Modeling approach(es) for multivariate binary response

Steve and Jonathan

2019-09-27

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

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
- We want explore modeling approaches, but we need to understand what we think is working.

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
- We want explore modeling approaches, but we need to understand what we think is working.
  - Simulation-based validation

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.

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

- Each household has its own year effect on the intercepts (random-intercept)
  - These are correlated
- There is a single measured covariate (corresponding to wealth)

Model

$$y_{hts} = \beta_{0ts} + \beta_{xs} x_{ht} + \epsilon_{hs}$$

- y_{hts} is the simulated value of household `h` in year `t` of service `s`

Model

$$y_{hts} = \beta_{0ts} + \beta_{xs} x_{ht} + \epsilon_{hs}$$

- y_{hts} is the simulated value of household `h` in year `t` of service `s`
  - h goes from 1 to the number of households

Model

$$y_{hts} = \beta_{0ts} + \beta_{xs} x_{ht} + \epsilon_{hs}$$

- y_{hts} is the simulated value of household $\mathtt{h}$ in year $\mathtt{t}$ of service $\mathtt{s}$
  - h goes from 1 to the number of households
  - $t = 1, 2, \cdots, 30$ years

Model

$$y_{hts} = \beta_{0ts} + \beta_{xs} x_{ht} + \epsilon_{hs}$$

- y_{hts} is the simulated value of household `h` in year `t` of service `s`
  - h goes from 1 to the number of households
  - $t = 1, 2, \cdots, 30$ years
  - $s = 1, 2, 3$ indexes the services

Model

$$y_{hts} = \beta_{0ts} + \beta_{xs} x_{ht} + \epsilon_{hs}$$

- y_{hts} is the simulated value of household `h` in year `t` of service `s`
  - h goes from 1 to the number of households
  - $t = 1, 2, \cdots, 30$ years
  - $s = 1, 2, 3$ indexes the services
- β_{0ts} is the effect of year and service type on the response

Model

$$y_{hts} = \beta_{0ts} + \beta_{xs} x_{ht} + \epsilon_{hs}$$

- y_{hts} is the simulated value of household `h` in year `t` of service `s`
  - h goes from 1 to the number of households
  - $t = 1, 2, \cdots, 30$ years
  - $s = 1, 2, 3$ indexes the services
- β_{0ts} is the effect of year and service type on the response
  - these are correlated

Model

$$y_{hts} = \beta_{0ts} + \beta_{xs} x_{ht} + \epsilon_{hs}$$

- y_{hts} is the simulated value of household `h` in year `t` of service `s`
  - h goes from 1 to the number of households
  - $t = 1, 2, \cdots, 30$ years
  - $s = 1, 2, 3$ indexes the services
- β_{0ts} is the effect of year and service type on the response
  - these are correlated
  - $\beta_{0ts} \sim MVN(\mu_0, \Sigma_0)$

Model

$$y_{hts} = \beta_{0ts} + \beta_{xs} x_{ht} + \epsilon_{hs}$$

- y_{hts} is the simulated value of household `h` in year `t` of service `s`
  - h goes from 1 to the number of households
  - $t = 1, 2, \cdots, 30$ years
  - $s = 1, 2, 3$ indexes the services
- β_{0ts} is the effect of year and service type on the response
  - these are correlated
  - $\beta_{0ts} \sim MVN(\mu_0, \Sigma_0)$
- β_{xs} is the (fixed) effect of x on service `s`

Model

$$y_{hts} = \beta_{0ts} + \beta_{xs} x_{ht} + \epsilon_{hs}$$

- y_{hts} is the simulated value of household $\mathtt{h}$ in year $\mathtt{t}$ of service $\mathtt{s}$
  - h goes from 1 to the number of households
  - $t = 1, 2, \cdots, 30$ years
  - $s = 1, 2, 3$ indexes the services
- β_{0ts} is the effect of year and service type on the response
  - these are correlated
  - $\beta_{0ts} \sim MVN(\mu_0, \Sigma_0)$
- β_{xs} is the (fixed) effect of x on service $\mathtt{s}$
  - $\beta_{xs} \sim (\beta_{1s}); s = 1, \cdots, 3$

Model

$$y_{hts} = \beta_{0ts} + \beta_{xs} x_{ht} + \epsilon_{hs}$$

- y_{hts} is the simulated value of household `h` in year `t` of service `s`
  - h goes from 1 to the number of households
  - $t = 1, 2, \cdots, 30$ years
  - $s = 1, 2, 3$ indexes the services
- β_{0ts} is the effect of year and service type on the response
  - these are correlated
  - $\beta_{0ts} \sim MVN(\mu_0, \Sigma_0)$
- β_{xs} is the (fixed) effect of x on service `s`
  - $\beta_{xs} \sim (\beta_{1s}); s = 1, \cdots, 3$
- ϵ_{hs} is the household-level random effect

Model

$$y_{hts} = \beta_{0ts} + \beta_{xs} x_{ht} + \epsilon_{hs}$$

- y_{hts} is the simulated value of household `h` in year `t` of service `s`
  - h goes from 1 to the number of households
  - $t = 1, 2, \cdots, 30$ years
  - $s = 1, 2, 3$ indexes the services
- β_{0ts} is the effect of year and service type on the response
  - these are correlated
  - $\beta_{0ts} \sim MVN(\mu_0, \Sigma_0)$
- β_{xs} is the (fixed) effect of x on service `s`
  - $\beta_{xs} \sim (\beta_{1s}); s = 1, \cdots, 3$
- ϵ_{hs} is the household-level random effect
  - $\epsilon_{hs} \sim MVN(\mathbf{0}, \Sigma)$

Model

$$y_{hts} = \beta_{0ts} + \beta_{xs} x_{ht} + \epsilon_{hs}$$

- y_{hts} is the simulated value of household $\mathtt{h}$ in year $\mathtt{t}$ of service $\mathtt{s}$
  - h goes from 1 to the number of households
  - $t = 1, 2, \cdots, 30$ years
  - $s = 1, 2, 3$ indexes the services
- β_{0ts} is the effect of year and service type on the response
  - these are correlated
  - $\beta_{0ts} \sim MVN(\mu_0, \Sigma_0)$
- β_{xs} is the (fixed) effect of x on service $\mathtt{s}$
  - $\beta_{xs} \sim (\beta_{1s}); s = 1, \cdots, 3$
- ϵ_{hs} is the household-level random effect
  - $\epsilon_{hs} \sim MVN(\mathbf{0}, \Sigma)$
- Observed values ($\{0, 1\}$) are drawn from binomial distribution with probability $\mathrm{plogis}(y_{hts})$

Results

- The results are here