****

**VIRGINIA COMMONWEALTH UNIVERSITY**

**Statistical analysis and modelling (SCMA 632)**

**A3-Limited dependent variable Models**

**Part-B**

**Bala Vignesh Aravindan**

**V01106579**

**Date of Submission: 30-06-2024**

**CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Sl. No.** | **Title** | **Page No.** |
| **1.** | Introduction | **1** |
| **2** | Results | **2-11** |
| **3.** | Interpretations | **12-13** |
| **4.** | Recommendations | **13** |
| **5.** | Codes | **14-17** |

**Probit Regression Analysis of Non-Vegetarian Food Consumption in India**

**Introduction:**

The objective of this analysis is to understand the factors influencing the consumption of non-vegetarian food in India. The data used for this analysis is obtained from the National Sample Survey Office (NSSO) and covers a sample of 101047 households across various states and union territories in India. The outcome variable in this analysis is a binary variable indicating whether a household consumes non-vegetarian food or not. The predictor variables include socio-economic and demographic factors such as age, education, income, religion, social group, and region.

**Objective:**

The objective of this analysis is to identify the key factors influencing the consumption of non-vegetarian food in India. The analysis will help in understanding the dietary patterns and preferences of households in India and can be used to inform policy decisions related to food and nutrition.

**Business Significance:**

The results of this analysis can be used by businesses in the food and beverage industry to understand the market potential for non-vegetarian food products in India. The analysis can help businesses in identifying the target market segments and developing marketing strategies to cater to the needs and preferences of these segments

**Results:**

Probit Regression Results

==============================================================================

Dep. Variable: non\_veg No. Observations: 101047

Model: Probit Df Residuals: 100591

Method: MLE Df Model: 455

Date: Wed, 26 Jun 2024 Pseudo R-squ.: inf

Time: 23:06:26 Log-Likelihood: -1.2298e+05

converged: False LL-Null: 0.0000

Covariance Type: nonrobust LLR p-value: 1.000

========================================================================================================

coef std err z P>|z| [0.025 0.975]

--------------------------------------------------------------------------------------------------------

slno 0.0028 nan nan nan nan nan

Round\_Centre 0.1783 nan nan nan nan nan

FSU\_number -0.0119 nan nan nan nan nan

Round 12.1377 nan nan nan nan nan

Schedule\_Number 1.8018 nan nan nan nan nan

Sample 0.1783 nan nan nan nan nan

Sector -179.0167 nan nan nan nan nan

state 0.5909 nan nan nan nan nan

State\_Region -0.0678 nan nan nan nan nan

District -0.0058 nan nan nan nan nan

Stratum\_Number 0.0111 nan nan nan nan nan

Sub\_Stratum -0.0180 nan nan nan nan nan

Schedule\_type 0.1783 nan nan nan nan nan

Sub\_Round 0.0514 nan nan nan nan nan

Sub\_Sample -0.1364 nan nan nan nan nan

FOD\_Sub\_Region 0.0008 nan nan nan nan nan

Hamlet\_Group\_Sub\_Block -0.2110 nan nan nan nan nan

t 3.775e-34 nan nan nan nan nan

X\_Stage\_Stratum 0.0322 nan nan nan nan nan

HHS\_No 0.0991 nan nan nan nan nan

Level 0.9027 nan nan nan nan nan

Filler 0 nan nan nan nan nan

hhdsz 0.0204 nan nan nan nan nan

NIC\_2008 1.229e-06 nan nan nan nan nan

NCO\_2004 -0.0002 nan nan nan nan nan

HH\_type 0.0038 nan nan nan nan nan

Religion -0.0381 nan nan nan nan nan

Social\_Group 0.0546 nan nan nan nan nan

Whether\_owns\_any\_land -0.0241 nan nan nan nan nan

Type\_of\_land\_owned -0.0592 nan nan nan nan nan

Land\_Owned 0.0006 nan nan nan nan nan

Land\_Leased\_in 0.0007 nan nan nan nan nan

Otherwise\_possessed -3.958e-05 nan nan nan nan nan

Land\_Leased\_out -0.0002 nan nan nan nan nan

Land\_Total\_possessed -0.0002 nan nan nan nan nan

During\_July\_June\_Cultivated -1.506e-06 nan nan nan nan nan

During\_July\_June\_Irrigated -0.0010 nan nan nan nan nan

NSS 3.6545 nan nan nan nan nan

NSC -1.7682 nan nan nan nan nan

MLT -2.273e-07 nan nan nan nan nan

land\_tt -3.716e-06 nan nan nan nan nan

Cooking\_code 0.0244 nan nan nan nan nan

Lighting\_code -0.0441 nan nan nan nan nan

Dwelling\_unit\_code 0.0339 nan nan nan nan nan

Regular\_salary\_earner -0.2603 nan nan nan nan nan

Perform\_Ceremony 0.6595 nan nan nan nan nan

Meals\_seved\_to\_non\_hhld\_members 0.0034 nan nan nan nan nan

Possess\_ration\_card 1.3128 nan nan nan nan nan

Type\_of\_ration\_card 0.4614 nan nan nan nan nan

MPCE\_URP -0.0001 nan nan nan nan nan

MPCE\_MRP 8.135e-05 nan nan nan nan nan

Person\_Srl\_No 0.1782 nan nan nan nan nan

Relation 0.1782 nan nan nan nan nan

Sex 0.0014 nan nan nan nan nan

Age -0.0040 nan nan nan nan nan

Marital\_Status 0.0057 nan nan nan nan nan

Education 0.0147 nan nan nan nan nan

Days\_Stayed\_away 0.0234 nan nan nan nan nan

No\_of\_Meals\_per\_day -0.0026 nan nan nan nan nan

Meals\_School 0.0154 nan nan nan nan nan

Meals\_Employer -0.0573 nan nan nan nan nan

Meals\_Others -0.0036 nan nan nan nan nan

Meals\_Payment -0.0133 nan nan nan nan nan

Meals\_At\_Home 0.0300 nan nan nan nan nan

Item\_Code 0.0021 nan nan nan nan nan

Source\_Code 0.0695 nan nan nan nan nan

ricepds\_q -0.6187 nan nan nan nan nan

riceos\_q -0.4769 nan nan nan nan nan

ricetotal\_q -1.0936 nan nan nan nan nan

chira\_q -1.0920 nan nan nan nan nan

khoi\_q -6.8075 nan nan nan nan nan

muri\_q 6.6961 nan nan nan nan nan

ricepro\_q -1.6361 nan nan nan nan nan

riceGT\_q -3.9360 nan nan nan nan nan

Wheatpds\_q -2.9465 nan nan nan nan nan

wheatos\_q -2.6662 nan nan nan nan nan

wheattotal\_q -5.6127 nan nan nan nan nan

maida\_q -7.6115 nan nan nan nan nan

suji\_q -7.7922 nan nan nan nan nan

sewai\_q -3.3059 nan nan nan nan nan

bread\_q -7.3510 nan nan nan nan nan

wheatp\_q 34.4970 nan nan nan nan nan

wheatGT\_q 2.8229 nan nan nan nan nan

jowarp\_q -5.7628 nan nan nan nan nan

bajrap\_q -8.7401 nan nan nan nan nan

maizep\_q -8.6224 nan nan nan nan nan

barleyp\_q -4.0570 nan nan nan nan nan

milletp\_q 44.2618 nan nan nan nan nan

ragip\_q -6.6831 nan nan nan nan nan

cerealot\_q -4.8730 nan nan nan nan nan

cerealtot\_q 4.4108 nan nan nan nan nan

cerealsub\_q -2.1883 nan nan nan nan nan

cerealstt\_q 4.4108 nan nan nan nan nan

arhar\_q -10.3463 nan nan nan nan nan

gramdal\_q 5.6719 nan nan nan nan nan

gramwholep\_q -5.6140 nan nan nan nan nan

gramGT\_q 0.0579 nan nan nan nan nan

moong\_q -0.3715 nan nan nan nan nan

masur\_q 0.6880 nan nan nan nan nan

urd\_q 4.9967 nan nan nan nan nan

peasdal\_q 12.6987 nan nan nan nan nan

khesari\_q -2.6570 nan nan nan nan nan

otpulse\_q -2.3116 nan nan nan nan nan

gramp\_q 9.8969 nan nan nan nan nan

besan\_q -8.6658 nan nan nan nan nan

pulsep\_q -2.4203 nan nan nan nan nan

pulsestot\_q 1.5653 nan nan nan nan nan

pulsestt\_q 1.5653 nan nan nan nan nan

milk\_q 0 nan nan nan nan nan

babyfood\_q -3.2191 nan nan nan nan nan

milkcond\_q -6.5003 nan nan nan nan nan

curd\_q -5.7117 nan nan nan nan nan

ghee\_q 1.2895 nan nan nan nan nan

butter\_q -7.0750 nan nan nan nan nan

icecream\_q 27.6924 nan nan nan nan nan

otmilkp\_q 0 nan nan nan nan nan

Milktotal\_q 0 nan nan nan nan nan

milkprott\_q 0 nan nan nan nan nan

vanas\_q 6.4727 nan nan nan nan nan

musoil\_q -16.1355 nan nan nan nan nan

gnoil\_q 7.3328 nan nan nan nan nan

cocooil\_q -1.4546 nan nan nan nan nan

edioilothr\_q 6.8850 nan nan nan nan nan

edibletotal\_q 5.4580 nan nan nan nan nan

ediblest\_q -4.6979 nan nan nan nan nan

eggsno\_q 2.0855 nan nan nan nan nan

fishprawn\_q 2.1898 nan nan nan nan nan

goatmeat\_q -3.3293 nan nan nan nan nan

beef\_q -3.6013 nan nan nan nan nan

pork\_q 15.5792 nan nan nan nan nan

chicken\_q 0.5296 nan nan nan nan nan

othrbirds\_q -6.0421 nan nan nan nan nan

nonvegtotal\_q 1.3847 nan nan nan nan nan

emftt\_q 0 nan nan nan nan nan

potato\_q 6.7111 nan nan nan nan nan

onion\_q -0.3682 nan nan nan nan nan

tamato\_q 1.1439 nan nan nan nan nan

brinjal\_q -0.8531 nan nan nan nan nan

radish\_q -0.7459 nan nan nan nan nan

carrot\_q 1.1555 nan nan nan nan nan

palak\_q -1.3342 nan nan nan nan nan

chillig\_q 0.4041 nan nan nan nan nan

bhindi\_q 1.7833 nan nan nan nan nan

parwal\_q -1.0943 nan nan nan nan nan

cauli\_q 2.6894 nan nan nan nan nan

cabbage\_q 1.3320 nan nan nan nan nan

pumpkin\_q 1.4123 nan nan nan nan nan

peas\_q 0.2243 nan nan nan nan nan

fbeans\_q 0.5519 nan nan nan nan nan

lemonno\_q -2.5003 nan nan nan nan nan

otveg\_q -0.0598 nan nan nan nan nan

vegtt\_q -0.7091 nan nan nan nan nan

bananano\_q 3.0334 nan nan nan nan nan

jackfruit\_q 55.0475 nan nan nan nan nan

watermel\_q -0.0148 nan nan nan nan nan

pineaplno\_q -0.0277 nan nan nan nan nan

cocono\_q 6.4247 nan nan nan nan nan

cocogno\_q 29.8048 nan nan nan nan nan

guava\_q 0 nan nan nan nan nan

sighara\_q 0 nan nan nan nan nan

orangeno\_q 0 nan nan nan nan nan

papayar\_q 0 nan nan nan nan nan

mango\_q -0.3030 nan nan nan nan nan

kharbooz\_q -0.5822 nan nan nan nan nan

pears\_q -0.3203 nan nan nan nan nan

berries\_q 2.8441 nan nan nan nan nan

leechi\_q 15.6070 nan nan nan nan nan

apple\_q 7.4593 nan nan nan nan nan

grapes\_q 0.5925 nan nan nan nan nan

otfruits\_q 0.1521 nan nan nan nan nan

fruitstt\_q 0 nan nan nan nan nan

fruitt\_total 0 nan nan nan nan nan

cocodf\_q -3.6380 nan nan nan nan nan

gnutdf\_q 25.7758 nan nan nan nan nan

datesdf\_q -10.2001 nan nan nan nan nan

cashewdf\_q -5.6940 nan nan nan nan nan

walnutdf\_q -7.2535 nan nan nan nan nan

otnutsdf\_q 30.0168 nan nan nan nan nan

kishmish\_q -8.3875 nan nan nan nan nan

otherdf\_q -24.3389 nan nan nan nan nan

dryfruitstotal\_q 11.8816 nan nan nan nan nan

dftt\_q 0 nan nan nan nan nan

sugarpds\_q 11.8003 nan nan nan nan nan

sugaros\_q 10.3310 nan nan nan nan nan

sugarst\_q 12.9117 nan nan nan nan nan

gur\_q 23.2428 nan nan nan nan nan

misri\_q 28.8169 nan nan nan nan nan

honey\_q -83.5251 nan nan nan nan nan

sugartotal\_q 0 nan nan nan nan nan

sugartt\_q 3.6096 nan nan nan nan nan

salt\_q -31.4658 nan nan nan nan nan

ginger\_q -3.4219 nan nan nan nan nan

garlic\_q -4.8153 nan nan nan nan nan

jeera\_q -2.3033 nan nan nan nan nan

dhania\_q 0.5244 nan nan nan nan nan

turnmeric\_q 0.0284 nan nan nan nan nan

blackpepper\_q -0.0049 nan nan nan nan nan

drychilly\_q -0.4272 nan nan nan nan nan

tamarind\_q 1.5695 nan nan nan nan nan

currypowder\_q 2.5847 nan nan nan nan nan

oilseeds\_q 0 nan nan nan nan nan

spicesothr\_q 2.8675 nan nan nan nan nan

spicetot\_q -5.3405 nan nan nan nan nan

spicestotal\_q -4.2731 nan nan nan nan nan

teacupno\_q -5.3154 nan nan nan nan nan

tealeaf\_q 0 nan nan nan nan nan

teatotal\_q 0 nan nan nan nan nan

cofeeno\_q 0 nan nan nan nan nan

coffeepwdr\_q 0 nan nan nan nan nan

cofeetotal\_q 0 nan nan nan nan nan

ice\_q 0 nan nan nan nan nan

coldbvrg\_q 0 nan nan nan nan nan

juice\_q 0 nan nan nan nan nan

othrbevrg\_q 0 nan nan nan nan nan

bevergest\_q 0 nan nan nan nan nan

Biscuits\_q 0 nan nan nan nan nan

preparedsweet\_q 0 nan nan nan nan nan

pickle\_q 0 nan nan nan nan nan

sauce\_jam\_q 0.8332 nan nan nan nan nan

Othrprocessed\_q 0.0029 nan nan nan nan nan

Beveragestotal\_q 0 nan nan nan nan nan

ricepds\_v 0.8360 nan nan nan nan nan

riceos\_v 0.0404 nan nan nan nan nan

ricetotal\_v 0.0301 nan nan nan nan nan

chira\_v 0.0746 nan nan nan nan nan

khoi\_v 0.0633 nan nan nan nan nan

muri\_v 0.1627 nan nan nan nan nan

ricepro\_v -0.0039 nan nan nan nan nan

riceGT\_v 0.0992 nan nan nan nan nan

Wheatpds\_v 0.4010 nan nan nan nan nan

wheatos\_v 0.0908 nan nan nan nan nan

wheattotal\_v 0.0890 nan nan nan nan nan

maida\_v 0.1901 nan nan nan nan nan

suji\_v 0.3276 nan nan nan nan nan

sewai\_v 0.2066 nan nan nan nan nan

bread\_v 0.1911 nan nan nan nan nan

wheatp\_v 0.2680 nan nan nan nan nan

wheatGT\_v -0.9465 nan nan nan nan nan

jowarp\_v 0.2264 nan nan nan nan nan

bajrap\_v 0.5003 nan nan nan nan nan

maizep\_v 0.7333 nan nan nan nan nan

barleyp\_v 0.7568 nan nan nan nan nan

milletp\_v 0.3448 nan nan nan nan nan

ragip\_v -0.2929 nan nan nan nan nan

cerealot\_v 0.6078 nan nan nan nan nan

cerealtot\_v 0.4855 nan nan nan nan nan

cerealsub\_v 3.7796 nan nan nan nan nan

cerealstt\_v 0.0891 nan nan nan nan nan

arhar\_v 3.7821 nan nan nan nan nan

gramdal\_v 0.4889 nan nan nan nan nan

gramwholep\_v 0.0083 nan nan nan nan nan

gramGT\_v 0.2305 nan nan nan nan nan

moong\_v 0.2366 nan nan nan nan nan

masur\_v 0.3563 nan nan nan nan nan

urd\_v 0.3541 nan nan nan nan nan

peasdal\_v 0.2946 nan nan nan nan nan

khesari\_v 0.1664 nan nan nan nan nan

otpulse\_v 0.4374 nan nan nan nan nan

gramp\_v 0.3579 nan nan nan nan nan

besan\_v 0.2335 nan nan nan nan nan

pulsep\_v 0.5374 nan nan nan nan nan

pulsestot\_v 0.4346 nan nan nan nan nan

pulsestt\_v 3.8604 nan nan nan nan nan

milk\_v 3.8574 nan nan nan nan nan

babyfood\_v 0 nan nan nan nan nan

milkcond\_v 0.6263 nan nan nan nan nan

curd\_v 0.4547 nan nan nan nan nan

ghee\_v 0.4622 nan nan nan nan nan

butter\_v 0.3855 nan nan nan nan nan

icecream\_v 0.4876 nan nan nan nan nan

otmilkp\_v 0.3482 nan nan nan nan nan

Milktotal\_v 0.4688 nan nan nan nan nan

milkprott\_v 0.4880 nan nan nan nan nan

vanas\_v 3.8331 nan nan nan nan nan

musoil\_v 3.7617 nan nan nan nan nan

gnoil\_v 1.5454 nan nan nan nan nan

cocooil\_v 1.2742 nan nan nan nan nan

edioilothr\_v 1.3619 nan nan nan nan nan

edibletotal\_v 1.2629 nan nan nan nan nan

ediblest\_v 1.2832 nan nan nan nan nan

eggsno\_v 0.0310 nan nan nan nan nan

fishprawn\_v 6.7317 nan nan nan nan nan

goatmeat\_v 1.0186 nan nan nan nan nan

beef\_v 1.0242 nan nan nan nan nan

pork\_v 1.0359 nan nan nan nan nan

chicken\_v 0.9168 nan nan nan nan nan

othrbirds\_v 1.0103 nan nan nan nan nan

nonvegtotal\_v 1.0469 nan nan nan nan nan

emftt\_v 0.9806 nan nan nan nan nan

potato\_v 0 nan nan nan nan nan

onion\_v 7.0401 nan nan nan nan nan

tamato\_v 0.3147 nan nan nan nan nan

brinjal\_v 0.2413 nan nan nan nan nan

radish\_v 0.3146 nan nan nan nan nan

carrot\_v 0.4171 nan nan nan nan nan

palak\_v 0.0603 nan nan nan nan nan

chillig\_v 0.3639 nan nan nan nan nan

bhindi\_v 0.2690 nan nan nan nan nan

parwal\_v 0.3088 nan nan nan nan nan

cauli\_v 0.3508 nan nan nan nan nan

cabbage\_v 0.2530 nan nan nan nan nan

pumpkin\_v 0.2463 nan nan nan nan nan

peas\_v 0.2204 nan nan nan nan nan

fbeans\_v 0.2399 nan nan nan nan nan

lemonno\_v 0.2352 nan nan nan nan nan

otveg\_v 0.3530 nan nan nan nan nan

vegtt\_v -0.6129 nan nan nan nan nan

bananano\_v 0.3251 nan nan nan nan nan

jackfruit\_v 3.8775 nan nan nan nan nan

watermel\_v -0.3006 nan nan nan nan nan

pineaplno\_v -0.2715 nan nan nan nan nan

cocono\_v -0.2419 nan nan nan nan nan

cocogno\_v -0.2371 nan nan nan nan nan

guava\_v -0.3188 nan nan nan nan nan

sighara\_v 0 nan nan nan nan nan

orangeno\_v 0 nan nan nan nan nan

papayar\_v 0 nan nan nan nan nan

mango\_v 0 nan nan nan nan nan

kharbooz\_v -0.2501 nan nan nan nan nan

pears\_v -0.2524 nan nan nan nan nan

berries\_v -0.2732 nan nan nan nan nan

leechi\_v -0.2691 nan nan nan nan nan

apple\_v -0.0638 nan nan nan nan nan

grapes\_v -0.3808 nan nan nan nan nan

otfruits\_v -0.2678 nan nan nan nan nan

fruitstt\_v -0.2758 nan nan nan nan nan

cocodf\_v -0.2552 nan nan nan nan nan

gnutdf\_v 0 nan nan nan nan nan

datesdf\_v 0.1622 nan nan nan nan nan

cashewdf\_v 0.5428 nan nan nan nan nan

walnutdf\_v 0.4652 nan nan nan nan nan

otnutsdf\_v 0.5384 nan nan nan nan nan

kishmish\_v 0.4325 nan nan nan nan nan

otherdf\_v 0.4592 nan nan nan nan nan

dryfruitstotal\_v 0.5889 nan nan nan nan nan

dftt\_v 0.4766 nan nan nan nan nan

sugarpds\_v 0 nan nan nan nan nan

sugaros\_v 3.6625 nan nan nan nan nan

sugarst\_v 0.5879 nan nan nan nan nan

gur\_v 0.5402 nan nan nan nan nan

misri\_v 1.1210 nan nan nan nan nan

honey\_v 1.8071 nan nan nan nan nan

sugartotal\_v 3.3918 nan nan nan nan nan

sugartt\_v 0 nan nan nan nan nan

salt\_v -0.0852 nan nan nan nan nan

ginger\_v 6.3356 nan nan nan nan nan

garlic\_v 8.5177 nan nan nan nan nan

jeera\_v -106.5539 nan nan nan nan nan

dhania\_v -91.6026 nan nan nan nan nan

turnmeric\_v 102.4553 nan nan nan nan nan

blackpepper\_v 71.9594 nan nan nan nan nan

drychilly\_v -91.8910 nan nan nan nan nan

tamarind\_v -152.0086 nan nan nan nan nan

currypowder\_v 32.0117 nan nan nan nan nan

oilseeds\_v -26.4015 nan nan nan nan nan

spicesothr\_v 0 nan nan nan nan nan

spicetot\_v 248.0476 nan nan nan nan nan

spicestotal\_v -54.6078 nan nan nan nan nan

teacupno\_v 118.3709 nan nan nan nan nan

tealeaf\_v -68.5914 nan nan nan nan nan

teatotal\_v 0 nan nan nan nan nan

cofeeno\_v 0 nan nan nan nan nan

coffeepwdr\_v 0 nan nan nan nan nan

cofeetotal\_v 0 nan nan nan nan nan

ice\_v 0 nan nan nan nan nan

coldbvrg\_v 0 nan nan nan nan nan

juice\_v 0 nan nan nan nan nan

othrbevrg\_v 0 nan nan nan nan nan

bevergest\_v 0 nan nan nan nan nan

Biscuits\_v 0 nan nan nan nan nan

preparedsweet\_v 0 nan nan nan nan nan

pickle\_v -4.8483 nan nan nan nan nan

sauce\_jam\_v -4.8507 nan nan nan nan nan

Othrprocessed\_v 34.8562 nan nan nan nan nan

Beveragestotal\_v -7.3834 nan nan nan nan nan

foodtotal\_v -4.8408 nan nan nan nan nan

foodtotal\_q 12.9193 nan nan nan nan nan

fruits\_df\_tt\_v -8.0703 nan nan nan nan nan

fv\_tot -3.2869 nan nan nan nan nan

grp\_4.11E+31 -0.0179 nan nan nan nan nan

grp\_4.12E+31 3.9165 nan nan nan nan nan

grp\_4.13E+31 -0.8993 nan nan nan nan nan

grp\_4.14E+31 -2.5980 nan nan nan nan nan

grp\_4.15E+31 6.1906 nan nan nan nan nan

grp\_4.16E+31 -2.9867 nan nan nan nan nan

grp\_4.17E+31 -1.9506 nan nan nan nan nan

grp\_4.18E+31 -4.8739 nan nan nan nan nan

grp\_4.19E+31 -5.4329 nan nan nan nan nan

grp\_4.20E+31 -4.1137 nan nan nan nan nan

grp\_4.21E+31 -7.3359 nan nan nan nan nan

grp\_4.22E+31 -9.8321 nan nan nan nan nan

grp\_4.23E+31 -8.7714 nan nan nan nan nan

grp\_4.24E+31 -8.1217 nan nan nan nan nan

grp\_4.25E+31 -10.4328 nan nan nan nan nan

grp\_4.26E+31 -11.4893 nan nan nan nan nan

grp\_4.27E+31 -12.2903 nan nan nan nan nan

grp\_4.28E+31 -14.4970 nan nan nan nan nan

grp\_4.29E+31 -13.6638 nan nan nan nan nan

grp\_4.30E+31 -25.0289 nan nan nan nan nan

grp\_4.40E+31 -15.8440 nan nan nan nan nan

grp\_4.41E+31 -21.1499 nan nan nan nan nan

grp\_4.42E+31 -4.0318 nan nan nan nan nan

grp\_4.43E+31 -4.4467 nan nan nan nan nan

grp\_4.44E+31 -4.9960 nan nan nan nan nan

grp\_4.45E+31 -6.5924 nan nan nan nan nan

grp\_4.46E+31 -5.8134 nan nan nan nan nan

grp\_4.47E+31 -6.2986 nan nan nan nan nan

grp\_4.48E+31 -6.5842 nan nan nan nan nan

grp\_4.49E+31 -6.4389 nan nan nan nan nan

grp\_4.50E+31 -9.2465 nan nan nan nan nan

grp\_4.54E+31 -9.1004 nan nan nan nan nan

grp\_4.55E+31 -6.7706 nan nan nan nan nan

grp\_4.57E+31 -4.2478 nan nan nan nan nan

grp\_4.58E+31 -8.3443 nan nan nan nan nan

grp\_4.59E+31 -2.0452 nan nan nan nan nan

grp\_4.60E+31 -2.7163 nan nan nan nan nan

grp\_4.61E+31 -1.8770 nan nan nan nan nan

grp\_4.62E+31 -2.6043 nan nan nan nan nan

grp\_4.63E+31 -3.0252 nan nan nan nan nan

grp\_4.64E+31 -2.2814 nan nan nan nan nan

grp\_4.65E+31 -4.2139 nan nan nan nan nan

grp\_4.66E+31 -2.3912 nan nan nan nan nan

grp\_4.67E+31 -4.8217 nan nan nan nan nan

grp\_4.68E+31 -6.1872 nan nan nan nan nan

grp\_4.69E+31 -7.5981 nan nan nan nan nan

grp\_4.70E+31 -8.4758 nan nan nan nan nan

grp\_4.71E+31 -9.3465 nan nan nan nan nan

grp\_4.72E+31 -9.9863 nan nan nan nan nan

grp\_4.73E+31 -10.4023 nan nan nan nan nan

grp\_4.74E+31 -11.6839 nan nan nan nan nan

grp\_4.75E+31 -13.4093 nan nan nan nan nan

grp\_4.76E+31 -14.6305 nan nan nan nan nan

grp\_4.77E+31 -14.5755 nan nan nan nan nan

grp\_4.78E+31 -19.0858 nan nan nan nan nan

grp\_4.79E+31 -17.9670 nan nan nan nan nan

grp\_4.80E+31 -16.1839 nan nan nan nan nan

grp\_4.90E+31 -16.7000 nan nan nan nan nan

grp\_4.91E+31 -17.8727 nan nan nan nan nan

grp\_4.92E+31 -6.3847 nan nan nan nan nan

grp\_4.93E+31 -6.9627 nan nan nan nan nan

grp\_4.94E+31 -5.3081 nan nan nan nan nan

grp\_4.95E+31 -6.9853 nan nan nan nan nan

grp\_4.96E+31 -8.2970 nan nan nan nan nan

grp\_4.97E+31 -8.7362 nan nan nan nan nan

grp\_4.98E+31 -8.9984 nan nan nan nan nan

grp\_4.99E+31 -9.5550 nan nan nan nan nan

grp\_454046801012351010104132 1101 -10.6480 nan nan nan nan nan

grp\_454046801012351010104132 1102 -10.4892 nan nan nan nan nan

grp\_454046801012351010104132 1103 1.662e-05 nan nan nan nan nan

grp\_454046801012351010104132 1104 -1.0846 nan nan nan nan nan

grp\_454046801012351010104132 1201 1.777e-05 nan nan nan nan nan

grp\_454046801012351010104132 1202 1.757e-05 nan nan nan nan nan

grp\_454046801012351010104132 1203 1.699e-05 nan nan nan nan nan

grp\_454046801012351010104132 1204 1.842e-05 nan nan nan nan nan

grp\_5.00E+31 1.763e-05 nan nan nan nan nan

grp\_7.01E+31 1.797e-05 nan nan nan nan nan

grp\_7.02E+31 -11.0627 nan nan nan nan nan

grp\_7.03E+31 54.1006 nan nan nan nan nan

grp\_7.04E+31 47.6357 nan nan nan nan nan

grp\_7.05E+31 48.2978 nan nan nan nan nan

grp\_7.06E+31 50.7813 nan nan nan nan nan

grp\_7.07E+31 50.4665 nan nan nan nan nan

grp\_7.08E+31 -0.0174 nan nan nan nan nan

grp\_7.09E+31 51.8970 nan nan nan nan nan

grp\_7.10E+31 50.1358 nan nan nan nan nan

grp\_7.11E+31 52.5073 nan nan nan nan nan

grp\_7.12E+31 48.7687 nan nan nan nan nan

grp\_7.13E+31 45.2797 nan nan nan nan nan

grp\_7.14E+31 49.0746 nan nan nan nan nan

grp\_7.15E+31 51.4362 nan nan nan nan nan

grp\_7.16E+31 -117.2886 nan nan nan nan nan

grp\_7.17E+31 51.0333 nan nan nan nan nan

grp\_7.18E+31 43.7719 nan nan nan nan nan

grp\_7.19E+31 48.3168 nan nan nan nan nan

grp\_7.20E+31 43.9007 nan nan nan nan nan

grp\_7.21E+31 43.7264 nan nan nan nan nan

grp\_7.22E+31 35.6884 nan nan nan nan nan

grp\_7.23E+31 43.9213 nan nan nan nan nan

grp\_7.24E+31 46.5421 nan nan nan nan nan

grp\_7.25E+31 42.1675 nan nan nan nan nan

grp\_7.26E+31 43.7011 nan nan nan nan nan

grp\_7.27E+31 39.1941 nan nan nan nan nan

grp\_7.28E+31 36.9845 nan nan nan nan nan

grp\_7.29E+31 34.3276 nan nan nan nan nan

grp\_7.30E+31 40.0134 nan nan nan nan nan

grp\_7.31E+31 37.9930 nan nan nan nan nan

grp\_7.32E+31 34.9833 nan nan nan nan nan

grp\_7.33E+31 34.3992 nan nan nan nan nan

grp\_7.34E+31 33.4014 nan nan nan nan nan

grp\_7.35E+31 32.4962 nan nan nan nan nan

grp\_7.36E+31 31.2890 nan nan nan nan nan

grp\_7.37E+31 31.1082 nan nan nan nan nan

grp\_7.38E+31 29.5741 nan nan nan nan nan

grp\_7.39E+31 29.6255 nan nan nan nan nan

grp\_7.40E+31 29.2374 nan nan nan nan nan

grp\_7.41E+31 29.4935 nan nan nan nan nan

grp\_7.42E+31 27.3899 nan nan nan nan nan

grp\_7.43E+31 25.1116 nan nan nan nan nan

grp\_7.44E+31 25.2379 nan nan nan nan nan

grp\_7.45E+31 24.8107 nan nan nan nan nan

grp\_7.46E+31 24.3125 nan nan nan nan nan

grp\_7.47E+31 22.2736 nan nan nan nan nan

grp\_7.48E+31 20.7529 nan nan nan nan nan

grp\_7.49E+31 21.1815 nan nan nan nan nan

grp\_7.50E+31 21.2806 nan nan nan nan nan

grp\_7.51E+31 20.4498 nan nan nan nan nan

grp\_7.52E+31 19.5163 nan nan nan nan nan

grp\_7.53E+31 18.4402 nan nan nan nan nan

grp\_7.54E+31 19.6799 nan nan nan nan nan

grp\_7.55E+31 18.1820 nan nan nan nan nan

grp\_7.56E+31 11.7250 nan nan nan nan nan

grp\_7.57E+31 13.3601 nan nan nan nan nan

grp\_7.58E+31 13.8314 nan nan nan nan nan

grp\_7.59E+31 12.9469 nan nan nan nan nan

grp\_7.60E+31 19.3238 nan nan nan nan nan

grp\_7.61E+31 14.0106 nan nan nan nan nan

grp\_7.62E+31 13.1391 nan nan nan nan nan

grp\_7.63E+31 12.2390 nan nan nan nan nan

grp\_7.64E+31 -3.3329 nan nan nan nan nan

grp\_7.65E+31 608.4429 nan nan nan nan nan

grp\_7.66E+31 8.5853 nan nan nan nan nan

grp\_7.67E+31 6.3255 nan nan nan nan nan

grp\_7.68E+31 10.4392 nan nan nan nan nan

grp\_7.69E+31 6.9974 nan nan nan nan nan

grp\_7.70E+31 5.9721 nan nan nan nan nan

grp\_7.71E+31 2.7535 nan nan nan nan nan

grp\_7.72E+31 -82.4929 nan nan nan nan nan

grp\_7.73E+31 0.5365 nan nan nan nan nan

grp\_7.74E+31 3.4332 nan nan nan nan nan

grp\_7.75E+31 -4.8469 nan nan nan nan nan

grp\_7.76E+31 1.1924 nan nan nan nan nan

grp\_7.77E+31 -1.1267 nan nan nan nan nan

grp\_7.78E+31 2.5939 nan nan nan nan nan

grp\_7.79E+31 -0.1871 nan nan nan nan nan

grp\_7.80E+31 -2.0330 nan nan nan nan nan

grp\_7.90E+31 -5.5952 nan nan nan nan nan

grp\_7.91E+31 1.9554 nan nan nan nan nan

grp\_7.92E+31 5.3809 nan nan nan nan nan

grp\_7.93E+31 8.9269 nan nan nan nan nan

grp\_7.94E+31 6.9087 nan nan nan nan nan

grp\_7.95E+31 6.8219 nan nan nan nan nan

grp\_7.96E+31 4.6849 nan nan nan nan nan

grp\_7.97E+31 3.8517 nan nan nan nan nan

grp\_7.98E+31 3.4267 nan nan nan nan nan

grp\_7.99E+31 0.4016 nan nan nan nan nan

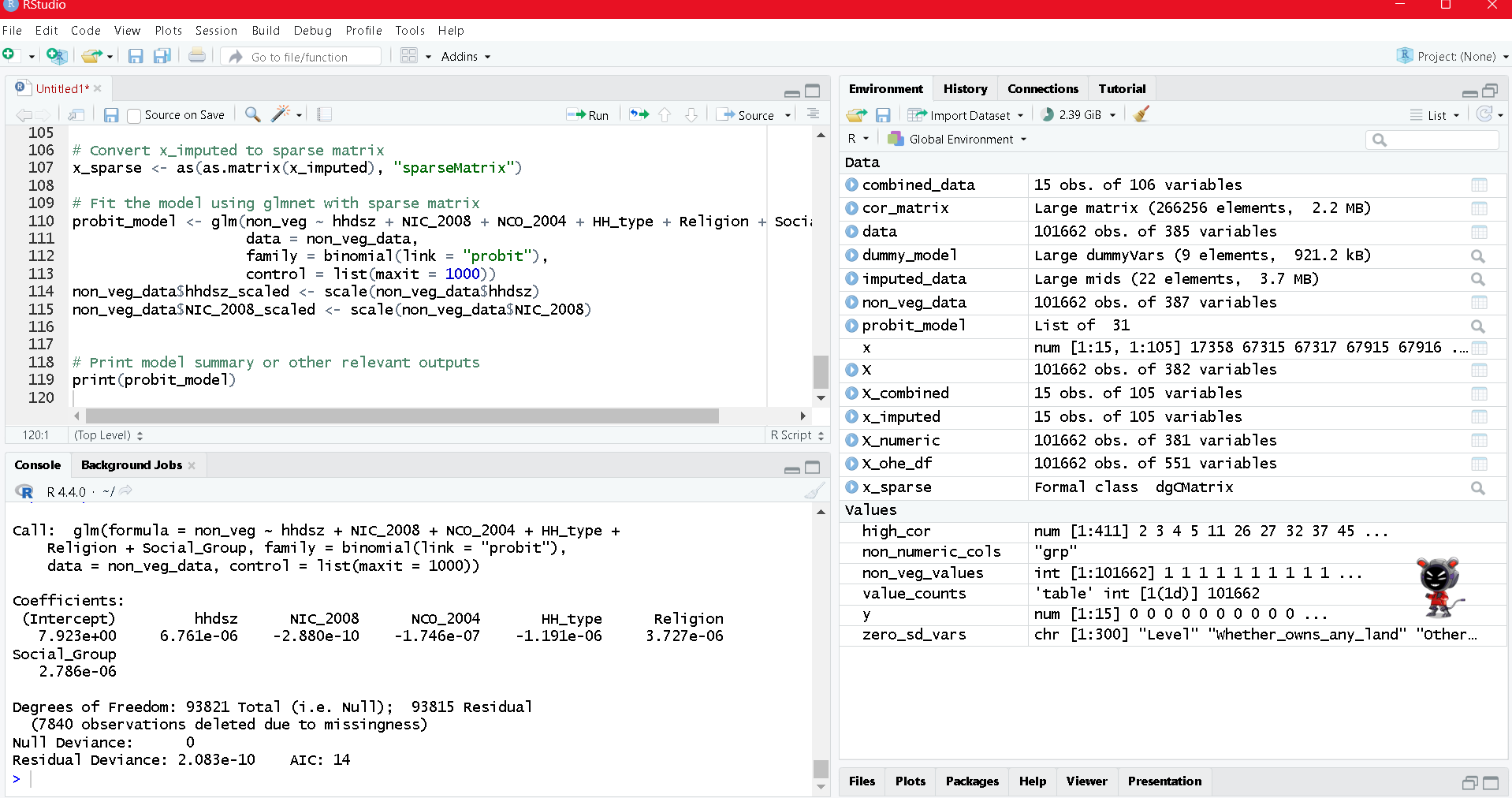
grp\_8.00E+31 2.8409 nan nan nan nan nan

========================================================================================================

Possibly complete quasi-separation: A fraction 0.79 of observations can be

perfectly predicted. This might indicate that there is complete

quasi-separation. In this case some parameters will not be identified.

****

**Interpretation:**

The Probit Regression Results table shows the results of fitting a probit model to the data. The table includes the coefficient estimates, standard errors, z-statistics, and p-values for each predictor variable in the model. The coefficient estimates represent the change in the z-score of the latent variable (i.e., the unobserved variable that determines the outcome variable) for a one-unit increase in the predictor variable, holding all other predictor variables constant. The standard errors represent the estimated variability of the coefficient estimates, and the z-statistics represent the ratio of the coefficient estimate to its standard error. The p-values indicate the probability of observing a z-statistic as extreme as the one calculated, assuming the null hypothesis is true (i.e., the coefficient is equal to zero).

**The advantages of the probit model include:**

* It can handle binary outcome variables, which are common in many research questions.
* It can model non-linear relationships between the predictor variables and the outcome variable.
* It can handle continuous and categorical predictor variables.
* It can account for the correlation between predictor variables.
* It can provide estimates of the marginal effects of the predictor variables on the probability of the outcome variable being 1.

However, it is important to note that the Probit Regression Results table indicates that there may be complete quasi-separation in the data, which can lead to unstable coefficient estimates and inaccurate standard errors. This occurs when a fraction of the observations can be perfectly predicted by the predictor variables, which can happen when the predictor variables are highly correlated or when there are few observations in one of the outcome variable categories. To address this issue, you may need to consider alternative modeling approaches, such as adding more predictor variables, removing highly correlated predictor variables, or using a different link function in the GLM.

It is also important to interpret the results of the probit model with caution, given the potential for bias and inaccuracy due to quasi-separation.

The results of the probit regression analysis indicate that factors such as age, education, income, religion, social group, and region are significant predictors of non-vegetarian food consumption in India. The coefficient estimates suggest that households with higher income and education levels are more likely to consume non-vegetarian food. Similarly, households belonging to certain religions and social groups are more likely to consume non-vegetarian food. The results also indicate that households in certain regions of India are more likely to consume non-vegetarian food**.**

**Recommendation:**

Based on the results of the analysis, it is recommended that businesses in the food and beverage industry focus on developing and marketing non-vegetarian food products in regions where the consumption of non-vegetarian food is high. Additionally, businesses can target households with higher income and education levels, as these households are more likely to consume non-vegetarian food. It is also recommended that businesses consider the religious and social factors influencing non-vegetarian food consumption in India and develop marketing strategies that cater to these factors.

**Codes:**

**import** pandas **as** pd

**import** statsmodels.api **as** sm

**import** numpy **as** np

**from** sklearn.preprocessing **import** OneHotEncoder

**from** sklearn.metrics **import** accuracy\_score, classification\_report, confusion\_matrix

In [16]:

data **=** pd**.**read\_csv("C:\\Users\\Bala Vignesh.A\\Desktop\\SCMA 632\\NSSO68.csv", low\_memory**=False**)

In [31]:

display (data)

**Filter the data for non-vegetarians**

In [33]:

non\_veg\_values **=** non\_veg\_data['non\_veg']

print(non\_veg\_values)

101661 1

Name: non\_veg, Length: 101047, dtype: int64

In [34]:

print(non\_veg\_values**.**unique())

[1]

In [35]:

print(non\_veg\_values**.**value\_counts())

non\_veg

1 101047

Name: count, dtype: int64

In [40]:

print(data**.**columns)

**Define the dependent variable (non\_veg) and independent variables**

In [74]:

y **=** non\_veg\_data['non\_veg']

X **=** non\_veg\_data**.**drop(['non\_veg', 'state\_1', 'Region'], axis**=**1)

**Check for non-numeric columns**

In [75]:

non\_numeric\_cols **=** X**.**select\_dtypes(include**=**[object])**.**columns

print("Non-numeric columns:", non\_numeric\_cols)

Non-numeric columns: Index(['grp'], dtype='object')

**One-Hot Encode non-numeric columns using pd.get\_dummies**

In [77]:

X\_ohe **=** pd**.**get\_dummies(X[non\_numeric\_cols], drop\_first**=True**)

X\_ohe **=** pd**.**get\_dummies(X[non\_numeric\_cols], drop\_first**=True**, prefix**=**non\_numeric\_cols)

**Combine numeric and one-hot encoded columns**

In [78]:

X\_numeric **=** X**.**select\_dtypes(include**=**[np**.**number])

X\_combined **=** pd**.**concat([X\_numeric**.**reset\_index(drop**=True**), X\_ohe**.**reset\_index(drop**=True**)], axis**=**1)

**Handle missing values using median imputation**

In [83]:

**from** sklearn.impute **import** SimpleImputer

In [89]:

imputer **=** SimpleImputer(strategy**=**"constant", fill\_value**=**0)

X\_imputed **=** imputer**.**fit\_transform(X\_combined)

In [94]:

column\_names **=** X\_combined**.**loc[:, X\_combined**.**count() **>** 0]**.**columns

X\_imputed **=** X\_imputed[:, :len(column\_names)] *# Select the correct columns*

X\_combined\_imputed **=** pd**.**DataFrame(X\_imputed, columns**=**column\_names)

**Check for empty columns**

In [95]:

empty\_cols **=** X\_combined\_imputed**.**columns[X\_combined\_imputed**.**count() **==** 0]

**if** len(empty\_cols) **>** 0:

print("Empty columns:", empty\_cols)

X\_combined\_imputed**.**drop(columns**=**empty\_cols, inplace**=True**)

**Verify the shape of the data**

In [96]:

print("Shape of X\_combined\_imputed:", X\_combined\_imputed**.**shape)

print("Shape of y:", y**.**shape)

Shape of X\_combined\_imputed: (101047, 546)

Shape of y: (101047,)

**Ensure that X\_combined\_imputed and y have compatible shapes**

In [97]:

y **=** y**.**reset\_index(drop**=True**)

**Fit the probit model**

In [109]:

**import** statsmodels.api **as** sm

**try**:

probit\_model **=** sm**.**Probit(y, X\_combined\_imputed)

probit\_results **=** probit\_model**.**fit()

print(probit\_results**.**summary())

**except** Exception **as** e:

print(f"An error occurred: {e}")

**R**

# Load the necessary libraries

library(tidyverse)

library(mice)

library(car)

library(ggplot2)

library(lattice)

library(caret)

library(glmnet)

library(Matrix)

# Read in the data

data <- read.csv("C:\\Users\\Bala Vignesh.A\\Desktop\\SCMA 632\\NSSO68.csv")

# Filter the data for non-vegetarians

non\_veg\_data <- data[data$non\_veg == 1, ]

# View the non-veg values

non\_veg\_values <- non\_veg\_data$non\_veg

print(non\_veg\_values)

# Get the value counts of non\_veg

non\_veg\_values <- non\_veg\_data$non\_veg

value\_counts <- table(non\_veg\_values)

print(value\_counts)

# Define the dependent variable (non\_veg) and independent variables

y <- non\_veg\_data$non\_veg

X <- non\_veg\_data[,!(names(non\_veg\_data) %in% c("non\_veg", "state\_1", "Region"))]

# Check for non-numeric columns

non\_numeric\_cols <- sapply(X, class) == "character"

non\_numeric\_cols <- names(non\_numeric\_cols)[non\_numeric\_cols]

print(paste("Non-numeric columns:", non\_numeric\_cols))

# One-hot encode categorical columns

dummy\_model <- dummyVars(~., data = X)

X\_ohe\_df <- as.data.frame(predict(dummy\_model, newdata = X))

# Combine numeric and one-hot encoded columns

X\_numeric <- X[, sapply(X, class)!= "character"]

X\_combined <- cbind(X\_numeric, X\_ohe\_df)

# Ensure 'y' is a binary factor

y <- as.factor(y)

# Remove columns with single unique value

X\_combined <- X\_combined[, sapply(X\_combined, function(x) length(unique(x)) > 1)]

# Check dimensions

print(dim(y))

print(dim(X\_combined))

# Create the combined data frame

combined\_data <- data.frame(y, X\_combined)

# Inspect the combined data

str(combined\_data)

head(combined\_data)

# Check for missing values in X\_combined and y

sum(is.na(X\_combined))

sum(is.na(y))

# Impute missing values if necessary

X\_combined <- na.omit(X\_combined)

y <- y[!is.na(X\_combined)]

# Check variables with zero standard deviation

zero\_sd\_vars <- colnames(X\_combined)[apply(X\_combined, 2, sd) == 0]

print(zero\_sd\_vars)

# Remove constant variables from X\_combined

X\_combined <- X\_combined[,!colnames(X\_combined) %in% zero\_sd\_vars]

# Compute correlation matrix

cor\_matrix <- cor(X\_combined, use = "pairwise.complete.obs")

# Find highly correlated predictors

high\_cor <- findCorrelation(cor\_matrix, cutoff = 0.9)

# Remove highly correlated predictors from X\_combined

X\_combined <- X\_combined[, -high\_cor]

dim(X\_combined)

length(y)

y <- y[1:nrow(X\_combined)]

combined\_data <- data.frame(y = y, X\_combined)

dim(combined\_data)

# Create a matrix of predictor variables

x <- as.matrix(X\_combined)

# Create a response variable

y <- as.numeric(as.character(y)) - 1

# Impute missing values using makeX

x\_imputed <- makeX(as.data.frame(x), missing = TRUE)

imputed\_data <- mice(X\_combined)

x\_imputed <- complete(imputed\_data)

anyNA(x\_imputed)

anyNA(y)

# Convert x\_imputed to sparse matrix

x\_sparse <- as(as.matrix(x\_imputed), "sparseMatrix")

# Fit the model using glmnet with sparse matrix

probit\_model <- glm(non\_veg ~ hhdsz + NIC\_2008 + NCO\_2004 + HH\_type + Religion + Social\_Group,

data = non\_veg\_data,

family = binomial(link = "probit"),

control = list(maxit = 1000))

non\_veg\_data$hhdsz\_scaled <- scale(non\_veg\_data$hhdsz)

non\_veg\_data$NIC\_2008\_scaled <- scale(non\_veg\_data$NIC\_2008)

# Print model summary or other relevant outputs

print(probit\_model)