# The Demographic and Socioeconomic Correlates of Living Arrangements of Elderly Individuals

## INTRODUCTION:

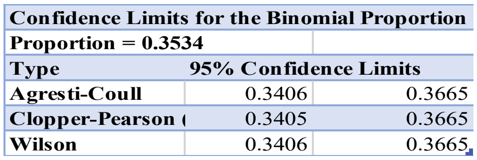
Living arrangement is viewed as a best indicator to understand the status and the wellbeing of the elderly in the society. The concept of the living arrangement is usually explained in terms of the type of family in which the elderly live, the headship they enjoy, the place they stay in and the people they stay with, the kind of relationship they maintain with their kith and kin, and on the whole, the extent to which they adjust to the changing environment. The paper addresses the demographic and socioeconomic correlates of living arrangements of elderly individuals in USA. Data for the reference group of this study came from the Second Longitudinal Study of Aging, Wave 3, conducted in 2000 (LSOA II). The LSOA II study design and sample selection procedures are well documented and is a collaboration effort of the National Center for Health Statistics (NCHS) and the National Institute on Aging (NIA). The LSOA II is a longitudinal study with a national representative sample consisting of 9447 civilian, non-institutionalized persons 70 years of age at the time of the LSOA II baseline interview. For the second follow-up interview (Wave 2), 8905 subjects were traced and located, and 7998 interviews were obtained; and the third follow-up interviews (Wave 3), 7936 were successfully traced and located and 6465 were interviewed. After deleting missing observations from our variables of interest, we finally came up with 5246 observations.

## EXPLORATORY ANALYSIS:

In the multivariate analyses, we first put a focus on a limited number of variables at the baseline that are likely to have already been determined during the time of the survey. To know the determinants of the living arrangement many studies have been conducted which indicate that living arrangements are influenced by a number of socio demographic variables. The selected socio-economic and demographic characteristics considered in the study include educational attainment, poverty index, disability, marital status, type of residence, frequency of driving (availability of public transport), get together with friends/relatives, employment status (information about labor participation of the individuals), unpaid volunteer work, felt sad/depressed, health status, interval since last doctor visit, family income less than $20,000, years at current residence, and region of the country.

At the multivariate level, the multinomial logistic regression models are employed to predict the characteristics of the individual in different living arrangements. In this attempt, the various types of living arrangements will be collapsed into two, namely, (i) living alone (Y=1); (ii) living with others (Y=0). In our sample 35.34% individuals are living alone. Since each of these two categories can be defined as dichotomous variables, the logistic regression methodology is suitable for providing these estimates.

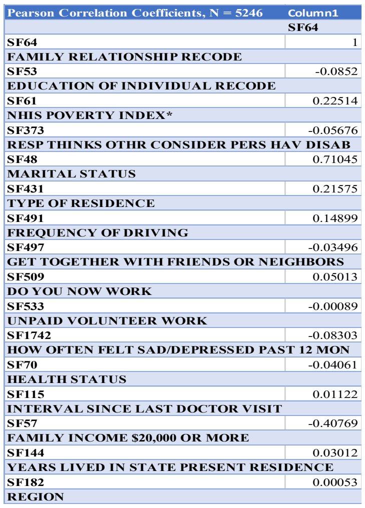
At the beginning of our explanatory analysis, we first check the confidence interval of our response variable at 95% confidence level. There are many methods that have proposed for the confidence limits for a binomial proportion. Among those methods, Agresti-Coull confidence limit and Clopper–Pearson interval are commonly used in the literature. From the Table 1, we see that the Agresti-Coull confidence limit of our response variable ranges from 0.3406 to 0.3665 at 95% confidence interval. The other two methods also show similar results.



| Confidence Limits for the Binomial Proportion | | |
|---|---|---|
| Proportion = 0.3534 | | |
| Type | 95% Confidence Limits | |
| Agresti-Coull | 0.3406 | 0.3665 |
| Clopper-Pearson ( | 0.3405 | 0.3665 |
| Wilson | 0.3406 | 0.3665 |

Table 1

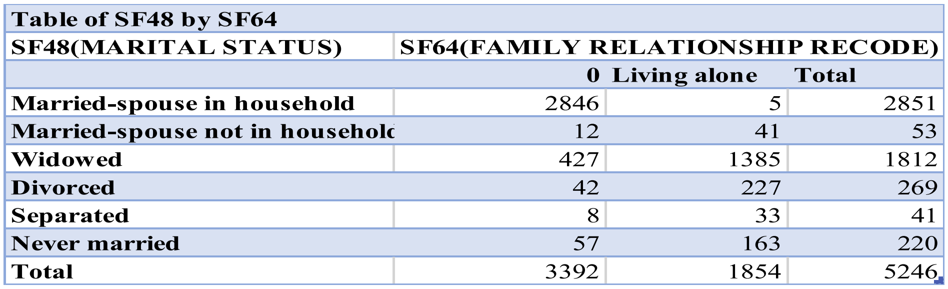
Table 2 shows the correlation matrix of our explanatory variables. From the table it is revealed that our response variable "living alone" is strongly positively correlated with 'NHIS Poverty Index', 'Marital Status', and 'Type of Residence'. On the other hand, ‘Health Status' and 'Family Income less than $20,000' are negatively correlated with our response variable.



| Pearson Correlation Coefficients, N = 5246 | Column1 |
|---|---|
| | SF64 |
| SF64 | 1 |
| FAMILY RELATIONSHIP RECODE | |
| SF53 | -0.0852 |
| EDUCATION OF INDIVIDUAL RECODE | |
| SF61 | 0.22514 |
| NHIS POVERTY INDEX* | |
| SF373 | -0.05676 |
| RESP THINKS OTHR CONSIDER PERS HAV DISAB | |
| SF48 | 0.71045 |
| MARITAL STATUS | |
| SF431 | 0.21575 |
| TYPE OF RESIDENCE | |
| SF491 | 0.14899 |
| FREQUENCY OF DRIVING | |
| SF497 | -0.03496 |
| GET TOGETHER WITH FRIENDS OR NEIGHBORS | |
| SF509 | 0.05013 |
| DO YOU NOW WORK | |
| SF533 | -0.00089 |
| UNPAID VOLUNTEER WORK | |
| SF1742 | -0.08303 |
| HOW OFTEN FELT SAD/DEPRESSED PAST 12 MON | |
| SF70 | -0.04061 |
| HEALTH STATUS | |
| SF115 | 0.01122 |
| INTERVAL SINCE LAST DOCTOR VISIT | |
| SF57 | -0.40769 |
| FAMILY INCOME $20,000 OR MORE | |
| SF144 | 0.03012 |
| YEARS LIVED IN STATE PRESENT RESIDENCE | |
| SF182 | 0.00053 |
| REGION | |

Table 2

In the survey data, marital status is categorized as married with couples, divorced, separated, widowed, never married and unresponsive individuals. From the contingency Table 3, we see that a vast majority of the individual are either widowed (34.54 percent) or married (54.34 percent), while the incidence of divorce is a rare occurrence in the country (5.13 percent). Among the widowed individual, 76.4% of them living alone whereas only 99.82% married people live with spouse.

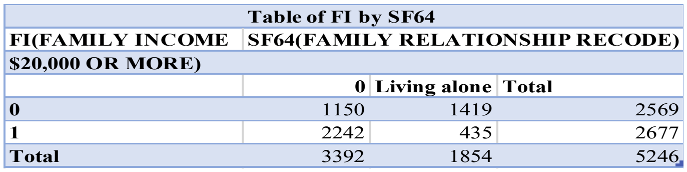


| Table of SF48 by SF64 | | | |
|---|---|---|---|
| SF48(MARITAL STATUS) | SF64(FAMILY RELATIONSHIP RECODE) | | |
| | 0 | Living alone | Total |
| Married-spouse in household | 2846 | 5 | 2851 |
| Married-spouse not in householc | 12 | 41 | 53 |
| Widowed | 427 | 1385 | 1812 |
| Divorced | 42 | 227 | 269 |
| Separated | 8 | 33 | 41 |
| Never married | 57 | 163 | 220 |
| Total | 3392 | 1854 | 5246 |

Table 3

Poverty index is defined as a binary variable where 90.25% individual live at or above the poverty level. The propensity to live with others (68.2%) is higher for people who are above the poverty level. It seems logical that people who live above the poverty level are able to afford larger dwellings, which in corresponds to more living room for multiple people. Intuitively, a large number of respondents living above the poverty level correspond to a smaller value of p-hat.

Family income is considered as one of the major determinants of living arrangement and is also defined as dichotomous variable. Table 4 shows that 48.97% individual has family income less than $20,000. However. 51.03% individual has family income more than $20,000. It is revealed that high income group individuals are less likely to live alone. Among the first group of individuals (family income less than $20,000), 16.2% individual live alone. This statistic is just opposite to some well-known studies from developed countries, such as the study by Costa (1997), have shown that income plays an important role enabling elderly to live alone. Privacy of both parents and children as a normal good has been modeled in studies of living arrangements in developed countries. A study conducted by Da Vanzo and Chan (1994) reveals that income is positively related to the concept of living alone in most of the developed countries. (Note: FI=0 for Income<$20,000)



| Table of FI by SF64 | | | |
|---|---|---|---|
| FI(FAMILY INCOME $20,000 OR MORE) | SF64(FAMILY RELATIONSHIP RECODE) | | |
| | 0 | Living alone | Total |
| 0 | 1150 | 1419 | 2569 |
| 1 | 2242 | 435 | 2677 |
| Total | 3392 | 1854 | 5246 |

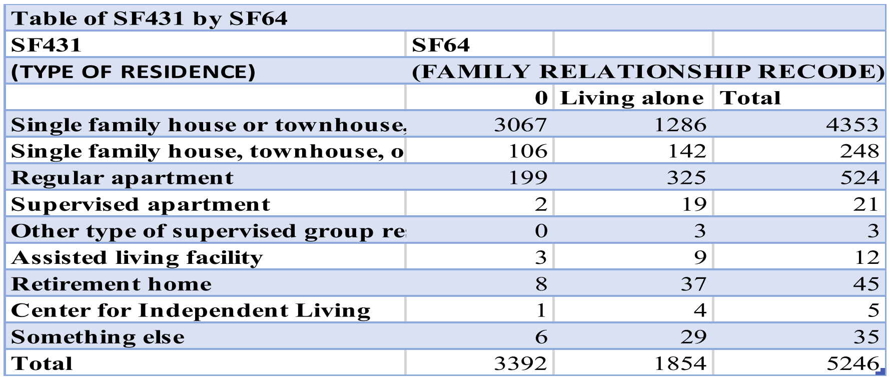
Table 4

Living arrangements are closely related to the health status. Six categories were used in asking about health status (such as excellent, very good, good, fair, poor, unknown) during the survey, 34.32% individual report their health status as "Good" whereas 8.9% individual reports as "Poor". The propensity to living alone is seen among individuals with good health. However, individuals with poor health prefer to live with others. Similar to health status, elderly persons who live alone have higher levels of depressive symptomatology; and this relationship is independent of the influence of expressive support from friends, face-to-face interaction with friends, undesirable life events, disability, and financial strain. About one-thirds (30.95%) of the individuals of our sample suffer from depression at least for some time whereas 34.78% sample individual reports that they do not suffer from any depression or sadness. As expected, people who suffers from depression are less likely to live alone whereas respondents who reports that they never suffer from depression are more likely to live alone (11.19%).

Although educational attainment can provide people with means and resources that can influence their economic and health status, it has the distinct effects of exposing them to new and non-traditional attitudes. According to literature, higher levels of educational attainment have significant negative correlation with the likelihood of kin co-residence. Among the nine categories of education level of individual, 33.37% individuals are high school graduate whereas individual with education of 5 years or more consists of only 4.76%. The second largest group of our sample completed only elementary education. Among the high school graduates, 11.04% live alone. Relatively small fraction of individual with college graduates live alone. This group of individuals prefer to live with others.

Regarding the distribution of individuals living different regions of United States, our contingency table shows that 32.30% individual lives in the southern part of the country whereas the second largest group of individuals live in the Midwest region. West is less densely populated. 18.77% of sample individual live in the western part. Among this group respondents, 11.01% individuals who live in the southern part of the country live alone which is highest among all regions.

We next examined the relationship between Living Arrangement and Type of Residence using a contingency table (Table 5). The relationship between the two variables is not as obvious from this table, but it still exists. People who live in a Supervised Apartment are more likely to live alone. A majority of the individuals (82.97%) live in a single family house or townhouse and among them 29.54% respondents live alone. However, only 9.99% of sampled individuals use regular apartments and 62.02% of them live alone.



| Table of SF431 by SF64 | | | |
|---|---|---|---|
| SF431 | SF64 | | |
| (TYPE OF RESIDENCE) | (FAMILY RELATIONSHIP RECODE) | | |
| | 0 | Living alone | Total |
| Single family house or townhouse, | 3067 | 1286 | 4353 |
| Single family house, townhouse, o | 106 | 142 | 248 |
| Regular apartment | 199 | 325 | 524 |
| Supervised apartment | 2 | 19 | 21 |
| Other type of supervised group re | 0 | 3 | 3 |
| Assisted living facility | 3 | 9 | 12 |
| Retirement home | 8 | 37 | 45 |
| Center for Independent Living | 1 | 4 | 5 |
| Something else | 6 | 29 | 35 |
| Total | 3392 | 1854 | 5246 |

Table 5

In the following sections we will discuss our regression results.

## REGRESSION ANALYSIS:

Since our response variable is a Bernoulli variable, we decided to consider logistic regression. However, in the original data set, Living Arrangement had four possible responses, so multi-nominal regression seemed to be a potential model choice as well. We then proceeded to perform regression analysis on the predicted probability of living alone using methods of binary logit and cumulative multi-nominal regressions.

From the former analysis, we know that among the 15 chosen exploratory variables, Marital status (SF48) and Region (SF182) are categorical data. We set them as class variables with the reference parameter for marital status be 'Windowed' and the reference parameter for region be 'West' and treat other variables as numerical data. After running regressions using both methods in SAS, we get some comparisons, and find that the cumulative multi-nominal regression model has a Pearson $\chi^2$ that is too large relative to the degrees of freedom (54.5958). With such a condition, we finally abandon the cumulative multi-nominal method and choose the binary logit regression as our regression method. In the following part, we will show you how we did such regressions and get our final results in detail.

At first, we prepared two choices to deal with our regression—binary logit and multi-nominal cumulative logit. However, after proceeding our primary regressions in SAS, we found the outcome of multi-nominal cumulative logit did not fit the data well. We then used binary logit regression method in our main regression process. Since binary logit regression is our main work, we are going to introduce it in detail. Here follows the explanations.

After proceeding a primary binary logit regression in SAS, we have get the following outcome:

$$logit(\hat{\pi}(x)) = 2.0343 + 0.2859*SF53 + 0.0631*PI - 0.2718*DIS + 0.2718*SF48_1 - 8.2182*SF48_2 + 0.2833*SF48_3 - 0.1577*SF48_4 -0.00779*SF48_5 +0.3688*SF431 - 0.3425*SF491 - 0.1449*GT + 0.2936*WN + 0.0458*UVW - 0.0706*SF1742 - 0.2628*SF70 - 0.1482*SF115 - 2.404*FI + 0.196*SF144 + 0.0136*SF182_1 - 0.3845*SF182_2 -0.1427*SF182_3 .$$

Where each variable has their meanings: SF53 Education; PI NHIS Poverty Index; DIS Disability; SF48 Marital Status ($SF48_1$ Divorced; $SF48_2$ Married-spouse in household; $SF48_3$ Married-spouse not in household; $SF48_4$ Never married; $SF48_5$ Separated ); SF431 Type of Residence; SF491 Frequency of Driving; GT Get together with friends or relatives; WN Work Now;UVW Unpaid Volunteer Work; SF1742 How Often Felt Sad/Depressed Past 12 Months; SF70 Health Status; SF115 Interval Since Last Doctors Visit; FI Family Income >$20,000; SF144 Years at Current Residence ; SF182 Region ($SF182_1$ Midwest; $SF182_2$ Northeast; $SF182_3$ South ).

The SAS output also displayed the p-values of each estimators. However, according to the p-values of the estimators, we find that not all of the estimators are significant. Thus, we need further explorations to select out the significant exploratory variables.

First, we assume that all 15 exploratory variables are non-significant and proceed a likelihood ratio test to find whether our assumptions are true. The likelihood ratio test with a null hypothesis that all estimators are zero is rejected by a LRT-$\chi^2$ = 4765.74 with d.f.=21. The p-value of this likelihood ratio test is less than 0.0001, which has a high evidence that among the 15 exploratory variables, at least some of them have significant effects. So, the next steps are to find these significant exploratory variables.

With this purpose, we then conduct a backwards selection process by SAS. The backward selection takes 9 steps to select for the needed variables. For the first step, it compared the model with and without variable SF533 (Unpaid Volunteer Work) and get a likelihood ratio test statistic 0.084 with 1 d.f., so the p-value is 0.7719 (>0.05), which means the variable SF533 does not have a significant effect on Living arrangement). So the step1 ended up by removing the variable SF533 (Unpaid Volunteer Work) with a non-significant p-value 0.7722. Then for step 2, it removed the variable PI (NHIS Poverty Index) with a non-significant p-value 0.7341. And for step 3, it removed SF1742 (How often felt sad/depressed past 12 months). Next step, it removed SF497 (Get Together With Friends or Neighbors). Step 5, removed SF509 (Work Now or Not). Step 6, removed SF115 (Interval Since Last Doctor Visit). Step 7, removed SF373 (Disability). Step 8, removed SF144 (Years at Current

Residence). Finally for step 9, it removed variable SF182 (Region). After all of these 9 steps of backward selection, we finally remove 9 non-significant variables and keep the remaining 6 ones as our final exploratory variables. The resulting model after backward selection is as follows:

$$logit(\hat{\pi}(x))=1.0518 + 0.2793*SF53 + 1.6112*SF48_1 -6.8750*SF48_2 + 1.5362*SF48_3 + 1.1051*SF48_4 + 1.3132*SF48_5 + 0.3684*SF431 - 0.3361*SF491 - 0.2043*SF70 - 2.4074*FI.$$

From this result, we can look at an example to interpret some of the predicted effects the variables have on the estimated probability of an individual living alone. We'll look at the predicted effect of family income on living conditions. The model predicts that, holding other variables constant, then the predicted odds of someone living alone will be multiplied by $e^{-2.4074}$ =0.09 as the level of family income increases 1 level.

Now that we have selected the model that we are going to use, we can make an analysis of how well the model fits the data we have chosen. From Table 9, we had 6,288,768 pairs. Of these pairs, 96.7 percent of them were concordant. This means that a large portion of the pairs we observed had a higher probability of the desired effect occurring compared to the non-event. Only 3.3 percent of pairs were discordant or tied according to our results. Also, the value for Somers' D is 0.936, or very close to 1, which implies that most of the pairs agree with each other. This implies that the model is a good fit.

We then calculated the sensitivity and the specificity of the results, using the rule to predict $\hat{Y} = 1$ if $\hat{\pi} > 0.5$ and $\hat{Y} = 0$ otherwise. The sensitivity, or the number of positive events truly identified as positive, was .952. The specificity, or the number of negative events truly identified as negative, was .902. This meant that our predicted outcomes ended up agreeing with what actually happened approximately 92 percent of the time, which once again shows the strength of the model. The residual plots also showed that the model fit the data well, despite the presence of some outliers. In the appendix we have shown the residual plots for SF64 (Family Relationship).

Finally, we can compare the tables entitled "Criteria for Assessing Goodness of Fit" between our fitted model and the model using only the intercept. For the model using only the intercept, the value for AIC was 6816.9037, and BIC was a very similar 6823.4869. The fitted model on the other hand produces an AIC of 2092.4256 and the value for BIC is 2164.643. The AIC and BIC values for our fitted model are much lower than the ones from the model only using the intercept. We also tried to add interaction terms between Marital Status and Family Income as well as between Family

Income and Type of Residence. As a result, the new AIC value was 2098.5278 and the new BIC value was 2210.1365. Because these values are slightly higher than the original fitted model without interaction that we produced, we decided to stick with that model since it seemed to provide a slightly better fit.

CONCLUSION:

After going through our regression analysis using backwards selection, we were able to come up with the following fitted model:

$logit(\hat{\pi}(x))=1.0518 + 0.2793*SF53 + 1.6112*SF48_1 -6.8750*SF48_2 + 1.5362*SF48_3 + 1.1051*SF48_4 + 1.3132*SF48_5 + 0.3684*SF431 - 0.3361*SF491 - 0.2043*SF70 - 2.4074*FI.$

It is worth noting that above, SF53 is education, SF48$_1$ is Divorced, SF48$_2$ is Married-spouse in household,SF48$_3$ is Married-spouse not in household, SF48$_4$ is Never married SF48$_5$ is Separated, SF431 is Type of Residence, SF491 is Frequency of Driving, SF70 is Health Status, and FI is Family income >$20,000, where

$$FI = \begin{cases} 0, \ income < \$20{,}000 \\ 1, \ income > \$20{,}000 \end{cases}$$

These are all of the explanatory variables that we found to be significant enough to include in the final model.

The amount of variables present in the model is reduced significantly from the original 15 we tested. The variables that were kept in the model all seem to make sense when testing for living arrangement. For example, one's health status would probably greatly affect whether or not they would feel comfortable living alone or if they feel they would benefit from the company of others. It was rather surprising to see variables like poverty index and work status however not make it to the final model.

Overall, despite the fact that the final model is quite smaller than the original one, the model fits the data rather well. This is based on the tests we ran in our regression analysis and the results they produced. Because of this, all the explanatory variables that were left in the model appear to provide a good picture of the choice of living arrangement of an elderly individual residing in the United States.

## APPENDIX:

```
/*Create a data set with only the variables of interest*/

data clean1;

set w3sf21;

keep sf64 sf53 sf61 sf373 sf48 sf431 sf491 sf497

sf509 sf533 sf1742 sf70 sf115 sf57 sf144 sf182;

run;


/*Format the data for sf64 to be binomial*/

data clean2;

set clean1;

if sf64=1 then sf64=1;

else sf64=0;

run;


/*remove all unknown/nonresponse entries*/

data clean3;

set clean2;

if sf53=7 then delete;

if sf61=3 then delete;

if sf373=1 or sf373=2;

if sf48=7 then delete;

if sf431=98 or sf431=99 then delete;

if sf491=8 or sf491=9 then delete;

if sf497=8 or sf497=9 then delete;

if sf509=1 or sf509=2;

if sf533=1 or sf533=2;
if sf1742=9 or sf1742=8 or sf1742=0 then delete;
if sf70=6 then delete;
if sf115=5 then delete;
if sf57=3 then delete;
if sf144=1 or sf144=2 or sf144=3 or sf144=4 or sf144=5;
run;

data clean4;
set clean3;
if sf57=1 then FI=0;
else FI=1;
if sf61=1 then PI=0;
else PI=1;
if sf373=1 then Dis=0;
else Dis=1;
if sf497=1 then GT=0;
else GT=1;
if sf509=1 then WN=0;
else WN=1;
if sf533=1 then UVW=0;
else UVW=1;
run;

data clean5;
set clean4;
keep sf64 sf53 sf48 sf431 sf491 sf1742 sf70 sf115 sf144 sf182
FI PI Dis GT WN UVW;
label FI="FAMILY INCOME $20,000 OR MORE" PI="NHIS POVERTY INDEX"
Dis="RESP THINKS OTHR CONSIDER PERS HAV DISAB" GT="GET TOGETHER WITH FRIENDS OR NEIGHBORS"
WN="DO YOU WORK NOW" UVW="UNPAID VOLUNTEER WORK";
run;
```

```
proc freq data=clean5;

tables sf64 / binomial (level=2 ac wilson exact p=.5) alpha=.05;

run;


proc means data=clean5;

run;


proc corr data=clean5 noprob outp=OutCorr nomiss cov;

var sf64 sf53 sf48 sf431 sf491 sf1742 sf70 sf115 sf144 sf182

FI PI DIS GT WN UVW;

run;


proc freq data=clean5;

tables sf48*sf64

/nopercent norow nocol chisq relrisk riskdiff;

run;


proc freq data=clean5;

tables PI*sf64

/nopercent norow nocol chisq relrisk riskdiff;

run;


proc freq data=clean5;

tables FI*sf64

/nopercent norow nocol chisq relrisk riskdiff;

run;


proc freq data=clean5;

tables sf431*sf64

/nopercent norow nocol chisq relrisk riskdiff;
```

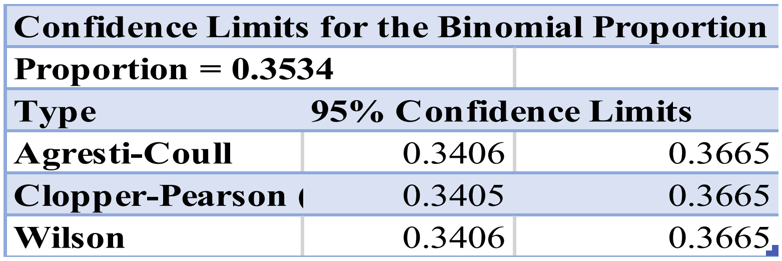
```
run;

proc univariate data=clean5 noprint;
      histogram sf64 sf48 PI FI;
run;

proc logistic data=clean5 descending /*plot=all plots(maxpoints=none)*/;
class sf182 sf48/para=ref;
model sf64=sf53 PI Dis sf48 sf431 sf491 GT
WN UVW sf1742 sf70 sf115 FI sf144 sf182/ covb aggregate scale=none;
output out=new1 p=pred;
run;

proc genmod data=clean5 descending;
model sf64=/ dist=binomial link=logit type1 type3 lrci;
output out=new1 p=pred;
run;

proc genmod data=clean5 descending;
class  sf182 sf48;
model sf64=sf53 PI Dis sf48 sf431 sf491 GT
WN UVW sf1742 sf70 sf115 FI sf144 sf182 /dist=binomial link=logit type1 type3
lrci;
output out=new1 p=pred;
run;

ods graphics on;
proc logistic data=clean5 descending /*plot=all plots(maxpoints=none)*/;
class sf182 sf48;
model sf64=sf53 PI Dis sf48 sf431 sf491 GT
WN UVW sf1742 sf70 sf115 FI sf144 sf182/ covb aggregate selection=backward
scale=none iplots;
output out=modified p=pred STDRESCHI=std_res RESCHI=Pearson_res;
```

```
run;
ods graphics off;

ods graphics on;
proc logistic data=clean5 descending /*plot=all plots(maxpoints=none)*/;
class sf182 sf48;
model sf64=sf53 PI Dis sf48 sf431 sf491 GT
WN UVW sf1742 sf70 sf115 FI sf144 sf182/ covb aggregate selection=backward
scale=none iplots;
output out=modified p=pred STDRESCHI=std_res RESCHI=Pearson_res;
run;
ods graphics off;

data clean6;
set clean5;
keep sf48 sf64 sf53 sf431 sf491 sf70 FI;
run;

proc logistic data=clean6 descending;
      class sf48;
      model sf64=sf53 sf48 sf431 sf491 sf70 FI/ covb aggregate scale=none;
      output out=new1 p=pred STDRESCHI=std_res RESCHI=Pearson_res;
run;

ods graphics on;
proc genmod data=clean6 descending plots=all;
class sf48;
model sf64=sf53 sf48 sf431 sf491 sf70 FI/ dist=binomial link=logit LRCI covb
type1 type3;
output out=out2 p=pred STDRESCHI=std_res RESCHI=Pearson_res;
run;
ods graphics off;
ods graphics on;
proc genmod data=clean6 descending plots=all;
class sf48;
model sf64=sf53 sf48 sf431 sf491 sf70 FI sf48*FI FI*sf431/ dist=binomial
link=logit LRCI covb type1 type3;
output out=out2 p=pred STDRESCHI=std_res RESCHI=Pearson_res;
run;
ods graphics off;


data new1;
set new1;
if pred >0.5 then Y_pred=1;
else Y_pred=0;
run;


proc freq order=data;
tables SF64*Y_pred/ nopercent norow nocol;
title "classification table";
run;
/*multinomail analysis*/
data new1;
set w3sf21;
keep sf64 sf53 sf61 sf373 sf48 sf431 sf491 sf497
sf509 sf533 sf1742 sf70 sf115 sf57 sf144 sf182;
run;
data new2;
set new1;
if sf64=1 then sf64=1;
else sf64=0;
if sf53=7 then delete;
if sf61=3 then delete;
if sf373=1 or sf373=2;
if sf48=7 then delete;
if sf431=98 or sf431=99 then delete;
if sf491=8 or sf491=9 then delete;
if sf497=8 or sf497=9 then delete;
if sf509=1 or sf509=2;
if sf533=1 or sf533=2;
if sf1742=9 or sf1742=8 or sf1742=0 then delete;
if sf70=6 then delete;
if sf115=5 then delete;
if sf57=3 then delete;
if sf144=9 or sf144='.' then delete;

run;


 /*binary logit*/

proc logistic data=new2;
title "baseline-category logit regression";
class   sf48(ref='Widowed' param=ref) sf182 (ref='West' param=ref)/
order=data ;
model sf64 (ref= 'Living alone') = sf53 sf61 sf373 sf48 sf431 sf491
sf497

sf509 sf533 sf1742 sf70 sf115 sf57 sf144 sf182 / selection=backward ;
output out=outgeneral p=pred STDRESCHI=std_res RESCHI=Pearson_res;
run;

/*multinominal cumulative logit*/
proc genmod data=new2 descending plots=all;
title "cumulative logit regression";
class  sf48(ref='Widowed' param=ref) sf182 (ref='West' param=ref)/
order=data ;
model sf64 (ref= 'Living alone') = sf53 sf61 sf373 sf48 sf431 sf491
sf497

sf509 sf533 sf1742 sf70 sf115 sf57 sf144 sf182 /dist=multinomial
                               link=cumlogit type1 type3 ;
output out=outmultinominal p=pred STDRESCHI=std_res
RESCHI=Pearson_res;;
run;


data clean1 (keep=sf64 sf53 sf373 sf48 sf431 sf491 sf70 sf57 sf182);
set outgeneral;
if Pearson_res>2 then delete;
if Pearson_res<-2 then delete;
```

```
run;

data clean2 (keep=sf64 sf53 sf61 sf373 sf48 sf431 sf491 sf497

sf509 sf533 sf1742 sf70 sf115 sf57 sf144 sf182);
set outmultinominal;
if Pearson_res>2 then delete;
if Pearson_res<-2 then delete;
run;

proc logistic data=clean1 DESCENDING;
model sf64=sf53 sf373 sf48 sf431 sf491 sf70 sf57
sf182/selection=backwards;
output out=out1 p=pred STDRESCHI=std_res RESCHI=Pearson_res;
run;
```

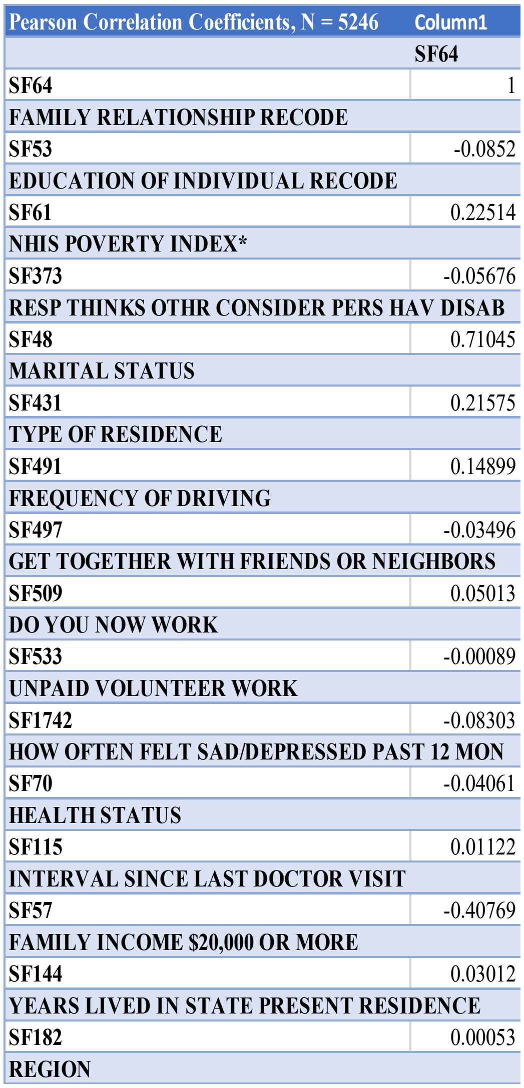
Finally, here is some of the SAS output:

Results from test of Binomial proportion (Table 1):



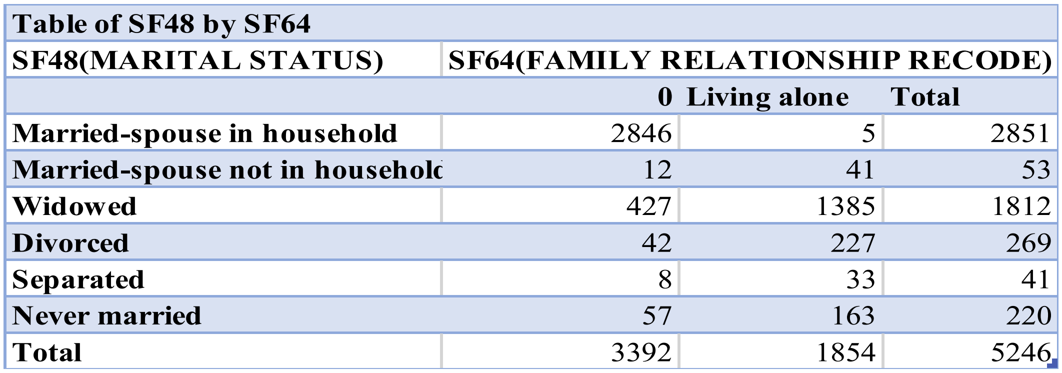
| **Confidence Limits for the Binomial Proportion** | | |
|---|---|---|
| **Proportion = 0.3534** | | |
| **Type** | **95% Confidence Limits** | |
| **Agresti-Coull** | 0.3406 | 0.3665 |
| **Clopper-Pearson (** | 0.3405 | 0.3665 |
| **Wilson** | 0.3406 | 0.3665 |

Exploratory Analysis-Correlations (Table 2):



| Pearson Correlation Coefficients, N = 5246 | Column1 |
|---|---|
| | SF64 |
| SF64 | 1 |
| FAMILY RELATIONSHIP RECODE | |
| SF53 | -0.0852 |
| EDUCATION OF INDIVIDUAL RECODE | |
| SF61 | 0.22514 |
| NHIS POVERTY INDEX* | |
| SF373 | -0.05676 |
| RESP THINKS OTHR CONSIDER PERS HAV DISAB | |
| SF48 | 0.71045 |
| MARITAL STATUS | |
| SF431 | 0.21575 |
| TYPE OF RESIDENCE | |
| SF491 | 0.14899 |
| FREQUENCY OF DRIVING | |
| SF497 | -0.03496 |
| GET TOGETHER WITH FRIENDS OR NEIGHBORS | |
| SF509 | 0.05013 |
| DO YOU NOW WORK | |
| SF533 | -0.00089 |
| UNPAID VOLUNTEER WORK | |
| SF1742 | -0.08303 |
| HOW OFTEN FELT SAD/DEPRESSED PAST 12 MON | |
| SF70 | -0.04061 |
| HEALTH STATUS | |
| SF115 | 0.01122 |
| INTERVAL SINCE LAST DOCTOR VISIT | |
| SF57 | -0.40769 |
| FAMILY INCOME $20,000 OR MORE | |
| SF144 | 0.03012 |
| YEARS LIVED IN STATE PRESENT RESIDENCE | |
| SF182 | 0.00053 |
| REGION | |

Exploratory Analysis-Contingency Table for Living Arrangement by Marital Status (Table 3):



| Table of SF48 by SF64 | | | |
|---|---|---|---|
| SF48(MARITAL STATUS) | SF64(FAMILY RELATIONSHIP RECODE) | | |
| | 0 | Living alone | Total |
| Married-spouse in household | 2846 | 5 | 2851 |
| Married-spouse not in householc | 12 | 41 | 53 |
| Widowed | 427 | 1385 | 1812 |
| Divorced | 42 | 227 | 269 |
| Separated | 8 | 33 | 41 |
| Never married | 57 | 163 | 220 |
| Total | 3392 | 1854 | 5246 |

Exploratory Analysis-Contingency Table for Living Arrangement by Family Income, where

$$FI = \begin{cases} 0, \; income < \$20{,}000 \\ 1, \; income > \$20{,}000 \end{cases}$$

(Table 4):

| Table of FI by SF64 | | | |
|---|---|---|---|
| FI(FAMILY INCOME $20,000 OR MORE) | SF64(FAMILY RELATIONSHIP RECODE) | | |
| | 0 | Living alone | Total |
| 0 | 1150 | 1419 | 2569 |
| 1 | 2242 | 435 | 2677 |
| Total | 3392 | 1854 | 5246 |

Exploratory Analysis-Contingency Table for Living Arrangement by Type of Residence (Table 5):

| Table of SF431 by SF64 | | | |
|---|---|---|---|
| SF431 | SF64 | | |
| (TYPE OF RESIDENCE) | (FAMILY RELATIONSHIP RECODE) | | |
| | 0 | Living alone | Total |
| Single family house or townhouse, | 3067 | 1286 | 4353 |
| Single family house, townhouse, o | 106 | 142 | 248 |
| Regular apartment | 199 | 325 | 524 |
| Supervised apartment | 2 | 19 | 21 |
| Other type of supervised group re | 0 | 3 | 3 |
| Assisted living facility | 3 | 9 | 12 |
| Retirement home | 8 | 37 | 45 |
| Center for Independent Living | 1 | 4 | 5 |
| Something else | 6 | 29 | 35 |
| Total | 3392 | 1854 | 5246 |

Exploratory Analysis - Simple Statistics (Table 6):

| Variable | Label | N | Mean | Std Dev | Min | Maxi |
|---|---|---|---|---|---|---|
| SF48 | MARITAL STATUS | 5246 | 2.095692 | 1.35148 | 1 | 6 |
| SF53 | EDUCATION OF INDIVIDUAL RECODE | 5246 | 2.9348075 | 1.4001 | 0 | 6 |
| SF64 | FAMILY RELATIONSHIP RECODE | 5246 | 0.3534121 | 0.47807 | 0 | 1 |
| SF70 | HEALTH STATUS | 5246 | 2.7335112 | 1.11882 | 1 | 5 |
| SF115 | INTERVAL SINCE LAST DOCTOR VISIT | 5246 | 1.1776592 | 0.59075 | 1 | 4 |
| SF144 | YEARS LIVED IN STATE PRESENT RESIDENCE | 5246 | 4.8879146 | 0.50054 | 1 | 5 |
| SF182 | REGION | 5246 | 2.464735 | 1.02987 | 1 | 4 |
| SF431 | TYPE OF RESIDENCE | 5246 | 1.4767442 | 1.66754 | 1 | 14 |
| SF491 | FREQUENCY OF DRIVING | 5246 | 1.9805566 | 1.28495 | 1 | 4 |
| SF1742 | HOW OFTEN FELT SAD/DEPRESSED PAST 12 MON | 5246 | 3.1736561 | 0.82117 | 1 | 4 |
| FI | FAMILY INCOME $20,000 OR MORE | 5246 | 0.5102936 | 0.49994 | 0 | 1 |
| PI | NHIS POVERTY INDEX | 5246 | 0.0974075 | 0.29654 | 0 | 1 |
| DIS | RESP THINKS OTHR CONSIDER PERS HAV DISAB | 5246 | 0.8644682 | 0.34232 | 0 | 1 |
| GT | GET TOGETHER WITH FRIENDS OR NEIGHBORS | 5246 | 0.2552421 | 0.43604 | 0 | 1 |
| WN | DO YOU WORK NOW | 5246 | 0.9027831 | 0.29628 | 0 | 1 |
| UVW | UNPAID VOLUNTEER WORK | 5246 | 0.7933664 | 0.40493 | 0 | 1 |

Regression Analysis – Fitting a model with all the variables using a logit link (Table 7):

| Analysis of Maximum Likelihood Estimates | | | | | | |
|---|---|---|---|---|---|---|
| Parameter | | DF | Estimate | Standard | Wald | Pr > ChiSq |
| | | | | Error | Chi-Square | |
| INTERCEPT | | 1 | 2.0343 | 0.7129 | 8.1418 | 0.0043 |
| SF53 | | 1 | 0.2859 | 0.0498 | 32.9908 | <.0001 |
| PI | | 1 | 0.0631 | 0.1836 | 0.1182 | 0.731 |
| DIS | | 1 | -0.2718 | 0.1663 | 2.6725 | 0.1021 |
| SF48 | Divorced | 1 | 0.2718 | 0.1986 | 1.8727 | 0.1712 |
| SF48 | Married-spouse in household | 1 | -8.2182 | 0.4975 | 272.8662 | <.0001 |
| SF48 | Married-spouse not in housel | 1 | 0.2833 | 0.3697 | 0.5871 | 0.4435 |
| SF48 | Never married | 1 | -0.1577 | 0.1907 | 0.6842 | 0.4082 |
| SF48 | Separated | 1 | -0.00779 | 0.4405 | 0.0003 | 0.9859 |
| SF431 | | 1 | 0.3688 | 0.0563 | 42.9377 | <.0001 |
| SF491 | | 1 | -0.3425 | 0.0479 | 51.2034 | <.0001 |
| GT | | 1 | -0.1449 | 0.1345 | 1.1613 | 0.2812 |
| WN | | 1 | 0.2936 | 0.2058 | 2.0357 | 0.1536 |
| UVW | | 1 | 0.0458 | 0.158 | 0.0838 | 0.7722 |
| SF1742 | | 1 | -0.0706 | 0.069 | 1.0485 | 0.3059 |
| SF70 | | 1 | -0.2628 | 0.0566 | 21.5594 | <.0001 |
| SF115 | | 1 | -0.1482 | 0.0937 | 2.4994 | 0.1139 |
| FI | | 1 | -2.404 | 0.1379 | 303.8198 | <.0001 |
| SF144 | | 1 | 0.196 | 0.1065 | 3.3909 | 0.0656 |
| SF182 | Midwest | 1 | 0.0136 | 0.1762 | 0.0059 | 0.9386 |
| SF182 | Northeast | 1 | -0.3845 | 0.1773 | 4.7008 | 0.0301 |
| SF182 | South | 1 | -0.1427 | 0.168 | 0.7217 | 0.3956 |

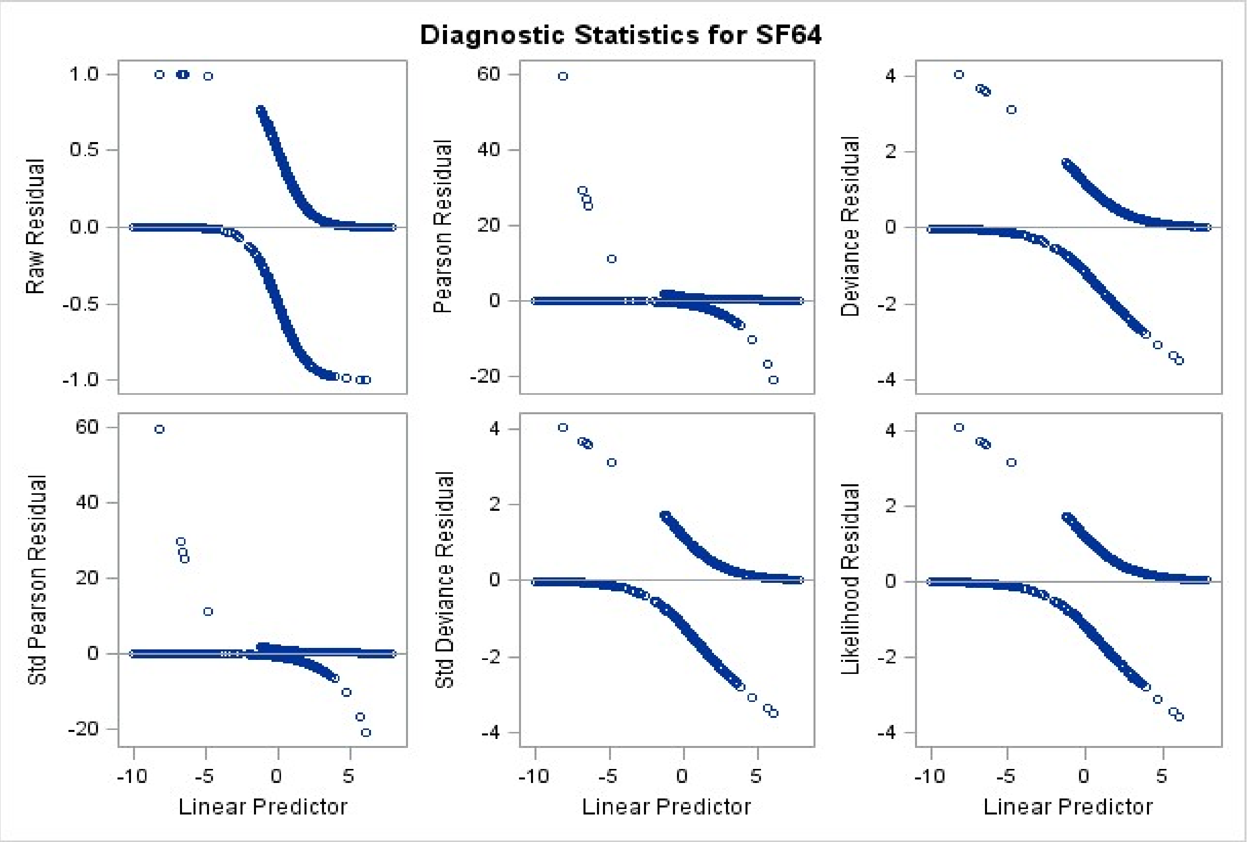
After running proc logistic with backwards regression, we are left with 6 significant explanatory variables – education, marital status, type of residence, frequency of driving, health status, and family income greater than $20,000 (Table 8):

| Analysis of Maximum Likelihood Estimates | | | | | | |
|---|---|---|---|---|---|---|
| Parameter | | DF | Estimate | Standard | Wald | Pr > ChiSq |
| | | | | Error | Chi-Square | |
| INTERCEPT | | 1 | 1.0518 | 0.2803 | 14.0861 | 0.0002 |
| SF53 | | 1 | 0.2793 | 0.0475 | 34.583 | <.0001 |
| SF48 | Divorced | 1 | 1.6112 | 0.2003 | 64.7212 | <.0001 |
| SF48 | Married-spouse in household | 1 | -6.875 | 0.4159 | 273.2464 | <.0001 |
| SF48 | Married-spouse not in household | 1 | 1.5362 | 0.3215 | 22.8339 | <.0001 |
| SF48 | Never married | 1 | 1.1051 | 0.1939 | 32.4894 | <.0001 |
| SF48 | Separated | 1 | 1.3132 | 0.379 | 12.0086 | 0.0005 |
| SF431 | | 1 | 0.3684 | 0.0561 | 43.1467 | <.0001 |
| SF491 | | 1 | -0.3361 | 0.0446 | 56.8931 | <.0001 |
| SF70 | | 1 | -0.2043 | 0.0512 | 15.9424 | <.0001 |
| FI | | 1 | -2.4074 | 0.1301 | 342.63 | <.0001 |

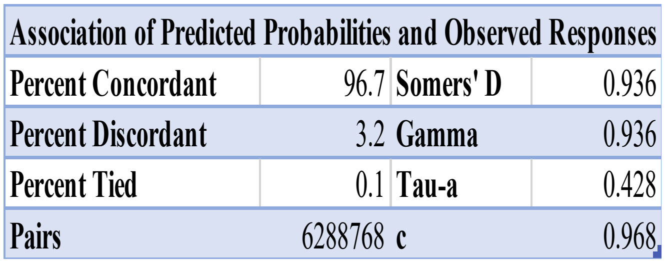
Summary of backwards selection (Table 9):

| Summary of Backward Elimination | | | | | | |
|---|---|---|---|---|---|---|
| Step | Effect | DF | Number | Wald | Pr > ChiSq | Variable |
| | Removed | | In | Chi-Square | | Label |
| 1 | UVW | 1 | 14 | 0.0838 | 0.7722 | UNPAID VOLUNTEER WORK |
| 2 | PI | 1 | 13 | 0.1154 | 0.7341 | NHIS POVERTY INDEX |
| 3 | SF1742 | 1 | 12 | 1.0219 | 0.3121 | HOW OFTEN FELT SAD/DEPRESSED PAST 12 MON |
| 4 | GT | 1 | 11 | 1.033 | 0.3095 | GET TOGETHER WITH FRIENDS OR NEIGHBORS |
| 5 | WN | 1 | 10 | 2.1634 | 0.1413 | DO YOU WORK NOW |
| 6 | SF115 | 1 | 9 | 2.6792 | 0.1017 | INTERVAL SINCE LAST DOCTOR VISIT |
| 7 | DIS | 1 | 8 | 3.3315 | 0.068 | RESP THINKS OTHR CONSIDER PERS HAV DISAB |
| 8 | SF144 | 1 | 7 | 3.4663 | 0.0626 | YEARS LIVED IN STATE PRESENT RESIDENCE |
| 9 | SF182 | 3 | 6 | 7.5903 | 0.0553 | REGION |

Residual plots and the following tables show that the model is a good fit (Graph 1):

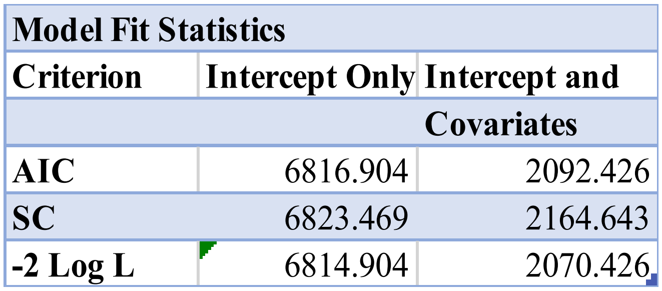


Predicted probabilities and Observed Responses(Table 9):



| Association of Predicted Probabilities and Observed Responses | | | |
|---|---|---|---|
| Percent Concordant | 96.7 | Somers' D | 0.936 |
| Percent Discordant | 3.2 | Gamma | 0.936 |
| Percent Tied | 0.1 | Tau-a | 0.428 |
| Pairs | 6288768 | c | 0.968 |

Model Fit Statistics (Table 10):



| Model Fit Statistics | | |
|---|---|---|
| Criterion | Intercept Only | Intercept and Covariates |
| AIC | 6816.904 | 2092.426 |
| SC | 6823.469 | 2164.643 |
| -2 Log L | 6814.904 | 2070.426 |