4.2 Measurement scales and modeling:

1. General

There are two types of scales, pure scales and compound scales. A bivariate responses with one response ordinal and the other continuous is an example of compound scales. For pure scales, there are several types:

(a) nominal scales: the categories are regarded as exchangeable and totally devoid of structure.

(b) ordinal scales: the categories are ordered much like the ordinal number, "first", "second",…. It does not make sense to talk of "distance" or "spacing" between "first" and "second" nor to compare "spacings" between pairs of response categories.

(c) interval scales: the categories are ordered and numerical labels or scores are attached. The scores are treated as category averages, median or mid-points. Differences between scores are therefore interpreted as a measure of separation of the categories.

Note:

In applications, the distinction between nominal or ordinal scales is usually but not always clear. For example, hair color and eye color can be ordered to a large extent on the grey-scale from light to dark and are therefore ordinal. However, unless there is a clear connection with electromagnetic spectrum or a grey-scale, colors are best regarded as nominal.

2. Models for ordinal scales

Ordinal scales occur more frequently in applications than the other types. The applications include food testing (bad, good, excellent,…), classification of radiographs, determination of physical or mental well-being, ….

Note:

It is essential the same conclusion can be arrived even though the number or choice of response categories has been changed. As a consequence, if a new category is formed by combining adjacent categories of the old scale, the form of the conclusions should be unaffected. This is an important non-mathematical point that is difficult to make mathematically rigorous. This point lead fair directly to models based on the cumulative probabilities r_{ij} rather than the category probabilities π_{ij}.

Commonly used models:

There are two commonly used models that are found to work well in practice. They are:

(a) logistic scale:

It is the simplest model. The form is

$$\log\left[\frac{r_{ij}(x)}{1-r_{ij}(x)}\right] = \theta_j - x\beta,$$

where the covariate x depends on the ith subject. This model is also known as the proportional-odds model since the ratio of the odds is

$$\frac{\left[\frac{r_{ij}(x_1)}{1-r_{ij}(x_1)}\right]}{\left[\frac{r_{ij}(x_2)}{1-r_{ij}(x_2)}\right]} = exp[-(x_1 - x_2)\beta]$$

which is independent of the choice of category j.

In addition, if

$$X = \begin{cases} 1, treatment\ \ group \\ 0, control\ \ group, \end{cases}$$

Then

$$\frac{\left[\frac{r_{ij}(1)}{1-r_{ij}(1)}\right]}{\left[\frac{r_{ij}(0)}{1-r_{ij}(0)}\right]} = exp(-\beta).$$

(b) complementary log-log scale:

The form is

$$log\{-log[1 - r_{ij}(x)]\} = \theta_j - x\beta.$$

Note:

The model based on logistic scale may be derived from the notion of a tolerance distribution or an underlying unobserved continuous random variable Z,

$$Z = x\beta + \varepsilon,$$

ε is distributed as the logistic distribution. If the unobserved variable lies in the interval $\theta_{j-1} < Z \le \theta_j$, then $Y_{ij} = y_{ij}$ is recorded. That is,

$$\begin{aligned} r_{ij}(x) &= P(Y_{i1} = y_{i1}\ \ or\ \ \cdots or\ \ Y_{ij} = y_{ij}) \\ &= P(Z \le \theta_j) \\ &= P(Z - x\beta \le \theta_j - x\beta) \\ &= P(\varepsilon \le \theta_j - x\beta) \end{aligned}$$

$$= \frac{exp(\theta_j - x\beta)}{1 + exp(\theta_j - x\beta)}$$

$$\Leftrightarrow \quad \log\left[\frac{r_{ij}(x)}{1 - r_{ij}(x)}\right] = \theta_j - x\beta \quad .$$

Note:

It is sometimes claimed that the models based on logistic scale and complementary log-log scale and related models are appropriate only if there exists a latent variable Z. This claim seems to be too strong and, in any case, the existence of Z is usually unverifiable in practice.

Note:

The model,

$$\frac{Z - x\beta}{exp(x\tau)} = \varepsilon$$

is worthy of serious consideration, where τ is a parameter and ε is distributed as the logistic distribution. The model will lead to

$$\log\left[\frac{r_{ij}(x)}{1 - r_{ij}(x)}\right] = \frac{\theta_j - x\beta}{exp(x\tau)},$$

where $x\beta$ plays the role of linear predictor for the mean and in the denominator $x\tau$ plays the role of linear predictor for the dispersion or variance. If

$$X = \begin{cases} 1, treatment \;\; group \\ 0, control \;\; group, \end{cases}$$

Then

$$\frac{\left[\frac{r_{ij}(1)}{1 - r_{ij}(1)}\right]}{\left[\frac{r_{ij}(0)}{1 - r_{ij}(0)}\right]} = exp\left(\frac{\theta_j - \beta}{\sigma} - \theta_j\right) = exp\left(-\frac{\beta}{\sigma}\right) exp\left[\theta_j\left(\frac{1}{\sigma} - 1\right)\right]$$

,

where $\sigma = exp(\tau)$. If $\sigma < 1$, then the odds ratio is increasing in θ_j and decreasing otherwise. This model is useful for testing the proportional-odds assumption $(\tau = 0)$ against the alternative that the odds ratio is systematically increasing or systematically decreasing in θ_j.

Note:

Models in which the $k - 1$ regression lines are not parallel can be specified by

$$\log\left[\frac{r_{ij}(x)}{1 - r_{ij}(x)}\right] = \theta_j - x\beta_j.$$