STAT 545/GS01 1233 193 5528: Generalized Linear Models and Categorical Data Analysis (Part II)
4: Regression model for multinomial data

Ziyi Li

Department of Biostatistics
The University of Texas MD Anderson Cancer Center

ZLi16@mdanderson.org

Fall 2022

Sampling Distributions of Contingency Tables

Seat Belt Example: to record seat-belt use and results of automobile crashes

	Result of crash		
Seat-Belt use	Fatality	Major injury	Minor injury
Always			
Sometimes			
Never			

- Treat the numbers of observations at the nine combinations of seat-belt use and outcome of crash as independent random variables. The sampling distribution is Poisson distribution.
- Randomly sample 400 police records of crashes and classify each according to seat-belt use and outcome of crash. The sampling distribution is Multinomial distribution with total sample n fixed.

Sampling Distributions of Contingency Tables

Seat Belt Example:

	Result of crash		
Seat-Belt use	Fatality	Major injury	Minor injury
Always			
Sometimes			
Never			

- Randomly sample 100 records of accidents with fatality, major injury, and minor injury, respectively. The sampling distribution is independent multinomial distribution for cells within each column with column totals n_{+j} fixed.
- Randomly sample 300 subjects and randomly assign 100 of them to always, sometimes and never wear seat belts; then force them all to have car accident. The sampling distribution is independent multinomial distribution for cells within each row with row totals n_{i+} fixed. For this and many medical studies, this design is not ethical.

Models for Multinomial Responses (§ 8)

- Nominal data vs. ordinal data: presence or absence of intrinsic order
  - Table 8.1: outcome = primary food choice of alligators (fish, invertebrate, reptile, bird, other); covariates = lake, gender, size

Table 8.1 Primary Food Choice of Alligators, by Lake, Gender, and Size of the Alligator

Lake	Gender	Size (m)	Fish	Invertebrate	Reptile	Bird	Other
			Primary Food Choice				
Hancock	Male	≤ 2.3	7	1	0	0	5
		> 2.3	4	0	0	1	2
	Female	≤ 2.3	16	3	2	2	3
		> 2.3	3	0	1	2	3
Oklawaha	Male	≤ 2.3	2	2	0	0	1
		> 2.3	13	7	6	0	0
	Female	≤ 2.3	3	9	1	0	2
		> 2.3	0	1	0	1	0
Trafford	Male	≤ 2.3	3	7	1	0	1
		> 2.3	8	6	6	3	5
	Female	≤ 2.3	2	4	1	1	4
		> 2.3	0	1	0	0	0
George	Male	≤ 2.3	13	10	0	2	2
		> 2.3	9	0	0	1	2
	Female	≤ 2.3	3	9	1	0	1
		> 2.3	8	1	0	0	1

Source: Data courtesy of Clint Moore, from an unpublished manuscript by M. F. Delaney and C. T. Moore.

Models for Multinomial Responses

- Nominal data vs. ordinal data: presence or absence of intrinsic order
  - Table 8.5: outcome = happiness (pretty, very, not too); covariates = total number of traumatic events experienced recently, race, age

Table 8.5 Four Observations from Data Set on Happiness, Number of Traumatic Events, and Race

Observation	Happiness	Number of Traumatic Events	Race
1	Pretty happy	2	White
2	Pretty happy	3	Black
3	Very happy	0	White
4	Not too happy	5	White

Source: 1984 General Social Survey; complete data at www.stat.ufl.edu/~aa/cda/cda.html.

Models for Multinomial Responses

- Nominal data vs. ordinal data: presence or absence of intrinsic order
- In this part, we start with the multinomial sampling scheme, and introduce two types of models for multinomial responses:
  - Nominal Responses: Baseline-Category logit models
  - Ordinal Responses: Cumulative logit models

Models for Multinomial Responses

- For nominal outcome variable $Y = 1, 2, ..., J$. We need to model $\pi_j(\boldsymbol{x}) = P(Y = j|\boldsymbol{x})$ under constraint $\sum_j \pi_j(\boldsymbol{x}) = 1$.
- Y follows a multinomial distribution with probabilities $\{\pi_1(\boldsymbol{x}), ..., \pi_J(\boldsymbol{x})\}$.
- Baseline-category logit model (e.g., pick J as the reference/baseline category): the logit of conditional prob. $P(Y = j | Y = j \text{ or } Y = J)$

$$\log \frac{\pi_j(\boldsymbol{x})}{\pi_J(\boldsymbol{x})} = \alpha_j + \boldsymbol{\beta}_j^T \boldsymbol{x} \quad , \qquad j = 1, 2, ..., J-1$$

- Baseline logit models pair each response category with a baseline category (in both SAS proc logistic and the R function vglm it is the last one J).
- These $J - 1$ equations imply logit-linear model for other pairs of response categories, e.g., the log odds of category **a** vs. **b**

$$\log \frac{\pi_a(\boldsymbol{x})}{\pi_b(\boldsymbol{x})} = \log \frac{\pi_a(\boldsymbol{x})}{\pi_J(\boldsymbol{x})} - \log \frac{\pi_b(\boldsymbol{x})}{\pi_J(\boldsymbol{x})}$$

Models for Multinomial Responses: baseline-category logit model

- The constraint leads to: $\pi_J(x) = \dfrac{1}{1 + \sum_{j=1}^{J-1} \exp\left(\alpha_j + \boldsymbol{\beta}_j^T \boldsymbol{x}\right)}$
- The probability for the j-th category ($j = 1, 2, ..., J-1$):

$$\pi_j(x) = \frac{\exp\left(\alpha_j + \boldsymbol{\beta}_j^T \boldsymbol{x}\right)}{1 + \sum_{j=1}^{J-1} \exp\left(\alpha_j + \boldsymbol{\beta}_j^T \boldsymbol{x}\right)}$$

- With more than two response categories ($J > 2$), the probability of a given category need not continuously increase or decrease (e.g., $\pi_j(x)$ may not be a monotone function of x)
- The model is fit by maximum likelihood. Let $\boldsymbol{y}_i = (y_{i1}, y_{i2}, ..., y_{iJ})$, where $y_{ij} = 1$ when the response is in category j and 0 otherwise, so that $\sum_j y_{ij} = 1$. The log likelihood is:

$$\sum_{i=1}^{n} \log \left[\prod_{j=1}^{J} \pi_j(\boldsymbol{x}_i)^{y_{ij}} \right]$$

Estimation of baseline-category logit model

- The log-likelihood function is concave and the Newton-Raphson method yields the MLE
- The MLEs have large-sample normal distributions.
- The standard errors are square roots of diagonal elements of the inverse information matrix
- Quasi-complete separation can happen; need large sample size; avoid too many categories
- In SAS, use Proc Logistic with "link=glogit" or Proc Catmod; in R, use vglm function in VGAM package.
- SAS code for Table 8.4 on textbook website.

Analysis of Alligator Food Data

- Let f denote food type $f \in \{F(1), I(2), B(3), R(4), O(5)\}$. F for fish (reference group), B for Bird, I for Invertebrate, R for Reptile, and O for other.
- Let l denote lake $l \in \{H(1), O(2), T(3), G(4)\}$. H for Hancock, O for Oklawaha, T for Trafford, and G for George (reference group).
- The equation for the log odds of food type $f \in \{I, B, R, O\}$ relative to F (fish) for size s in lake l is

$$log \frac{\pi_{fsl}}{\pi_{Fsl}} = \alpha_f + \beta^S_{fs} + \beta^L_{fl}$$

where $\alpha_F = 0$ and $\beta^S_{f0} = \beta^L_{fG} = 0$ for $f = F, I, B, R, O$.

Analysis of Alligator Food Data

Table 8.4 Estimated Parameters in Baseline-Category Logit Model for Alligator Food Choice, Based on Indicator Variable for Size (1 = Small, 0 = Large) and for Each Lake except Lake George[a]

			Lake		
Logit[b]	Intercept	Size ≤ 2.3	Hancock	Oklawaha	Trafford
$\log(\pi_I/\pi_F)$	−1.55	1.46 (0.40)	−1.66 (0.61)	0.94 (0.47)	1.12 (0.49)
$\log(\pi_R/\pi_F)$	−3.31	−0.35 (0.58)	1.24 (1.19)	2.46 (1.12)	2.94 (1.12)
$\log(\pi_B/\pi_F)$	−2.09	−0.63 (0.64)	0.70 (0.78)	−0.65 (1.20)	1.09 (0.84)
$\log(\pi_O/\pi_F)$	−1.90	0.33 (0.45)	0.83 (0.56)	0.01 (0.78)	1.52 (0.62)

[a] *SE* values in parentheses.
[b] Response categories: *I*, invertebrate; *R*, reptile; *B*, bird; *O*, other; *F*, fish.

- How to interpret?

Analysis of Alligator Food Data

Table 8.4 Estimated Parameters in Baseline-Category Logit Model for Alligator Food Choice, Based on Indicator Variable for Size (1 = Small, 0 = Large) and for Each Lake except Lake George[a]

			Lake		
Logit[b]	Intercept	Size ≤ 2.3	Hancock	Oklawaha	Trafford
$\log(\pi_I/\pi_F)$	−1.55	1.46 (0.40)	−1.66 (0.61)	0.94 (0.47)	1.12 (0.49)
$\log(\pi_R/\pi_F)$	−3.31	−0.35 (0.58)	1.24 (1.19)	2.46 (1.12)	2.94 (1.12)
$\log(\pi_B/\pi_F)$	−2.09	−0.63 (0.64)	0.70 (0.78)	−0.65 (1.20)	1.09 (0.84)
$\log(\pi_O/\pi_F)$	−1.90	0.33 (0.45)	0.83 (0.56)	0.01 (0.78)	1.52 (0.62)

[a] *SE* values in parentheses.
[b] Response categories: *I*, invertebrate; *R*, reptile; *B*, bird; *O*, other; *F*, fish.

The interpretation for the first row:

- For any given lake, for small alligators the estimated odds that the primary food choice was invertebrates instead of fish are $exp(1.46) = 4.3$ times the odds for large alligators. Wald 95% CI (2.0, 9.3).
- For a given size, there are lake effects on the log odds of invertebrates as the primary food source instead of fish. The estimated odds are relatively higher at Trafford and Oklawaha and relatively lower at Hancock than they are at Lake George.

Analysis of Alligator Food Data

The log odds relative to the reference level also determine the log odds for other food choice pairs:

- For example, the log odds that invertebrates (I) are the primary food choice relative to birds (B) for lake l and size s is

$$\begin{aligned}\log(\pi_{Isl}/\pi_{Bsl}) &= \log(\pi_{Isl}/\pi_{Fsl}) - \log(\pi_{Bsl}/\pi_{Fsl}) \\ &= \alpha_I + \beta^S_{Is} + \beta^L_{Il} - (\alpha_B + \beta^S_{Bs} + \beta^L_{Bl})\end{aligned}$$

- For small alligator ($s = 1$) at lake Trafford (l = T), this is estimated to be

$$\log(\widehat{\pi}_{I1T}/\widehat{\pi}_{B1T}) = -1.55 + 1.46 + 1.12 - (-2.09) - (-0.63) - 1.09 =$$

- For large alligators ($s = 0$) at lake Trafford ($l = T$) this is estimated to be

$$\log(\widehat{\pi}_{I0T}/\widehat{\pi}_{B0T}) = -1.55 + 0 + 1.12 - (-2.09) - 0 - 1.09 = 0.58$$

Analysis of Alligator Food Data

- Subtracting these two estimated log odds gives us an estimate for the log odds ratio that the primary food choice is invertebrates rather than birds for small versus large alligators at lake Trafford:

$$\log\left(\frac{\widehat{\pi}_{I1T}/\widehat{\pi}_{B1T}}{\widehat{\pi}_{I0T}/\widehat{\pi}_{B0T}}\right) = 2.67 - 0.58 = 2.09$$

- Interpretation: The estimated odds of the primary food choice being invertebrates as opposed to birds for small alligators at lake Trafford are $\exp(2.09) = 8.08$ times the estimated odds of the primary food choice being invertebrates as opposed to birds for large alligators at the same lake.

Analysis of Alligator Food Data

Estimating Response Probabilities:

- Response probabilities are given by

$$\pi_j(x) = \frac{\exp(\alpha_j + \beta_j^T x)}{1 + \Sigma_{h=1}^{J-1} \exp(\alpha_h + \beta_h^T x)}, j = 1, 2, \cdots, J-1,$$

  where $\alpha_J = 0$ and $\beta_J = 0$

- For example, the estimated probability that a small alligator ($s = 1$) in Lake Hancock ($l = H$) has invertebrates ($f = I$) as the primary food choice is

$$\widehat{\pi}_{I1H}$$

$$= \frac{e^{-1.55+1.46-1.66}}{1 + e^{-1.55+1.46-1.66} + e^{-3.31-0.35+1.24} + e^{-2.09-0.63+0.7} + e^{-1.\text{!`}}}$$

$$= 0.093$$

Analysis of Alligator Food Data

- Presenting the effect of continuous covariates graphically
- Relationship with probability may not be monotone (different from binary outcome)
- Use 3 instead of 5 response categories

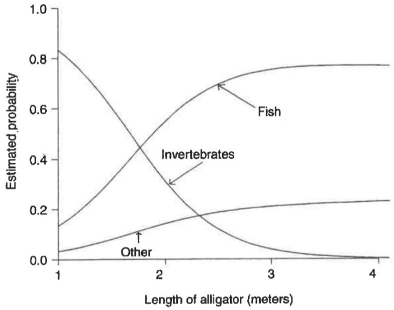


Figure 8.1 Estimated probabilities for primary food choice.

Models for Ordinal Responses

- In the presence of an ordinal response, utilizing its ordinal nature results in more parsimonious models and more powerful inference.
- $Y = 1, 2, ..., J$. Model $\pi_j(\boldsymbol{x}) = P(Y = j|\boldsymbol{x})$ under constraint $\sum_j \pi_j(\boldsymbol{x}) = 1$.
- Due to the intrinsic ordering of the response categories, we model the cumulative probabilities:

$$P(Y \leq j|\boldsymbol{x}) = \pi_1(\boldsymbol{x}) + ... + \pi_j(\boldsymbol{x}), j = 1, \cdots, J.$$

- The cumulative logits are defined, for $j = 1, \cdots, J-1$, as

$$\begin{aligned} logit[P(Y \leq j|\boldsymbol{x})] &= \log \frac{P(Y \leq j|\boldsymbol{x})}{1 - P(Y \leq j|\boldsymbol{x})} \\ &= \log \frac{\pi_1(\boldsymbol{x}) + ... + \pi_j(\boldsymbol{x})}{\pi_{j+1}(\boldsymbol{x}) + ... + \pi_J(\boldsymbol{x})} \end{aligned}$$

Model for Ordinal Responses

- Cumulative logit model

$$logit[P(Y \leq j|\boldsymbol{x})] = \alpha_j + \beta^T \boldsymbol{x} \quad , \quad j = 1, 2, ..., J - 1$$

- The cumulative logit is monotone in $\boldsymbol{x}$ because we require $\alpha_1 \leqslant \alpha_2 \leqslant ... \leqslant \alpha_{J-1}$; this feature not available in baseline-category logit model (simpler, and more interpretable)
- A model for $logit[\Pr(Y \leq j)]$ for a single j is an ordinary logit model for a binary response, $Y \leq j$ vs $Y > j$.
- One approach is to fit each ordinary logit model separately, resulting in parameter estimates may not be “compatible” with each other.
- Alternatively, we can include all $J - 1$ cumulative logits in one single model, the proportional odds model, and estimate all parameters simultaneously.

Proportional Odds Model

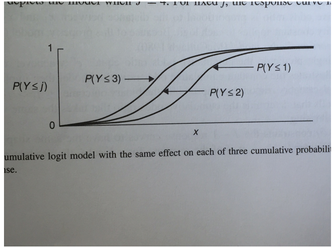
- For a fixed j and at two different levels of the predictors, x_1 and x_2, we have

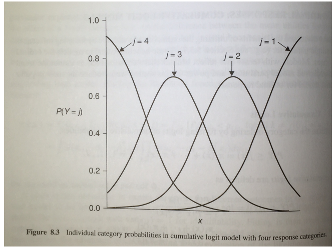
$$
\begin{aligned}
& logit[P(Y \leq j|\boldsymbol{x}_2)] - logit[P(Y \leq j|\boldsymbol{x}_1)] \\
= & \log \frac{\Pr(Y \leq j|\boldsymbol{x}_2)/\Pr(Y > j|\boldsymbol{x}_2)}{\Pr(Y \leq j|\boldsymbol{x}_1)/\Pr(Y > j|\boldsymbol{x}_1)} \\
= & \beta^T(\boldsymbol{x}_2 - \boldsymbol{x}_1)
\end{aligned}
$$

  This is an odds ratio of cumulative probabilities, also called a cumulative odds ratio.

- When we use the same β for each logit, *cumulative logit model* is also called the *proportional odds model* because *cumulative odds ratio* $\exp(\beta)$ does not depend on j.
- Interpretations: The odds of having response $\leq j$ at $\boldsymbol{x} = \boldsymbol{x}_2$ are $exp[\beta^T(\boldsymbol{x}_2 - \boldsymbol{x}_1)]$ times the odds of response at $\boldsymbol{x} = \boldsymbol{x}_2$.

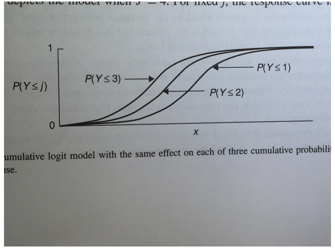
Proportional Odds Model

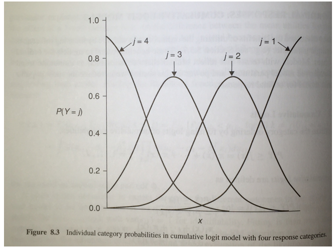




- The name proportional odds model comes from the fact that the log cumulative odds ratio is proportional to the distance between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.
- We cannot make the model more generalizable by replacing $\boldsymbol{\beta}$ with $\boldsymbol{\beta}_j$: the different cumulative probabilities may cross ($\alpha_j + \boldsymbol{\beta}_j^T \boldsymbol{x}$ is not monotone) implying negative probability of individual categories.

Proportional Odds Model





- The cumulative probability curves in Figure 8.2 have the same shape, just with a shift left or right. The individual category probability is more difficult to interpret (but with ordinal data, cumulative probability is interpretable).

$$P(Y \leq j|x) = expit(\alpha_j + \beta x)$$

$$P(Y \leq j+1|x) = expit(\alpha_{j+1} + \beta x) = expit\left[\alpha_j + \beta\left(x + \frac{\alpha_{j+1} - \alpha}{\beta}\right.\right.$$

Proportional Odds Model

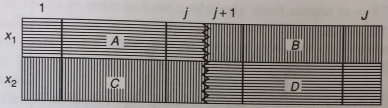


Figure 8.4 Uniform odds ratios AD/BC whenever $x_1 - x_2 = 1$, for all binary collapsings of the response in cumulative logit model of proportional odds form.

likelihood function is

$$\prod_{i=1}^{n}\left[\prod_{j=1}^{J}\pi_j(\mathbf{x}_i)^{y_{ij}}\right] = \prod_{i=1}^{n}\left\{\prod_{j=1}^{J}[P(Y \le j|\mathbf{x}_i) - P(Y \le j-1|\mathbf{x}_i)]^{y_{ij}}\right\}$$

$$= \prod_{i=1}^{n}\left\{\prod_{j=1}^{J}\left[\frac{\exp(\alpha_j + \boldsymbol{\beta}^T\mathbf{x}_i)}{1+\exp(\alpha_j + \boldsymbol{\beta}^T\mathbf{x}_i)} - \frac{\exp(\alpha_{j-1} + \boldsymbol{\beta}^T\mathbf{x}_i)}{1+\exp(\alpha_{j-1} + \boldsymbol{\beta}^T\mathbf{x}_i)}\right]^{y_{ij}}\right\}, \tag{8.6}$$

viewed as a function of $(\{\alpha_j\}, \boldsymbol{\beta})$. This can be maximized to obtain the ML estimates us-

- Maximum likelihood estimation, Newton-Raphson, non-canonical link model

Probit and Logit: Latent Variable Motivation for Binary Outcome Model

- A continuous latent variable y^* with $y^* = \tilde{\boldsymbol{\beta}}^T \boldsymbol{x} + \epsilon$ (no intercept) and the distribution function of ϵ is $G(.)$
- Threshold $-\infty < \tilde{\alpha} < \infty$. The observed response y satisfies $y = 1$ if $y^* \leq \tilde{\alpha}$ and $y = 0$ otherwise

$$P(Y = 1|\boldsymbol{x}) = P(y^* \leq \tilde{\alpha}|\boldsymbol{x}) = G(\tilde{\alpha} - \tilde{\boldsymbol{\beta}}^T \boldsymbol{x})$$

- The logistic regression model:
  $P(Y = 1|\boldsymbol{x}) = expit\left(\alpha + \boldsymbol{\beta}^T \boldsymbol{x}\right)$.
  $$G(.) \sim expit(.),\ G^{-1}(.) \sim logit(.),\ \tilde{\alpha} = \alpha,\ \tilde{\boldsymbol{\beta}} = -\boldsymbol{\beta}$$
- The probit regression model: $P(Y = 1|\boldsymbol{x}) = \Phi\left(\alpha + \boldsymbol{\beta}^T \boldsymbol{x}\right)$.
  $$G(.) \sim \Phi(.),\ G^{-1}(.) \sim \Phi^{-1}(.),\ \tilde{\alpha} = \alpha,\ \tilde{\boldsymbol{\beta}} = -\boldsymbol{\beta}$$

Latent Variable Motivation for Proportional Odds Model

- A continuous latent variable y^* with $y^* = \tilde{\boldsymbol{\beta}}^T \boldsymbol{x} + \epsilon$ and the distribution function of ϵ is $G(.)$ (not mean zero)
- Thresholds $-\infty = \tilde{\alpha}_0 < \tilde{\alpha}_1 < ... < \tilde{\alpha}_J = \infty$
- The observed response y satisfies $y = j$ if $\tilde{\alpha}_{j-1} < y^* \leq \tilde{\alpha}_j$

$$P(Y \leq j|\boldsymbol{x}) = P(y^* \leq \tilde{\alpha}_j|\boldsymbol{x}) = G(\tilde{\alpha}_j - \tilde{\boldsymbol{\beta}}^T \boldsymbol{x})$$

- The proportional odds model

$$P(Y \leq j|\boldsymbol{x}) = expit\left(\alpha_j + \boldsymbol{\beta}^T \boldsymbol{x}\right)$$

- $G(.) \sim expit(.)$, $G^{-1}(.) \sim logit(.)$, $\tilde{\alpha}_j = \alpha_j$, $\tilde{\boldsymbol{\beta}} = -\boldsymbol{\beta}$
- Logistic regression is a special case

Proportional Odds Model: Interpretation using latent variable

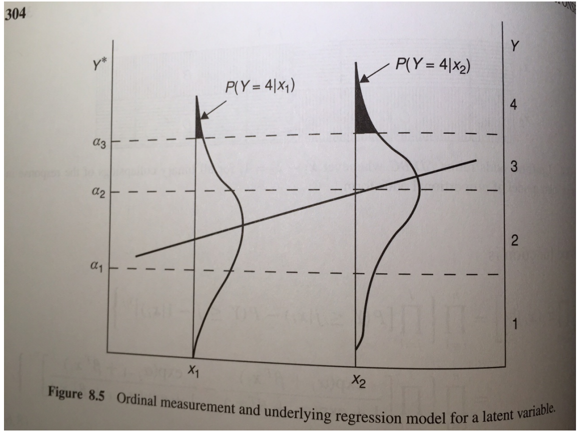


Figure 8.5 Ordinal measurement and underlying regression model for a latent variable.

Check the Proportional Odds Assumption

$$P(Y \leq j|\boldsymbol{x}) = expit\left(\alpha_j + \boldsymbol{\beta}^T\boldsymbol{x}\right)$$

- Proportional odds model is parsimonious and easy to interpret due to the proportional odds assumption $\beta_j \equiv \beta$ for all j.
- Replacing β with β_j may cause the cumulative probabilities to cross (will not cross within the observed range of $\boldsymbol{x}$ if the deviation is small). Check through separate logistic regression for binary collapsings of the outcome variable
- A score test of proportional odds model is available (SAS PROC LOGISTIC)
- Retain proportional odds model unless there is strong deviation from this assumption
- What to do when the proportional odds assumption is violated:
  - Adding additional terms, such as interaction between covariates
  - alternative ordinal model (next slides)

Alternative Models for Ordinal Data

- Cumulative link model: $G^{-1}\left[P(Y \leq j|\boldsymbol{x})\right] = \alpha_j + \boldsymbol{\beta}^T\boldsymbol{x}$
  - When the link is logit, we have odds ratio interpretation
- Cumulative probit model: $\Phi^{-1}\left[P(Y \leq j|\boldsymbol{x})\right] = \alpha_j + \boldsymbol{\beta}^T\boldsymbol{x}$
  - We can interpret the model with underlying normal distribution
- Cumulative complementary log-log model: $\log\left\{-\log\left[1 - P(Y \leq j|\boldsymbol{x})\right]\right\} = \alpha_j + \boldsymbol{\beta}^T\boldsymbol{x}$
  - Equivalent to a latent variable model with extreme value distribution for the residuals
  - Equivalent to Cox proportional hazard model in for discrete survival data analysis (e.g., year of death at 1, 2, 3, ...); need survival analysis background

Alternative Models for Ordinal Data (cont.)

- Adjacent category logit model
  - Equivalent to baseline-category logit model but make use of the ordering
  - Focus on change in the probability of individual outcome categories. Baseline-category logit model focuses on cumulative probability
  - Parsimonious interpretation with a single β; unlike baseline-category logit model, using β_j instead of a common β does not lead to conflict
- Continuation ratio logit model
  - Connection with discrete survival data model (modeling the hazard function)
  - Separate estimation at each level j