Ordinal Responses

- Ordinal responses are common in areas such as marketing research, opinion polls and psychiatry where “soft” measures are common.
- If there is an obvious order among the response categories, this should be taken into account.
- Conclusions should not be much affected by the number or choice of categories.

- It's natural to look at cumulative probabilities

$$
\begin{aligned}
\pi_j &= \mathbb{P}(y_j = 1), \\
\gamma_j &= \mathbb{P}(\sum_{k=1}^{j} y_k = 1) = \sum_{k=1}^{j} \pi_k .
\end{aligned}
$$

- γ_j is monotonically increasing with j
- Different models are available: cumulative model, adjacent category model, continuation ratio model.

Cumulative Model

Model cumulative probabilities γ_j $(j = 1, \cdots, J-1)$ where

$$g(\gamma_j) = \eta_j = \boldsymbol{x}^T \boldsymbol{\beta}_j = \beta_{0j} + \beta_{1j} x_1 + \cdots + \beta_{pj} x_p$$

- Several options for the link function
- We will focus on the logit link
- Logit: $g(\gamma_j) = \log \frac{\gamma_j}{1-\gamma_j} = \log \frac{\pi_1 + \cdots + \pi_j}{\pi_{j+1} + \cdots + \pi_J}$

A special case: proportional odds model

$$\log \frac{\gamma_j}{1-\gamma_j} = \log \frac{\pi_1 + \cdots + \pi_j}{\pi_{j+1} + \cdots + \pi_J} = \beta_{0j} + \beta_1 x_1 + \cdots + \beta_p x_p$$

where $-\infty < \beta_{01} < \cdots < \beta_{0,J-1} < \infty$.

- γ_j's are guaranteed to be monotonically increasing.
- $J - 1 + p$ parameters in the model.
- Easy interpretation: β_k is the log odds ratio (of lower categories vs higher categories) for a unit change in x_k.

Example:

When there is only one covariate variable (e.g., indicator of treatment, $X = 0, 1$), then

$$\exp(\beta) = \frac{\text{odds}(Y \le j | x = 1)}{\text{odds}(Y \le j | x = 0)}$$

- If $\beta = 0$, $\mathbb{P}(Y \le j | x = 0) = \mathbb{P}(Y \le j | x = 1)$
- If $\beta > 0$, treatment group has larger odds of $Y \le j$
- If $\beta < 0$, control group has larger odds of $Y \le j$

$$\log \frac{\gamma_j}{1-\gamma_j} = \beta_{0j} + \boldsymbol{x}^T \boldsymbol{\beta}, \quad j = 1, \cdots, J-1$$

- Intuitively, this model assumes that the effects of the covariates are the same for all categories on the log odds scale.
- If some of the consecutive categories are combined, this does not change the true parameters $\beta_1, \ldots, \beta_{p-1}$ except the intercept.
- If the labelling of the categories is reversed, only the signs of the parameters will be changed.
- This is usually the default model in common statistical softwares like SAS, R/S-PLUS.

Alternative Models

- The adjacent-category logit model is

$$\log(\frac{\pi_j}{\pi_{j+1}}) = \boldsymbol{x}^T \boldsymbol{\beta}_j.$$

  which models the odds for each pair of adjacent response categories.

- The continuation-ratio logit model is

$$\log(\frac{\pi_j}{\pi_{j+1} + \cdots + \pi_J}) = \boldsymbol{x}^T \boldsymbol{\beta}_j,$$

  which models the odds of the response being in category j conditional on $\geq j$. This model is useful for hierarchical responses.

- Can be fitted using R pacakge VGAM

Example: Car Preference

		Response			
Sex	Age	Unimportant	Import	Very Import	Total
Women	18-23	26 (58%)	12 (27%)	7 (16%)	45
	24-40	9 (20%)	21 (47%)	15 (33%)	45
	> 40	5 (8%)	14 (23%)	41 (68%)	60
Men	18-23	40 (62%)	17 (26%)	8 (12%)	65
	24-40	17 (39%)	15 (34%)	12 (27%)	44
	> 40	8 (20%)	15 (37%)	18 (44%)	41
Total		105	94	101	300

The following proportional odds model was fitted to the data:

$$\log\left(\frac{\pi_1}{\pi_2 + \pi_3}\right) = \beta_{01} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3,$$

$$\log\left(\frac{\pi_1 + \pi_2}{\pi_3}\right) = \beta_{02} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3.$$