$ the diagonal matrix of the variance of the element of $\mathbf{Y_i}$. The matrix $\mathbf{R_i}$ is the 'working' correlation matrix that expresses the dependence among repeated observations over the subjects.

Missingness

Missing data patterns

- Drop out / attrition
- Non-monotone missingness

Missing data mechanism (Little and Rubin, 1987)

- MCAR: Missing completely at random
- MAR: Missing at random
- MNAR: Missing not at random

Example : Appetite loss - (1) Treatment effect

Model

- Consider the model:

$$logit[Pr(Y_{ij} \leq k)] = \theta_k + (\beta_1 + \boldsymbol{\gamma_{k1}}) t_{ij} + (\beta_2 + \boldsymbol{\gamma_{k2}}) Treat_i + (\beta_3 + \boldsymbol{\gamma_{k3}}) t_{ij} \times Treat_i$$

- $k = 1, 2, 3$
- t_{ij}: j^{th} time of measurement on subject i
- $Treat_i$: treatment group (1= RT+CT vs 0=RT)

Assumption

- Missing data mechanism is MCAR (GEE)
- Proportional odds assumption is verified for t, $Treat$ and $t \times Treat$.
  $\boldsymbol{\gamma_{k,t} = 0} \quad (p = 0.86)$
  $\boldsymbol{\gamma_{k,Treat} = 0} \quad (p = 0.21)$
  $\boldsymbol{\gamma_{k,t \times Treat} = 0} \quad (p = 0.17)$

Example : Appetite loss - (1) Treatment effect

Model becomes

$$logit[Pr(Y_{ij} \leq k)] = \theta_k + \beta_1 t_{ij} + \beta_2 Treat_i + \beta_3 (t_{ij} \times Treat_i) \quad , k = 1, 2, 3$$

Estimation

Table1: GEE parameter estimates for the appetite loss scale - Proportional odds model

Covariates	Estimate	SE	p-value
θ_1	1.21	0.14	
θ_2	2.48	0.16	
θ_3	3.81	0.21	
t_{ij}	0.08	0.04	0.033
$Treat_i$	-0.39	0.19	0.034
$t_{ij} \times Treat_i$	-0.12	0.05	0.009

A significant difference between treatment arms was found in favor of the RT alone treatment.

Example : Appetite loss - (2) Tumor cell effect

Model

- Consider the model:

$$logit[Pr(Y_{ij} \leq k)] = \theta_k + (\beta_1 + \boldsymbol{\gamma_{k1}}) t_{ij} + (\beta_2 + \boldsymbol{\gamma_{k2}}) Tumor_i + (\beta_3 + \boldsymbol{\gamma_{k3}}) t_{ij} \times Tumor_i$$

- $k = 1, 2, 3$
- t_{ij}: j^{th} time of measurement on subject i
- $Tumor_i$: type of diagnosed tumor (1=pure vs 0=mixed)

Assumption

- Missing data mechanism is MCAR (GEE)
- Proportional odds assumption is not met for t, $Tumor$ and $t \times Tumor$.
  $\boldsymbol{\gamma_{k,t}} = \mathbf{0} \quad (p = 0.015)$
  $\boldsymbol{\gamma_{k,Tumor}} = \mathbf{0} \quad (p = 0.044)$
  $\boldsymbol{\gamma_{k,t \times Tumor}} = \mathbf{0} \quad (p = 0.008)$

Example : Appetite loss - (2) Tumor cell effect

Estimations

Table2: GEE parameter estimates for the appetite loss scale - Partial proportional odds model

Covariates	k	Estimate	SE	p-value
θ_1	1	-0.75	0.25	
θ_2	2	1.58	0.41	
θ_3	3	1.93	0.78	
t_{ij}	1	0.49	0.06	$<$0.0001
t_{ij}	2	-0.10	0.12	0.39
t_{ij}	3	0.53	0.22	0.015
$Tumor_j$	1	1.30	0.20	$<$0.0001
$Tumor_j$	2	0.45	0.33	0.18
$Tumor_j$	3	1.14	0.65	0.079
$t_{ij} \times Tumor_j$	1	-0.34	0.04	$<$0.0001
$t_{ij} \times Tumor_j$	2	0.092	0.097	0.34
$t_{ij} \times Tumor_j$	3	-0.32	0.16	0.04

Example : Appetite loss - (2) Tumor cell effect

$$logit[Pr(Y_{ij} \leq 1)] = -0.75 + 0.49 t_{ij} + 1.30\, Tumor_j - 0.34 t_{ij} \times Tumor_j$$
$$logit[Pr(Y_{ij} \leq 2)] = 1.58 - 0.10 t_{ij} + 0.45\, Tumor_j + 0.092 t_{ij} \times Tumor_j$$
$$logit[Pr(Y_{ij} \leq 3)] = 1.93 + 0.53 t_{ij} + 1.14\, Tumor_j - 0.32 t_{ij} \times Tumor_j$$

where 1="Not at all', 2='A little', 3='Quite a bit', 4='Very much'

Interpretation

- At baseline, pure cell tumor patients have $e^{1.30} = 3.7$ time higher odds of having no appetite loss than mixed cells tumor patients.
- At baseline, pure cell tumor patients have $e^{0.45} = 1.6$ time higher odds of having at most little appetite loss than mixed cells tumor patients.
- At baseline, pure cell tumor patients have $e^{1.14} = 3.1$ time higher odds of having at most quite a bite appetite loss than mixed cells tumor patients.

Conclusion

We have explored the extension of the partial proportional odds model to the case of longitudinal data

- Estimation mechanism (GEE)
- Testing for the proportional odds assumption for each covariate
- Final model that
  - takes into account the ordinal nature of the variable under study
  - takes into account the correlation between repeated observations
  - allows relaxing the proportional odds assumption (when necessary)
- Missing data to be first investigated (GEE, WGEE, Mi-GEE)

Thank you for your attention