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 `as.factor()` to incorporate dummy variables for all the categorical variables: family income, education level and the type of population center. For each categorical variable with n levels, we need n-1 dummy variables to fully study its influence on our response variable.

The dummy variables setting ups are stated in table x, y and z.

Table 1: Dummy Variable Coding Set Up for Income Levels

income.level	inc1	inc2	inc3	inc4	inc5	inc6
greater than $125,000	1	0	0	0	0	0
between $25,000 to $49,999	0	1	0	0	0	0
between $50,000 to $74,999	0	0	1	0	0	0
between $75,000 to $99,999	0	0	0	1	0	0
less than $25,000	0	0	0	0	1	0
between $100,000 tp $124,999	0	0	0	0	0	0

Table 2: Dummy Variable Coding Set Up for Education Levels

education.level	HS	graduate	under	college
high school or less education	1	0	0	0
University Graduate	0	1	0	0
University Undergraduate	0	0	1	0
college/trade	0	0	0	0

Table 3: Dummy Variable Coding Set Up for Population Center

population.center	PEI	rural	urban
Poplation centered at PEI	1	0	0
Rural areas and small population centres(non CMA/CA)	0	1	0
Larger urban population centres (CMA/CA)	0	0	0

Results

Table x summaries our model results:

Table 4: 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

In an equation, this means

$$log(\frac{\hat{p}}{1-\hat{p}}) = \hat{\beta}_0 + \hat{\beta}_{age}x_{age} + \hat{\beta}_{inc_1}x_{inc_1} + \hat{\beta}_{inc_2}x_{inc_2} + \hat{\beta}_{inc_3}x_{inc_3}$$

$$+\hat{\beta}_{inc_4}x_{inc_4} + \hat{\beta}_{inc_5}x_{inc_5} + \hat{\beta}_{HS}x_{HS} + \hat{\beta}_{graduate}x_{graduate} + \hat{\beta}_{under}x_{under}$$

$$+\hat{\beta}_{PEI}x_{PEI} + \hat{\beta}_{rural}x_{rural}$$

Notice that we have incorporated our categorical variables using dummy variables. “inc1”, “inc2”, “inc3”, “inc4”, “inc5” represent the six income categories; “HS”, “graduate” and “under” describe the four education levels; “PEI” and “rural” represent the three population center.

The model predicts the following result for the probability a woman has children p, rounded to three decimal places:

$$log(\frac{\hat{p}}{1-\hat{p}}) = -1.51 + 0.061x_{age} + 0.161x_{inc_1} - 0.68x_{inc_2} - 0.449x_{inc_3}$$

$$-0.197x_{inc_4} - 0.873x_{inc_5} - 0.103x_{HS} - 0.78x_{graduate} - 0.468x_{under}$$

$$+0.276x_{PEI} + 0.526x_{rural}$$

Note that the interpretation of the dummy variables' prediction results is by comparing it to a certain level that is not in the equation above. A $\hat{\beta}_{HS} = -0.103$ does not mean the coefficient for a high school level education is -0.103. It represents the difference of influence (on childbirth decision) between a woman who has a college level of education and a high school level of education. -0.103 indicates that a woman who has a high school degree or below is less likely to have children than a woman who has a college degree.

```
##
## Call:
## glm(formula = child ~ age + as.factor(income_family) + as.factor(education_level) +
##     as.factor(pop_center), family = "binomial", data = my_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6819  -0.7605   0.4728   0.7270   1.7675
##
## Coefficients:
##                                                                  Estimate
## (Intercept)                                                     -1.513991
## age                                                              0.060790
## as.factor(income_family)$125,000 and more                        0.161424
## as.factor(income_family)$25,000 to $49,999                      -0.680099
## as.factor(income_family)$50,000 to $74,999                      -0.449230
## as.factor(income_family)$75,000 to $99,999                      -0.197017
## as.factor(income_family)Less than $25,000                       -0.872898
## as.factor(education_level)HS or less                            -0.102521
## as.factor(education_level)University Graduate                   -0.779989
## as.factor(education_level)University Undergraduate              -0.468140
## as.factor(pop_center)Prince Edward Island                        0.276375
## as.factor(pop_center)Rural areas and small population centres (non CMA/CA)  0.526007
##                                                                 Std. Error
## (Intercept)                                                       0.107165
## age                                                               0.001515
## as.factor(income_family)$125,000 and more                         0.092512
## as.factor(income_family)$25,000 to $49,999                        0.095024
## as.factor(income_family)$50,000 to $74,999                        0.096823
## as.factor(income_family)$75,000 to $99,999                        0.100131
## as.factor(income_family)Less than $25,000                         0.103018
## as.factor(education_level)HS or less                              0.061989
## as.factor(education_level)University Graduate                     0.087481
## as.factor(education_level)University Undergraduate                0.065554
## as.factor(pop_center)Prince Edward Island                         0.135165
## as.factor(pop_center)Rural areas and small population centres (non CMA/CA)   0.069172
##                                                                 z value
## (Intercept)                                                     -14.128
## age                                                              40.133
```

```
## as.factor(income_family)$125,000 and more                                    1.745
## as.factor(income_family)$25,000 to $49,999                                  -7.157
## as.factor(income_family)$50,000 to $74,999                                  -4.640
## as.factor(income_family)$75,000 to $99,999                                  -1.968
## as.factor(income_family)Less than $25,000                                   -8.473
## as.factor(education_level)HS or less                                        -1.654
## as.factor(education_level)University Graduate                               -8.916
## as.factor(education_level)University Undergraduate                          -7.141
## as.factor(pop_center)Prince Edward Island                                    2.045
## as.factor(pop_center)Rural areas and small population centres (non CMA/CA)   7.604
##                                                                            Pr(>|z|)
## (Intercept)                                                                 < 2e-16
## age                                                                         < 2e-16
## as.factor(income_family)$125,000 and more                                    0.0810
## as.factor(income_family)$25,000 to $49,999                                  8.24e-13
## as.factor(income_family)$50,000 to $74,999                                  3.49e-06
## as.factor(income_family)$75,000 to $99,999                                   0.0491
## as.factor(income_family)Less than $25,000                                   < 2e-16
## as.factor(education_level)HS or less                                         0.0982
## as.factor(education_level)University Graduate                               < 2e-16
## as.factor(education_level)University Undergraduate                          9.25e-13
## as.factor(pop_center)Prince Edward Island                                    0.0409
## as.factor(pop_center)Rural areas and small population centres (non CMA/CA) 2.86e-14
##
## (Intercept)                                                                ***
## age                                                                        ***
## as.factor(income_family)$125,000 and more                                  .
## as.factor(income_family)$25,000 to $49,999                                 ***
## as.factor(income_family)$50,000 to $74,999                                 ***
## as.factor(income_family)$75,000 to $99,999                                 *
## as.factor(income_family)Less than $25,000                                  ***
## as.factor(education_level)HS or less                                       .
## as.factor(education_level)University Graduate                              ***
## as.factor(education_level)University Undergraduate                         ***
## as.factor(pop_center)Prince Edward Island                                  *
## as.factor(pop_center)Rural areas and small population centres (non CMA/CA) ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 12733  on 10915  degrees of freedom
## Residual deviance: 10501  on 10904  degrees of freedom
## AIC: 10525
##
## Number of Fisher Scoring iterations: 4
```

Discussion

References

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