Modeling approach(es) for multivariate binary response

Steve and Jonathan

2019-09-23

Background

- Longitudinal (2003 - 2015) NUHDSS covering Korogocho and Viwandani

Background

- Longitudinal (2003 - 2015) NUHDSS covering Korogocho and Viwandani
- Predictors

Background

- Longitudinal (2003 - 2015) NUHDSS covering Korogocho and Viwandani
- Predictors
  - Slum area

Background

- Longitudinal (2003 - 2015) NUHDSS covering Korogocho and Viwandani
- Predictors
  - Slum area
  - Interview year

Background

- Longitudinal (2003 - 2015) NUHDSS covering Korogocho and Viwandani
- Predictors
  - Slum area
  - Interview year
  - Age

Background

- Longitudinal (2003 - 2015) NUHDSS covering Korogocho and Viwandani
- Predictors
  - Slum area
  - Interview year
  - Age
  - Gender

Background

- Longitudinal (2003 - 2015) NUHDSS covering Korogocho and Viwandani
- Predictors
  - Slum area
  - Interview year
  - Age
  - Gender
  - Ethnicity

Background

- Longitudinal (2003 - 2015) NUHDSS covering Korogocho and Viwandani
- Predictors
  - Slum area
  - Interview year
  - Age
  - Gender
  - Ethnicity
  - Household size

Background

- Longitudinal (2003 - 2015) NUHDSS covering Korogocho and Viwandani
- Predictors
  - Slum area
  - Interview year
  - Age
  - Gender
  - Ethnicity
  - Household size
  - Wealth index

Background

- Longitudinal (2003 - 2015) NUHDSS covering Korogocho and Viwandani
- Predictors
  - Slum area
  - Interview year
  - Age
  - Gender
  - Ethnicity
  - Household size
  - Wealth index
  - Household expenditure

Background

- Longitudinal (2003 - 2015) NUHDSS covering Korogocho and Viwandani
- Predictors
  - Slum area
  - Interview year
  - Age
  - Gender
  - Ethnicity
  - Household size
  - Wealth index
  - Household expenditure
- Response(s): Three WASH variables were created as per WHO definition

Background

- Longitudinal (2003 - 2015) NUHDSS covering Korogocho and Viwandani
- Predictors
  - Slum area
  - Interview year
  - Age
  - Gender
  - Ethnicity
  - Household size
  - Wealth index
  - Household expenditure
- Response(s): Three WASH variables were created as per WHO definition
  - Drinking water source

Background

- Longitudinal (2003 - 2015) NUHDSS covering Korogocho and Viwandani
- Predictors
  - Slum area
  - Interview year
  - Age
  - Gender
  - Ethnicity
  - Household size
  - Wealth index
  - Household expenditure
- Response(s): Three WASH variables were created as per WHO definition
  - Drinking water source
  - Toilet facility type

Background

- Longitudinal (2003 - 2015) NUHDSS covering Korogocho and Viwandani
- Predictors
  - Slum area
  - Interview year
  - Age
  - Gender
  - Ethnicity
  - Household size
  - Wealth index
  - Household expenditure
- Response(s): Three WASH variables were created as per WHO definition
  - Drinking water source
  - Toilet facility type
  - Garbage disposal method

Objective

- The aim is to investigate the contribution of demographic, social and economic factors to improved water, sanitation and hygiene (WASH) among the urban poor.

Problems

- How do we account for the repeated measurements within the households across the years?
  - Model the wash variables separately
  - Pick one of the WASH indicator and treat the remaining two as fixed covariates
- The two approaches are not accounting for the unmeasured variations and correlation among the WASH variables

Problems

- How do we account for the repeated measurements within the households across the years?
  - Model the wash variables separately
  - Pick one of the WASH indicator and treat the remaining two as fixed covariates
- The two approaches are not accounting for the unmeasured variations and correlation among the WASH variables
- But we need some understanding of data generation process

Problems

- How do we account for the repeated measurements within the households across the years?
  - Model the wash variables separately
  - Pick one of the WASH indicator and treat the remaining two as fixed covariates
- The two approaches are not accounting for the unmeasured variations and correlation among the WASH variables
- But we need some understanding of data generation process
  - Some simulations

Data exploration

- WASH variables (services) are binary (`0 = unimproved` and `1 = improved`)
  - Each household was surveyed once per year
  - For some HH, the services have improved or unimproved for different years
- Aggregated by `year`, all HH have varying `wealth_index`

Simulations

Assumptions

- Use year as the grouping variable
- Each household has its own year effect on the intercepts (random-intercept)

Model

$$y_{ijk} = \beta 0_{jk} + \beta 1_{jk} x_i + \epsilon_{ik}$$

Simulations (Contd. . .)

- y_{ijk} is the simulated value of the ith household in the jth year of the kth service

Simulations (Contd. . .)

- y_{ijk} is the simulated value of the ith household in the jth year of the kth service
  - i goes from 1 to the number of households

Simulations (Contd. . .)

- y_{ijk} is the simulated value of the ith household in the jth year of the kth service
  - i goes from 1 to the number of households
  - $j = 1, 2, \cdots, 30$ indexes the number of years

Simulations (Contd. . .)

- y_{ijk} is the simulated value of the ith household in the jth year of the kth service
  - i goes from 1 to the number of households
  - $j = 1, 2, \cdots, 30$ indexes the number of years
  - $k = 1, 2, 3$ indexes the number of services

Simulations (Contd. . .)

- y_{ijk} is the simulated value of the ith household in the jth year of the kth service
  - i goes from 1 to the number of households
  - $j = 1, 2, \cdots, 30$ indexes the number of years
  - $k = 1, 2, 3$ indexes the number of services
- $\beta 0_{jk}$ is the random-intercept effect of the jth year of the kth service on the response

Simulations (Contd. . .)

- y_{ijk} is the simulated value of the ith household in the jth year of the kth service
  - i goes from 1 to the number of households
  - $j = 1, 2, \cdots, 30$ indexes the number of years
  - $k = 1, 2, 3$ indexes the number of services
- $\beta 0_{jk}$ is the random-intercept effect of the jth year of the kth service on the response
  - $\beta 0_{jk} \sim MVN(\mu 0_k, \Sigma_{tk})$; $t = 1, 2, 3$

Simulations (Contd. . .)

- $\beta 1_{jk}$ is the random-slope effect of the jth year of the kth service on the response

Simulations (Contd. . .)

- $\beta 1_{jk}$ is the random-slope effect of the jth year of the kth service on the response
  - $\beta 1_{jk} \sim MVN(\mu 1_k, \Sigma_{tk})$

Simulations (Contd. . .)

- $\beta 1_{jk}$ is the random-slope effect of the jth year of the kth service on the response
  - $\beta 1_{jk} \sim MVN(\mu 1_k, \Sigma_{tk})$
- ϵ_{ik} is the household-level random effect (residual error term) of the kth service

Simulations (Contd. . .)

- $\beta 1_{jk}$ is the random-slope effect of the jth year of the kth service on the response
  - $\beta 1_{jk} \sim MVN(\mu 1_k, \Sigma_{tk})$
- ϵ_{ik} is the household-level random effect (residual error term) of the kth service
  - $\epsilon_{ik} \sim MVN(\mathbf{0}, \Sigma_{tk})$

Simulations (Contd. . .)

- $\beta 1_{jk}$ is the random-slope effect of the jth year of the kth service on the response
  - $\beta 1_{jk} \sim MVN(\mu 1_k, \Sigma_{tk})$
- ϵ_{ik} is the household-level random effect (residual error term) of the kth service
  - $\epsilon_{ik} \sim MVN(\mathbf{0}, \Sigma_{tk})$
- x_j is the ith predictor value

Simulations (Contd. . .)

- $\beta 1_{jk}$ is the random-slope effect of the jth year of the kth service on the response
  - $\beta 1_{jk} \sim MVN(\mu 1_k, \Sigma_{tk})$
- ϵ_{ik} is the household-level random effect (residual error term) of the kth service
  - $\epsilon_{ik} \sim MVN(\mathbf{0}, \Sigma_{tk})$
- x_j is the ith predictor value
- y_{ijk} drawn from binomial distribution, with probability, $p_k = plogis(y_{ijk})$

Results

- See the results are here