Item Response Probit Model

Summer 2024

This report writes about the DIHOPIT model with the random effect deleted. It is also called the item response probit model in psychometric. I plan to write down the current information and try to integrate the psychometric and linear regression terms. This approach may help us further investigate how we can use this model to solve the urban-rural bias that exists in the current PCA approach.

Mathematical Formula

The DIHOPIT model without random effect can be written as

$$\begin{aligned} y_i^* &= X_i'\beta + \varepsilon_i \quad i = 1, \ldots, N \\ \varepsilon_i &\sim N(0,1) \end{aligned}$$

The observation mechanism for each indicator variable $a = 1, \ldots, A$:

$$\begin{aligned} y_i^a &= 0 \text{ if } \quad -\infty < y_i^* \leq \tau^a \\ y_i^a &= 1 \text{ if } \quad \tau^a < y_i^* \leq +\infty \end{aligned}$$

where τ^a is the threshold for the a^{th} indicator variable.

$$\begin{aligned} P(y_i^a = 1|X_i) &= P(X_i'\beta + \varepsilon_i > \tau^a) \\ &= P(\varepsilon_i > \tau^a - X_i'\beta) \\ &= 1 - P(\varepsilon_i \leq \tau^a - X_i'\beta) \\ &= 1 - \Phi(\tau^a - X_i'\beta) \\ &= \Phi(X_i'\beta - \tau^a) \end{aligned}$$

The same model has another form of expression in terms of

$$\begin{aligned} &\theta_i \sim N(0,1) \quad i = 1, \ldots, N \\ &P(y_i^a = 1|\theta_i, a_a, b_a) = \Phi(a_a(\theta_i - b_a)) \quad a = 1, \ldots, A \end{aligned}$$

where

θ_i is the latent trait (ability) of respondent i

y_i^a is the response of individual i to item a

$\Phi(\cdot)$ is the standard normal cumulative distribution function (probit function)

a_a is the discrimination parameter for item a

b_a is the difficulty parameter for item a