TBD

Arjun Dhatt, Benjamin Draskovic, Yiqu Ding, Gantavya Gupta

10/16/2020

Abstract

Introduction

Data

Model

We are interested in how variables such as age, family income, education level, and population center affect a woman's decision to have children. Due to the large sample size and the nature of GSS, the sample represents the population well. The model that we are using is logistic regression, it works well for our response variable, which is a categorical variable, and it incorporates both numerical and categorical explanatory variables.

Logistic regression estimates $\beta_0...\beta_k$ in the following equation:

$$log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k$$

In our case, it estimates $\beta_{age}, \beta_{inc}, \beta_{edu}, \beta_{pop}$ in:

$$log(\frac{p}{1-p}) = \beta_0 + \beta_{age}x_{age} + \beta_{inc}x_{inc} + \beta_{edu}x_{edu} + \beta_{pop}x_{pop}$$

We use `glm` from `stats`package to fir the model to our data. We use use `as.factor()` to incorporate dummy variables for all the categorical variables: family income, education level and the type of population center.

Results

Table x summaries our model results:

Table 1: Summary of Losgistic Estimates

variable	estimate
age	0.060790
income greater than $125,000	0.161424
income between $25,000 to $49,999	-0.680099
income between $50,000 to $74,999	-0.449230
income between $75,000 to $99,999	-0.197017
income less than $25,000	-0.872898
high school or less education	-0.102521
University Graduate	-0.779989
University Undergraduate	-0.468140
Poplation centered at PEI	0.276375
Rural areas and small population centres(non CMA/CA)	0.526007

Discussion

References

- https://mc-stan.org/rstanarm/articles/mrp.html
- https://www.monicaalexander.com/posts/2019-08-07-mrp/