Chapter 4 Models for Polytomous Data

4.1 Introduction:

Motivating example:

Cheese	I	II	III	IV	V	VI	VII	VIII	IX	Total
A	0	0	1	7	8	8	19	8	1	52
B	6	9	12	11	7	6	1	0	0	52
C	1	1	6	8	23	7	5	1	0	52
D	0	0	0	1	3	7	14	16	11	52
Total	7	10	19	27	41	28	39	25	12	208

Response category: I~IX ("strong dislike" to "excellent taste").

$y_{ij}, i = 1, 2, 3, 4; j = 1, 2, \cdots, 9$: the response frequencies for the cheese additives.

$i = 1$: additive A, $i = 2$: additive B, $i = 3$: additive C, $i = 4$: additive D.

$j = 1$: response I, $j = 2$: response II, $j = 3$: response III,

$j = 4$: response IV, $j = 5$: response V, $j = 6$: response VI,

$j = 7$: response VII, $j = 8$: response VIII, $j = 9$: response IX.

For example,

$$y_{11} = 0, y_{12} = 0, y_{13} = 1, y_{14} = 7, y_{15} = 8, y_{16} = 8, y_{17} = 19, y_{18} = 8, y_{19} = 1.$$

Also,

$\pi_{ij}, i = 1, 2, 3, 4; j = 1, 2, \cdots, 9$: the probability corresponding to the cheese additive and the response category.

$r_{ij} = \sum_{r \le j} \pi_{ij}$: the cumulative probability corresponding to the cheese additive and the response category.

That is,

$$r_{i1} = \pi_{i1}, r_{i2} = \pi_{i1} + \pi_{i2}, \cdots, r_{i9} = \pi_{i1} + \pi_{i2} + \cdots + \pi_{i9} = 1$$

Then, the distribution of

$$Y_i = (Y_{i1}, Y_{i2}, \cdots, Y_{i9})$$

of which data values are

$$y_i = (y_{i1}, y_{i2}, \cdots, y_{i9})$$

Is a multinomial distribution $M(m_i, \pi_{i1}, \cdots, \pi_{ik})$ with parameters $m_i = 52$ and $\pi_i = (\pi_{i1}, \pi_{i2}, \cdots, \pi_{i9})$.

Objective:

We are concerned with the effect on the taste of various cheese additives. That is, we want to evaluate the statistical significance of the differences among these

cheese additives. We want to find a model which is capable of describing these differences, for example, the interrelation among different cumulative probabilities such as

$$r_{4j} < r_{1j} < r_{3j} < r_{2j} \ \ (D, A, C, B \ \ from \ \ the \ \ best \ \ to \ \ the \ \ worst).$$

Definition of Polytomous Data:

If the response of an individual or item in a study is restricted to one of a fixed set of possible values, we say that the response is polytomous.

Examples of polytomous data include blood type $(\mathrm{A}, \mathrm{B}, \mathrm{AB}, \mathrm{O}, \ldots)$, food testing, measures of mental and physical well-being, variables arising in social science research.

Note:

If the categories are ordered and $j = 1, 2, \cdots, k$, we may prefer to work with the cumulative response probabilities

$$r_{i1} = \pi_{i1}, r_{i2} = \pi_{i1} + \pi_{i2}, \cdots, r_{ik} = \pi_{i1} + \pi_{i2} + \cdots + \pi_{ik} = 1$$

It makes little sense to work with a model specified in term of r_{ij} if the response categories are not ordered.