# Finding 1 :

From below variables:

running on all these variables

c("State","County","Urban","GroupQuartersFlag","LowIncomeTracts","lahunv1share","PCTGQTRS","MedianFamilyIncome","lawhite1","lablack1","laasian1","lahisp1","lanhopi1","laomultir1","laaian1","lakids10","lakids1","TractKids","laseniors1","laseniors10","TractKids","TractSeniors","TractWhite","TractBlack","TractAsian","TractNHOPI","TractAIAN","TractOMultir","TractHispanic","TractHUNV","TractSNAP", "PovertyRate")

AUC 0.961

Confusion matrix from Logit Model

FALSE        TRUE

0        58806        2149

1        284        6304

Good prediction

even though the logistic regression said all variables are significant

the odds ratio gave 0s and inf indicating the small variables are useless and some variables gives perfect separation.

this perfect separation may because the lila is derived from lowincome

multicolinearity problem

# Finding 2 :

From Chi sq test we could say GroupQuartersFlag and LILATracts\_1And10 is dependent.

model <- glm(LILATracts\_1And10 ~ PovertyRate + GroupQuartersFlag, family = binomial(link = "logit"), data = df)

AIC: 50201

VIF of these 3 variables :

PovertyRate GroupQuartersFlag Urban   
 1.049878 1.027239 1.022395

Povertyrate : Urban :

\* When Urban is 0 (Rural):

An increase in PovertyRate by one unit is associated with a larger increase in the log-odds of the response variable being 1.

\* When Urban is 1 (Urban):

An increase in PovertyRate by one unit is associated with a smaller increase in the log-odds of the response variable being 1.

Povertyrate : GroupQuartersFlag :

\* In areas Non group quarters (GroupQuartersFlag is 0),

an increase in PovertyRate is associated with a relatively larger increase in the log-odds of the response variable being 1.

\* In areas with group quarters (GroupQuartersFlag is 1),

the positive impact of PovertyRate on the log-odds is moderated, resulting in a smaller increase in the log-odds.

Findings 3 :

Model 1

model\_imp <- glm(LILATracts\_1And10 ~ GroupQuartersFlag + PovertyRate + HUNVFlag + MedianFamilyIncome + LATracts1 + LATractsVehicle\_20, family = "binomial", data = df\_subset)

AIC: 25482

Accuracy : 0.9534612

AUC : 0.9365

Model 2: mine

model\_glm2 <- glm(LILATracts\_1And10 ~ OHU2010 + LA1and20 + LATracts10 + LATractsVehicle\_20 + GroupQuartersFlag + PovertyRate + HUNVFlag + MedianFamilyIncome + lalowi1share + lakids1share + lahisp1share + laseniors10share + lawhite1share + lablack1share + laaian1share + laomultir10share, family = "binomial", data = df\_subset)

AIC: 10887

Accuracy : 0.9696421

AUC : 0.9922

Forward selection his: Model 4

glm(formula = LILATracts\_1And10 ~ LowIncomeTracts + lahunv1share + Urban + lakids10 + lawhite1 + laseniors10 + lablack1 + laomultir1 + laasian1 + TractAsian + laaian1 + TractHUNV + TractWhite + lakids1 + TractOMultir + GroupQuartersFlag + PovertyRate + TractSNAP, family = binomial(link = "logit"), data = cleaned\_data)

AIC: 32035

Accuracy : 0.914114

AUC : 0.9204

# Findings 4 :

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | White | black | nhopi | asian | aian | omultir | hisp |
|  | 10952  96.92  0.99 | 11015  96.85 | 11039  96.86  0.992 | 11004  96.89  0.992 | Not  significant | 10956  96.89 | 11024  96.8 |
|  | 22,35,53,265 | 3,89,29,319 | 5,40,013 | 1,46,74,252 | 29,32,248 | 2,81,16,441 | 5,04,77,594 |

Significantly there is no much difference in model performance wise according to demographic groups.

But when compared, White group give better performance. May be due to more in count compared to other groups.( having more count of non food deserted regions in dataset )