Affordable Housing Comparison between New Jersey, Alabama, and Florida

Written by: Group 7
Lina Gao, Yu-Fen Huang, Wei Tang, Rhea Vettithanam, Mingzhe Yu

November 21, 2018

Summary

Overall, the three states- New Jersey (NJ), Alabama (AL), and Florida (FL)- have various differences and similarities. By controlling factors like owner, rental, etc, NJ has the highest median income and rent value; Fl has the lowest proportion of low-income families; AL has the lowest rent value, the highest proportion of low-income families and units inhabited by low-income families that are affordable and uncrowded. In general, NJ has the highest economic level whereas AL has the lowest one. With regards to variables impacting the rental situation, these three states are affected by different factors (e.g. age, gender, rent payment, etc). For example, rental in AL is mostly affected by age: the older the people, the more difficult for them to get a satisfying rent living unit. However, in NJ and FL, rental is highly affected by average rent payment. Gender difference plays a huge role for the rental house in AL whereas in NJ and FL, its effect is limited.

Important Variables

Dataset: G7ALAFLO, Gnj

Important variables:
aff = if the house is affordable (For owner, if smocapi <= 30, aff = 1, otherwise aff = 0; for rental, if grapi <= 30, aff = 1, otherwise aff = 0)
affc = affordable * uncrowd (if aff =1 and uncrowd = 1 then affc = 1; otherwise, affc = 0)
age = age of the head of household
bldgsz = building size
cnum = uncrowd*new*hlinc;
denom = new;
denom2 = new*hlinc;
educ = education level of the head of household
edug = high education (if educ >7 then edug = 1; otherwise, edug = 0)
enum = aff*new*hlinc
finc = family income
grapi = percent of income going to rent
gra = log((grapi+2)/(102-grapi))
grent = rent per month
Hht = household type (1,2,3 = family)
hinc = household income
hinch = high household income (if lhinc > median of lhinc for rentals or owners in each states, then hinch =1; otherwise hinch = 0)
lhinc = log(1+hinc), log household income
hlinc = low income household (0=high household income, 1=low household income)
hweight = census weight
morth = high mortgage, (if lmort > median of the lmort for owners in each state, then morth =1; otherwise, morth = 0)
lmort = log(smocapi*hinc)
new = houses which are built in 1,2 or 3 years, yrbuilt, (0=old, 1=new)
persons = number of people in the household
Puma1 = puma 1% sample region
rat = snum/sdenom, among the new units, what is the proportion of units that are low income, and are affordable and are not crowded.
rat2 = snum/sdenom2, Among all new units that are occupied by low income families, what is the proportion of units that are affordable and are not crowded.
rat3 = slinc/stot, proportion of low income families

renth = high rent (if grent > median of household data for rentals in each state, then renth =1 otherwise renth = 0)
rooms = number of rooms
senior = senior house head (if age>=65 then senior = 1; otherwise, senior = 0)
sex = gender of the head of household
smi = rentals or owners (0=rentals, 1=owners)
smocapi = percent of income going to house
trvtime = travel time of the head of household
uncrowd = persons <= rooms (0=crowded house, 1=uncrowd house)
units= type of unit (0=housing unit,1,2=inst. Quarters)
Value = housing value code
Valuen = transformed house values
vehicl = number of cars
xnum = Aff*uncrowd*new*hlinc;
yrbuilt = year built 1,2,3 last ten years

Introduction

This project focuses on comparing low-income affordable housing in New Jersey (NJ) to that of Alabama (AL) and Florida (FL). The data includes information about household income, rents, house value, size of houses, the amount of income allocated to housing, family size, gender, age, education level, etc. We have three goals in this project: compare the income distributions (along with rent and housing values), compare the state basic statistics (e.g.median income), and produce logistic regression models for affordable and uncrowded housing in each state. The key results are summarized in the following sections and the appendix.

Methods and Results

Part I- Income, Rent, and Property Distributions in NJ, AL and FL

Because the data includes more than 530,000 observations across NJ (New Jersey, state 34), AL(Alabama, state 1) and FL(Florida, state 12), the percentile distribution of key variables (including household income, property values, and rents) in these three states are plotted to compare differences before further analysis. Using proc univariate, 1%, 2%, 3%, 4%, 5%, 7%, 10%, 20%, 30%, 40%, 60%, 80%, 90%, 93%, 95%, 96%, 97%, 98% and 99% of household income, property value for home owner, and rent for renters in NJ, AL and FL are calculated to draw percentile-income/property value/rents plots and then percentile-percentile plots to compare state distributions, as shown in Figures 1 through 6.

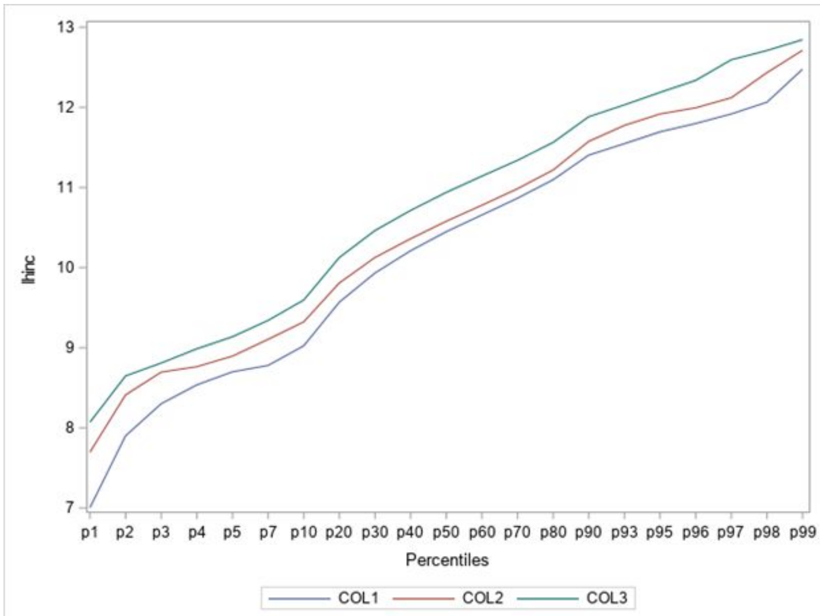


Figure 1. The percentiles of log household income distribution in AL(col1), FL(col2) and NJ(col3).

Similar percentile distribution of household income in AL, FL and, NJ is observed in Figure 1, with col3 (NJ) > col2 (FL) > col1(AL) across 1st through 99th percentiles.

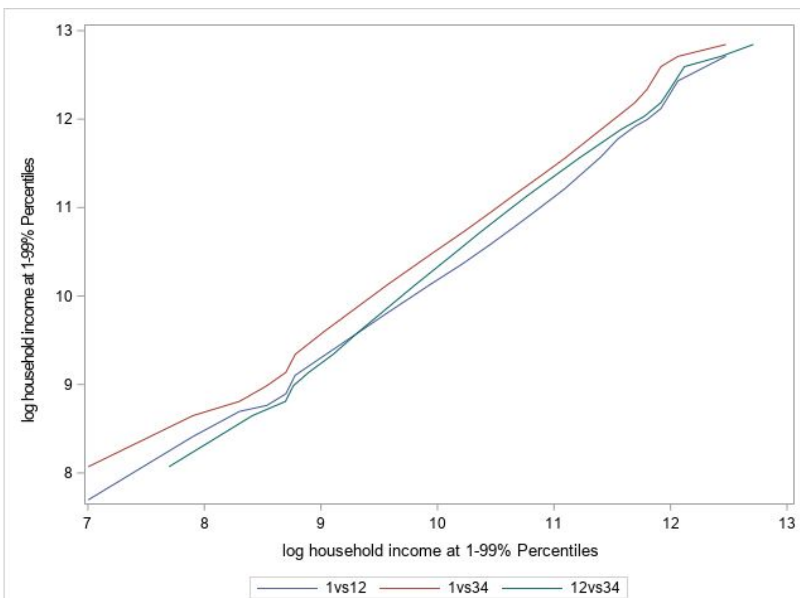


Figure 2. Percentile-Percentile household income distribution comparison among AL(1), FL(12) and NJ(34).

The pair comparison of household income distribution can be obtained from observing Figure 2. Although the percentile distribution of household income in all these three states was similar, the sequence for the difference between these states was NJ vs AL > NJ vs FL and FL vs AL - this is because NJ and AL had the biggest difference in distribution. The roughly linear lines on the percentile-percentile plot indicate that the states have similar household income distributions.

The same data analysis was performed for house rentals in NJ, AL, and FL. The corresponding results are shown in Figures 3 and 4. When the rent is less than $500 dollars/month, the percentile distribution of rent/month, grent, in FL and NJ is almost the same from 1st-10th percentiles. When rent is more than $500 dollars/month, the rent in NJ is consistently higher than FL from percentile 10% to 99%. As for AL, it has the lowest rent across the entire percentiles (1-99%) among these three states with the largest difference from NJ and FL observed at the highest percentiles. When pair comparison was performed, similar results can be obtained. The distribution at lower rent was minor between these three states, but differences increase as rent increased, especially from the 90th to 98th percentile. NJ has the highest rent compared to the other two states. The paired comparison found the largest difference for NJ vs AL, followed by AL vs FL and NJ vs FL. NJ and FL show a similar distribution pattern.

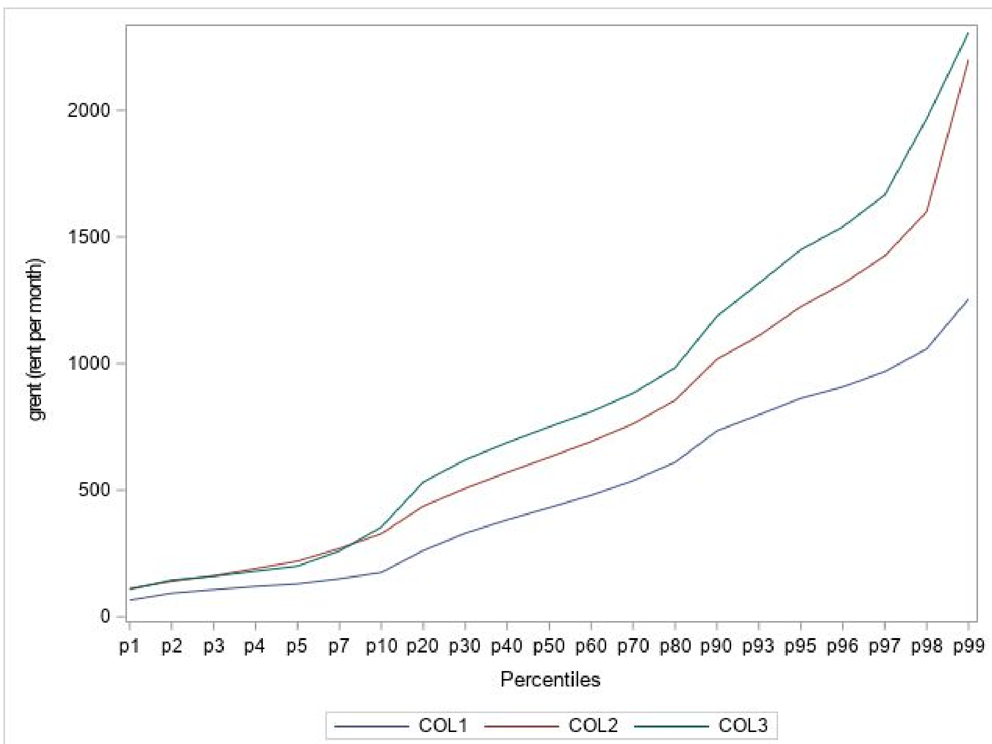


Figure 3. The percentile distribution of grent (rent per month) in AL(col1), FL(col2) and NJ(col3)

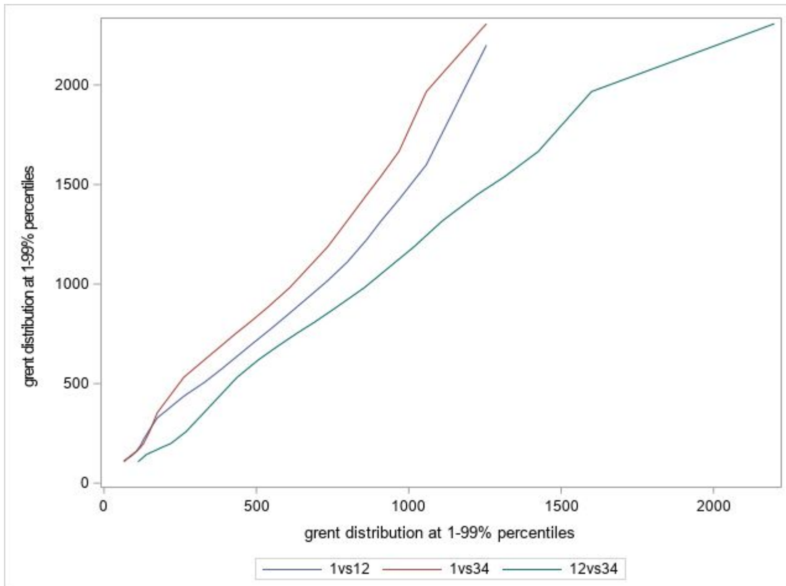


Figure 4. The grent (rent per month) comparison at 1-99% percentiles in AL(1), FL(12) and NJ(34)

Next, we compared house values across all three states. As indicated in Figure 5, although a higher house value was observed in NJ than FL from 1% to 98% percentile, the two states show almost the same pattern at 98%-99% percentile, further demonstrated by the y=x distribution of higher values in Figure 6. FL and AL show very similar distribution at 1-80% percentiles. The similarity and dissimilarities of the distributions are highlighted in Figure 6. From the paired comparison, the largest difference was observed for AL vs NJ, followed by NJ vs FL and AL vs FL. In summary, NJ shows the highest household income, rent, and house value among these three states. FL is the second, followed by AL.

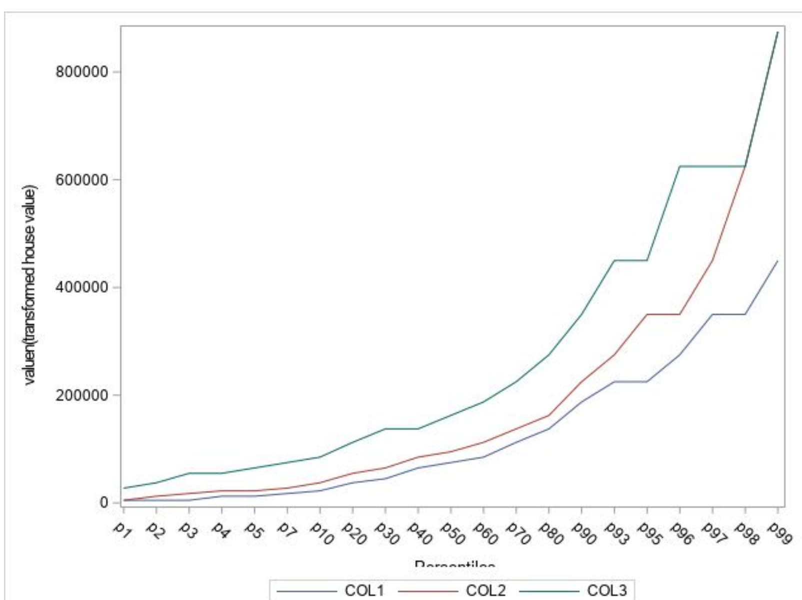


Figure 5. The percentile distribution of house value in AL(col1), FL(col2) and NJ(col3)

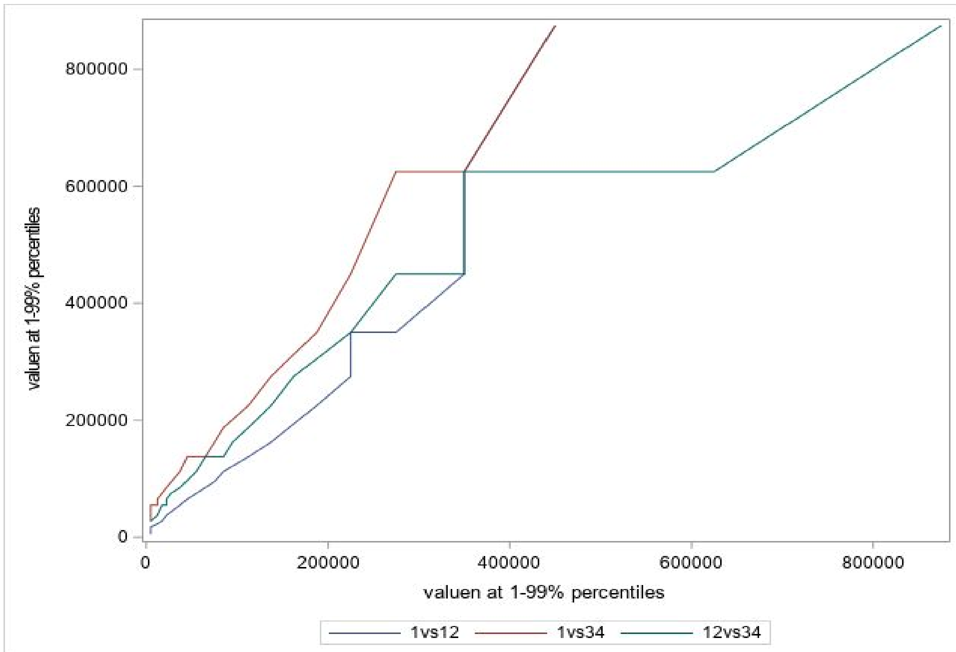


Figure 6. The pair comparison of house value distribution at 1-99% percentiles in AL(1), FL(12) and NJ(34).

Part II- Basic Statistics of NJ, AL, and FL

A. Median Income

We compared the median of family income (FINC) and household income (HINC) between the three states. AL has the lowest median income whereas NJ has the highest. The 2000 Census of Population and Housing's Technical Documentation states that because many households consist of only one person, average household income is usually less than the average family income. Below, it is evident that the HINC median is consistently larger than the FINC median across all three states. This indicates that in AL, FL, and NJ, many households consist of more than one person.

Table 1. Comparison of FINC and HINC medians across AL, FL, and NJ

	AL	FL	NJ
FINC median	$27,365	$28,000	$44,100
HINC median	$34,500	$39,300	$56,343

Detailed summary statistics for affordable housing in NJ, AL, and FL table can be found in the Appendix.

B. Proportion of Low-Income Families (rat3)

The proportion of low-income families (rat3) is calculated by slinc (sum of household income) / stot (sum of total). The box-plot of rat3, Figure 7, shows that AL has the highest proportion of low-income families, and FL has the lowest. A one-way ANOVA was performed to test if the rat3 across the three states were significantly different. With $p<.0004$, we confirm that there is a significant difference between AL, FL, and NJ. The qqplot and residual plot indicate that normality of ANOVA assumption is reasonable. The homogeneity of variance is confirmed by the Levene's Test ($p = 0.5745$). Multiple group comparison through a two-way ANOVA helps us identify the significant difference between these state pairs. Tukey-Kramer adjust test is applied to control EER. EER, the experimentwise error rate, is the probability of making at least one Type 1 error when we perform the entire set of comparisons. The results are shown in Figure 8 - the significant difference is identified between AL vs. FL ($p = 0.0011$) and NJ vs FL ($p = 0.0201$); AL vs NJ ($p = 0.3227$) is found to be insignificant with $\alpha = 0.05$.

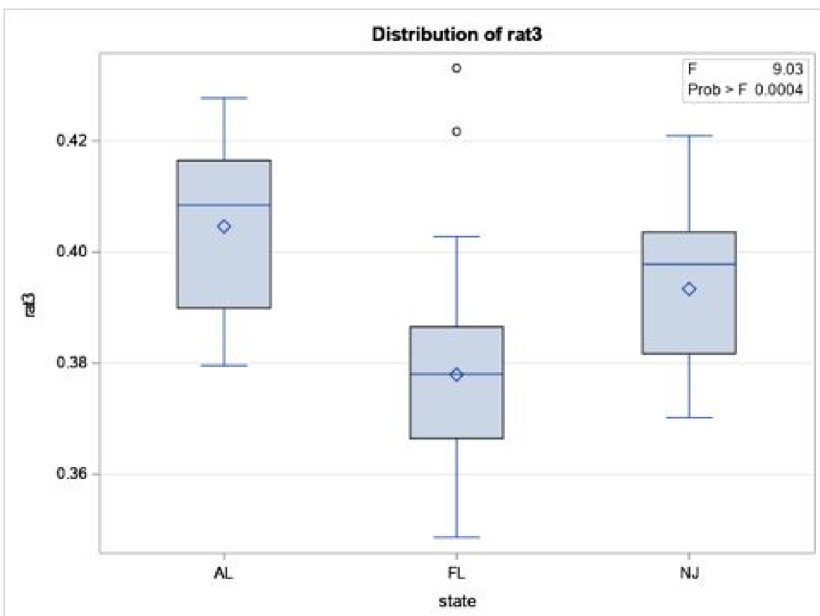


Figure 7. The boxplot of rat3 for AL, FL, and NJ

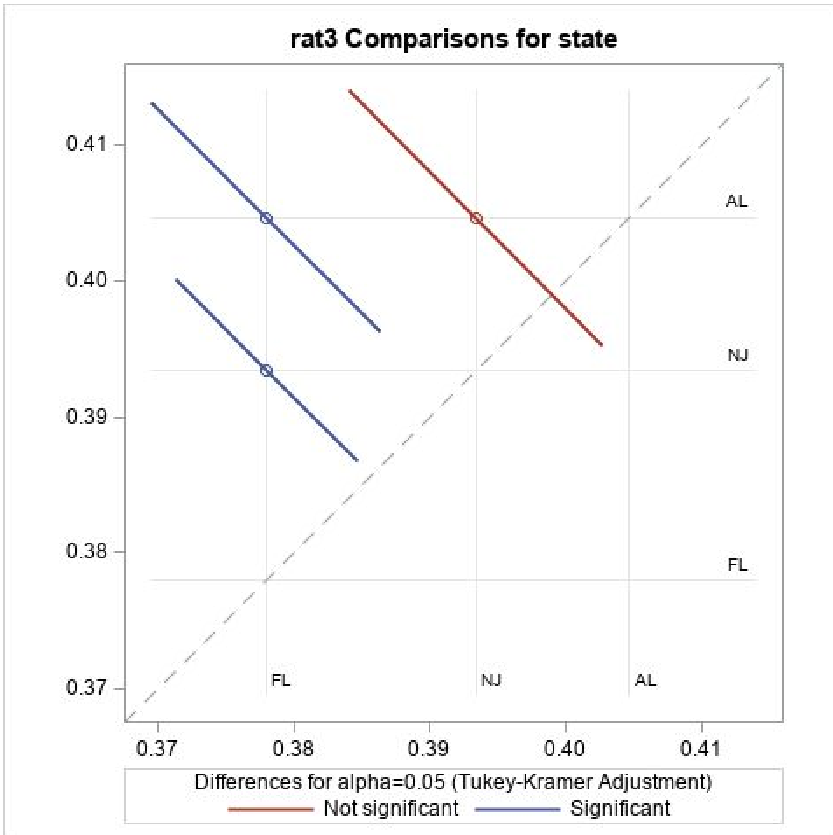


Figure 8. The Tukey's test of rat3 between NJ, AL and FL.

C. Proportion of Low-Income Units That are Affordable and Uncrowded (rat) Among New Units

The proportion of low-income unit that is affordable and uncrowded (rat) among new units is calculated by snum (sum of xnum= Aff*uncrowd*new*hline) / sdenom (sum of denom= new). AL has the highest proportion followed by NJ and FL, as shown in Figure 9. A one-way ANOVA test shows that there is a significant difference between these states ($p < 0.0001$). The diagnostic test for ANOVA analysis indicated that the variance of rat between these states is homogenous ($p = 0.3802$), and the normality of residual is confirmed by qqplot and predicted value-residual plot. The two-way ANOVA analysis identifies that a significant difference exists in AL vs FL ($p <.0001$) and AL vs NJ ($p = 0.0004$), but not FL vs NJ ($p = 0.9809$). The results are shown in Figure 10.

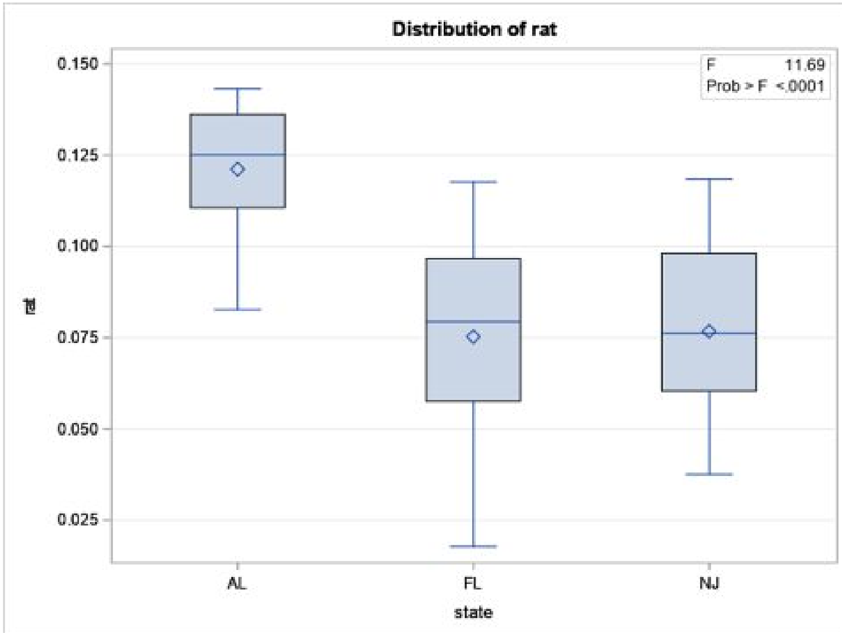


Figure 9. The boxplot of rat of NJ, FL and AL

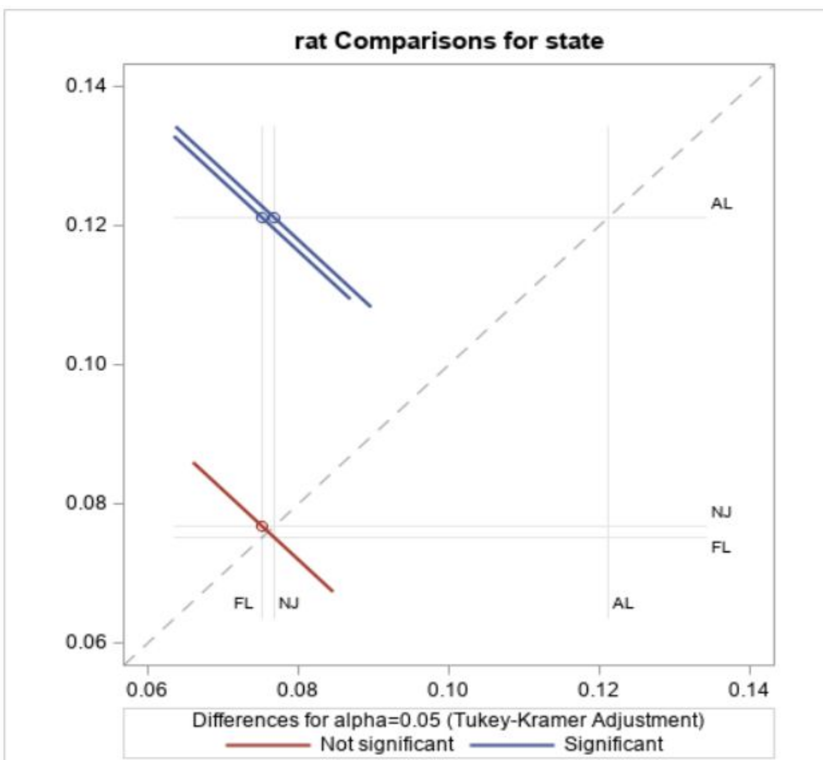


Figure 10. The pairwise comparison of rat in NJ AL and FL using Tukey's test

D. Proportion of New Units Inhabited by Low-Income Families that are Affordable and Uncrowded (rat2)

The proportion of new units inhabited by low-income families that are affordable and uncrowded (rat2) is calculated by snum (sum of xnum= Aff*uncrowd*new*hline) / sdenom2 (sum of denom2= new*hlinc). As shown in Figure 11, AL has the highest proportion, and NJ has the lowest. The one-way ANOVA results indicate a significant difference of rat2 between the states ($p < 0.0026$). The ANOVA assumptions are confirmed by the qqplot and Levene's Test ($p = 0.0583$). The two-way ANOVA analysis further identifies that a significant difference exists between AL vs FL ($p = 0.0035$) and AL vs NJ ($p = 0.0041$), but not FL vs NJ ($p = 0.9210$) at α=.05, as indicated in Figure 12.

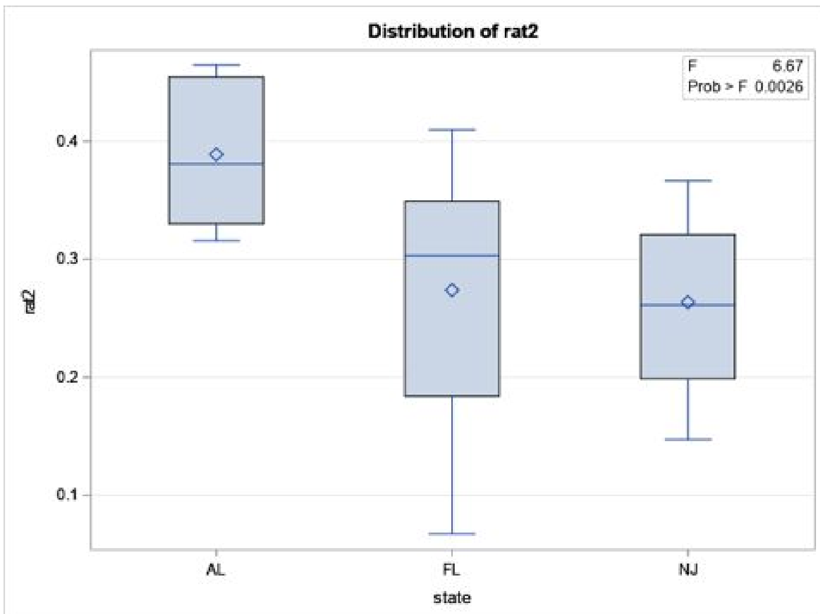


Figure 11. The boxplot of rat 2 of NJ, FL, and AL

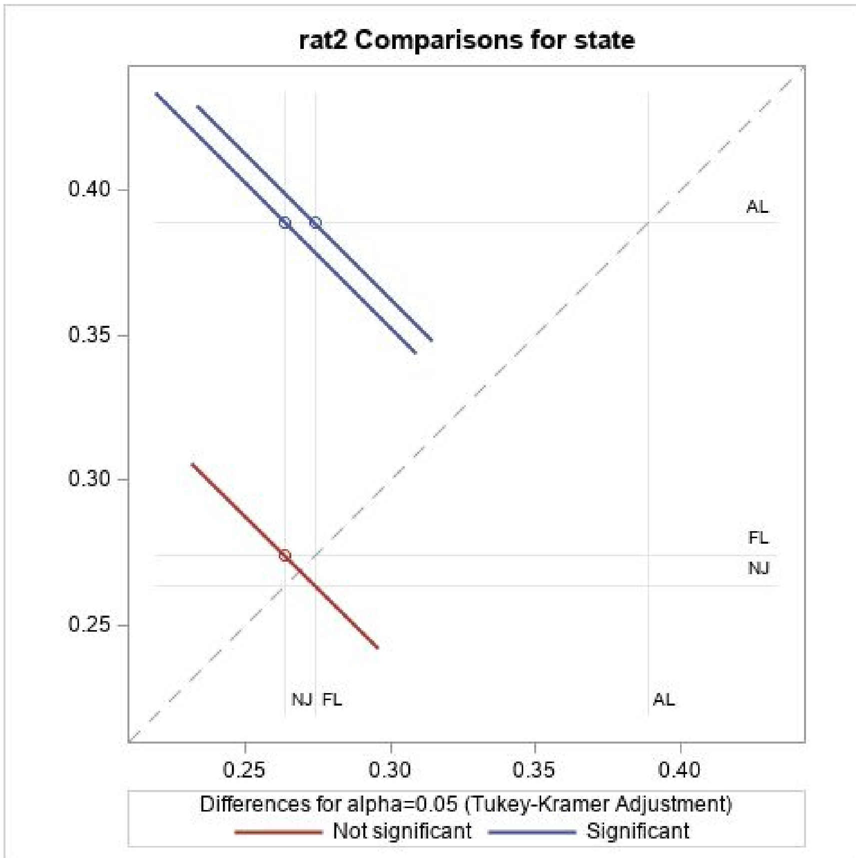


Figure 12. The pairwise comparison of rat2 using Tukey's test between NJ, FL and AL

In summary, AL has the lowest family income median and household income median whereas NJ has the highest median income, both in household and family. AL has the highest proportion of low-income families (rat3), and FL has the lowest. Among the new units, AL has the highest proportion of low-income units that are affordable and uncrowded (rat), and NJ and FL have the lower proportions. Furthermore, AL has the highest proportion of new units occupied by low-income families that are affordable and uncrowded (rat2), and NJ has the lowest.

Part III- Logistic Regression Models

In order to develop the logistic model for affordable and uncrowded housing (affc) in NJ, AL and FL, all possible related categorical variables are selected, including senior, renth, edug, hinch, sex, bldgsz, persons, rooms and vehicl. Considering possible collinearity between those predictors, lasso regression is applied for the initial variable selection. Compared to traditional forward, backward, and stepwise variable selection techniques, lasso provides several advantages. It is a shrinkage method that provides a unified variable selection and parameter estimation framework and more stable, and it does not suffer from high variability. By forcing the sum of the absolute value of the regression coefficients to be less than a fixed value, the coefficients of unimportant predictors can shrink to zero. Therefore, besides regularization during model development, lasso performs variable selection task simultaneously. Using glmnet in R, all predictors, senior, renth, edug, hinch, sex, bldgsz, persons, rooms and vehicl, are evaluated for the logistic model development for affc household controlling hlinc and smi in NJ, AL and FL (low household income owner (hlinc =1, smi=1), high household income owner (hlinc = 0, smi = 1), low household income rental (hlinc = 1, smi = 0), high household income rental (hlinc = 0, smi = 0)). After predictor selection, logistic models are developed in SAS using proc logistic. The convergence criterion are met for all developed models. The significant p-values are obtained in the testing global null hypothesis tests (BETA=0) for all models as well.

I. Model Fitting for NJ Data

a. Variable Selection by Lasso

Collinearity between these predictors is observed when all 9 variables (senior, renth, edug, hinch, sex, bldgsz, persons, rooms and vehicl) are included in the model development for NJ data. Using lasso, only the significant predictors are selected, as shown in Table 2**.** The selected predictors are displayed in red.

Table 2. Selection of predictors using lasso regression (selected shown in red)

Predictor Selection by Lasso for NJ Logistic Model Development

High income rentals (model1)	Low income rentals (model2)	high income owners (model3)	Low income owners (model4)
renth	renth	morth	morth
hinch	hinch	hinch	hinch
rooms	sex	sex	sex
edug	rooms	edug	edug
vehi	edug	senior	senior
persons	vehi	vehi	vehi
senior	persons	rooms	rooms
sex	bldgsz	persons	persons
bldgsz	senior	bldgsz	bldgsz

b. Logistic Model Fitting in SAS for NJ Data

Table 3. Coefficients of predictors in the logistic models for NJ data

predictor	Model 1 Rental,high income	Model 2 Rental,low income	Model 3 Owner,high income	Model 4 Owner,low income
hinch	-1.8082 (p<.0001)	-2.4881 (p<0.0001)	-2.6337 (p<0.0001)	-2.8897 p<0.0001)
sex	N/A	-0.5162 (p<0.0001)	0.3257 (p<0.0001)	0.3251 (p<0.0001)
rooms	-0.7486 (p<.0001)	-0.0419 (p<0.0001)	-0.5206 (p<0.0001)	0.0492 (p<0.001)
edug	0.3685 (p<.0001)	0.0780 (p=0.015)	-0.4815 (p<0.0001)	-0.2780 (p<0.0001)
vehi	-0.2424 (p<.0001)	-2.4881 (p<0.0001)	-0.1605 (p<0.0001)	-0.1294 (p<0.0001)
renth	1.8814 (p<.0001)	2.5433 (p<0.0001)	N/A	N/A
morth	N/A	N/A	2.5541 (p<0.0001)	4.3962 (p<0.0001)
persons	0.9432 (p<.0001)	N/A	-0.1777 (p<0.0001)	N/A
bldgsz	N/A	N/A	N/A	N/A
senior	1.4259 (p<.0001)	N/A	-0.6313 (p<0.0001)	-0.3876 (p<0.0001)

c. Model 1- Rentals with High Income in NJ

Using the selected predictors in Table 2, the hypothesis output in SAS gives a likelihood ratio chi-square of 10201.5568 with a p-value of 0.0001, indicating the fitting of Model 1 is significant. In addition, the p-values of all the predictors are less than 0.0001.

Regression model 1 is written as follows:

$$log(odds) = log(\frac{P}{(1-P)}) = -1.3033 + 1.4259*senior + 0.9432*persons$$
$$+ 1.8814*renth - 1.8082*hinch + 0.3685*edug - 0.7486*rooms - 0.2424*vehicl$$

The coefficients can be interpreted as the multiplicative change in the odds for a one unit change in the predictor. For one unit increase in renth, there is a 1.8814 unit increase observed in the log of odds that units aren't affordable. Log means natural log, rather than log with base 10. The odds ratio graph also indicates that all variables significantly contribute to the model fitting with renth as the most positive predictor and hinch as the most negative predictor.

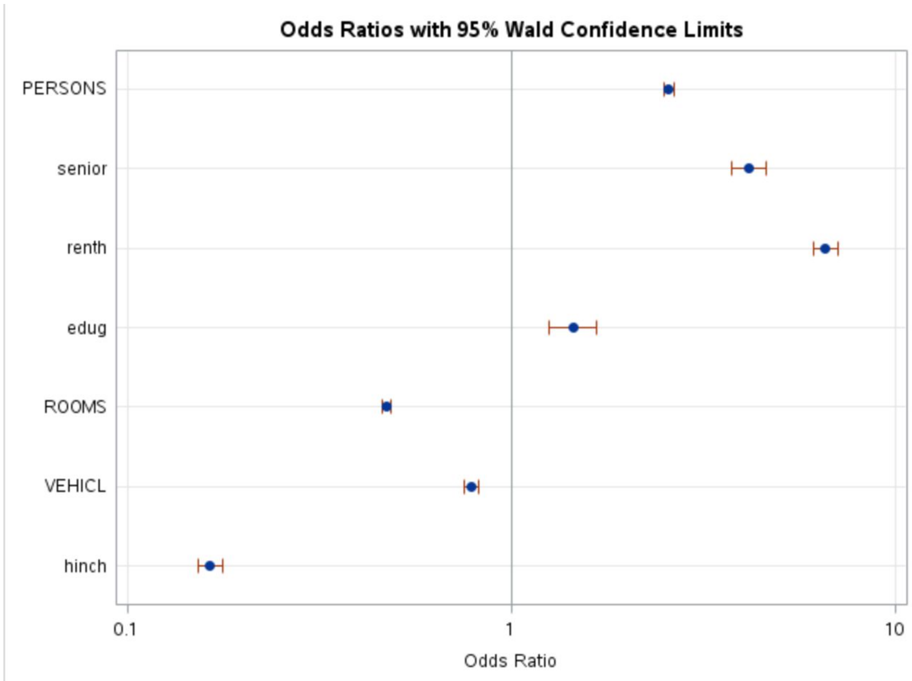


Figure 13. Odds ratios graph(Logbase=10) of affc for high income rentals in NJ

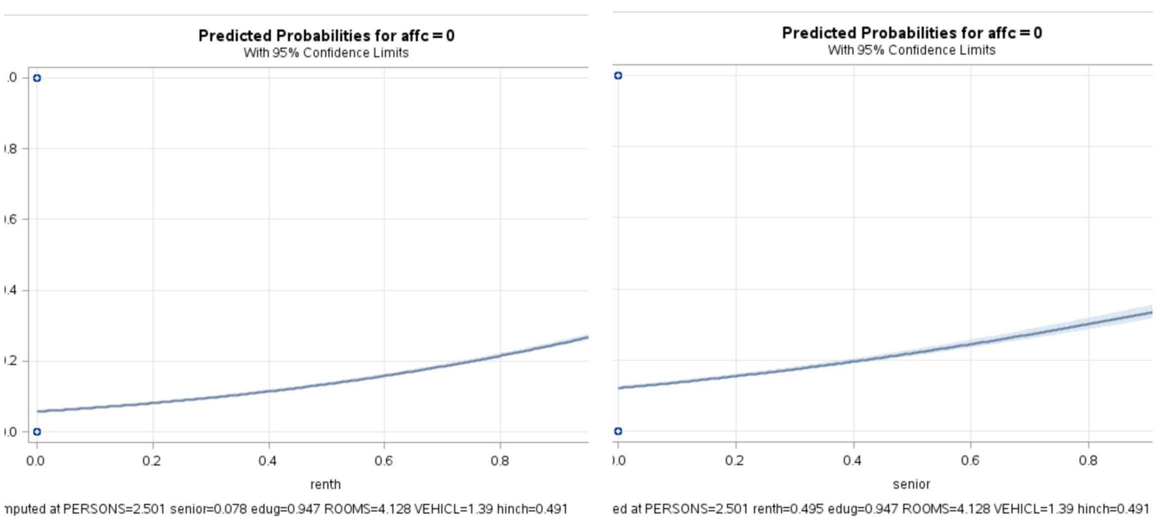


Figure 14. The relationship of predictor and affc for high income rentals in NJ; (a) predictor grent, (b) predictor senior

In this model, the predictor is odd of renth, 0/1(grent: rent per month is less than $1,100 or greater than $1,100 where $1,100 is the median grent). This figure indicates that with an increase in rent per month, the probability of unaffordable units increases, as shown in figure 14.a. Figure 14.b shows a similar pattern for the effect of predictor senior on affc.

d. Model 2- Rentals with Low Income in NJ

For Rental with low income, we use six predictors to fit the model by SAS. The likelihood ratio chi-square is 7828.2578 with a p-value less than 0.0001. The p-value of likelihood estimate for all predictor variables is less than 0.0001 except edug (p = 0.0150).

Model 2 is described as below:

$$log(odds) = log(\frac{P}{(1-P)}) = 2.0367 - 0.5162*sex - 0.0419*rooms$$
$$+2.5433*renth + 0.0780*edug - 2.4881*hinch - 0.1534*vehicl$$

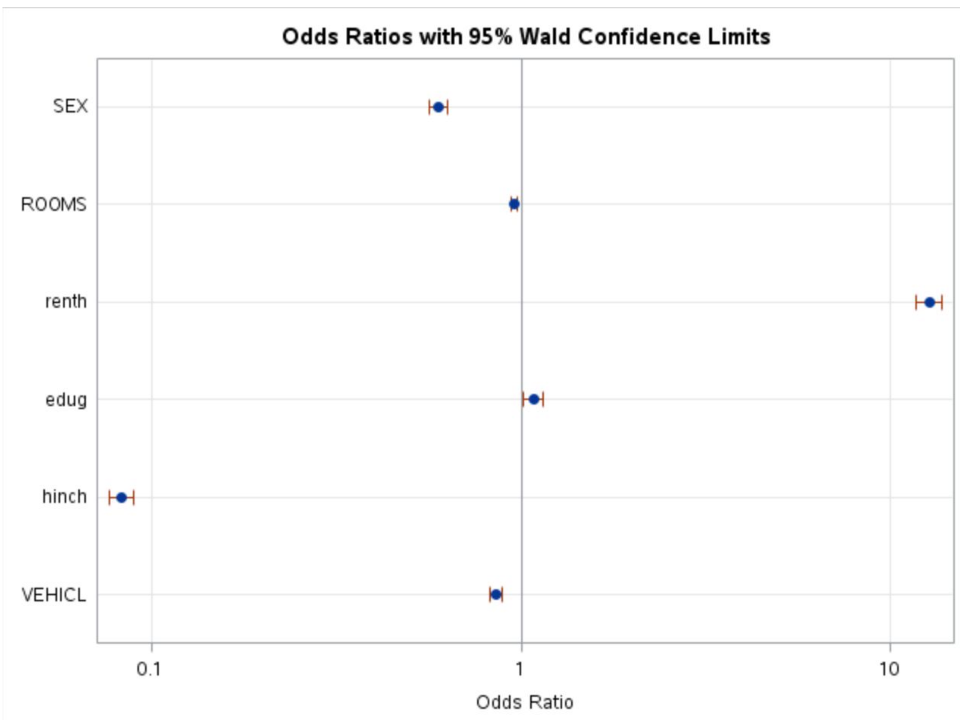


Figure 15. Odds ratios graph (Logbase=10) of affc for low income rentals in NJ

Similar to Model 1, renth and hinch are the most influential predictors.

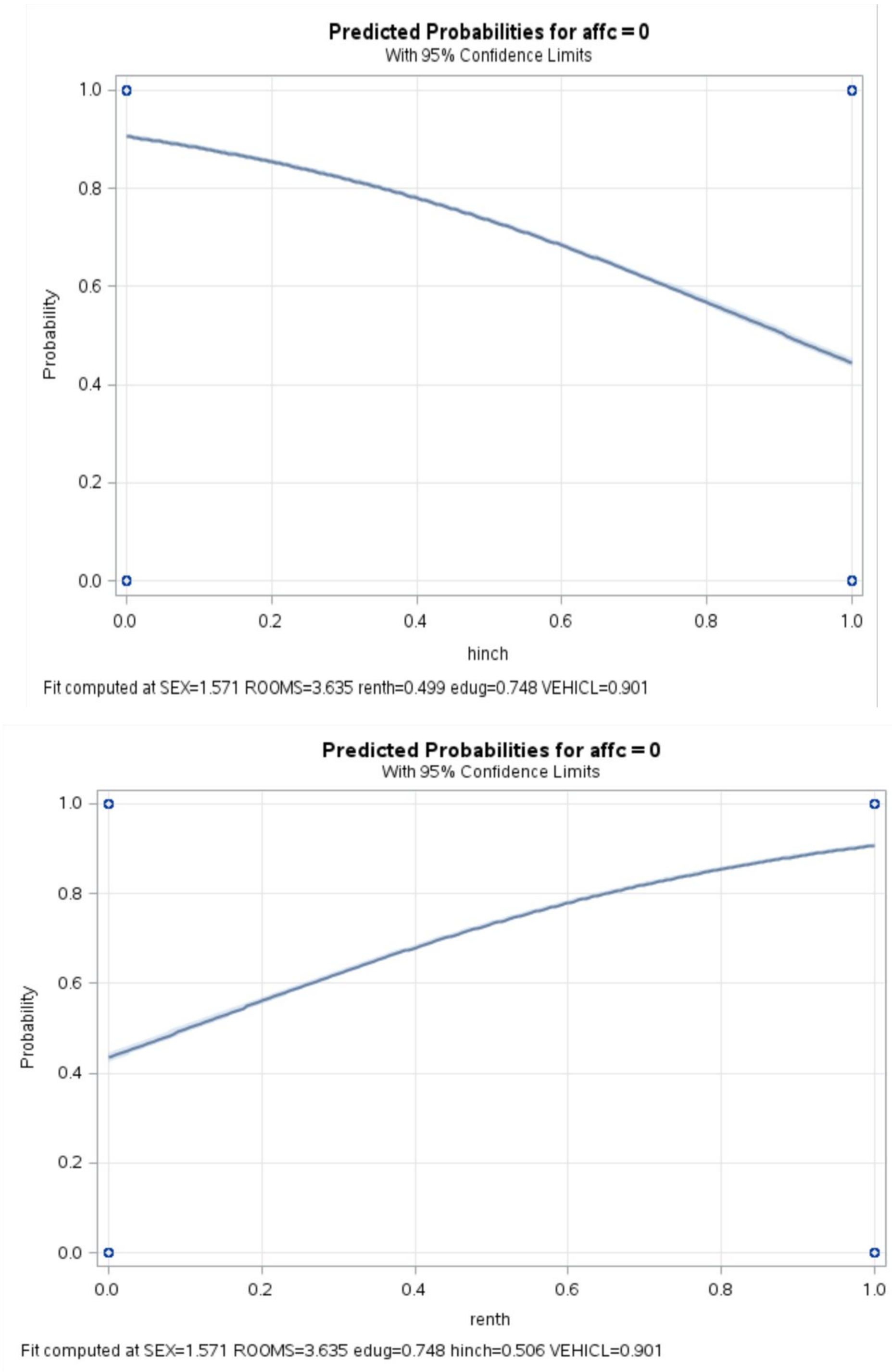


Figure 16. The relationship of predictor and response for low income rentals in NJ; (a) is of predictor hinch and (b) is of predictor renth

The relationship between response and some key predictors in Model 2 is described in Figure 16. The predicted variable is renth = 0/1(grent is less or greater than $710, $710 is the median of grent). As show in Figure 16.b, the probability of unaffordable units will be increase（affc=0) to about 1 as renth changes from below median to above median, indicating renth as a key predictor. Meanwhile, when hinch change from 0 to 1, the probability of affc = 0 will decrease from about 1 to below 0.5 as shown in Figure 16.a, implying a negative effect.

e. Model 3- Homeowners with High Income in NJ

The likelihood ratio chi-square of 44893.8111 with a p-value of 0.0001 indicates that Model 3 has a good fitting. Also, the p-values of all predictors are less than 0.0001.

The regression formula is as follows:

$$log(odds) = log(\frac{P}{(1-P)}) = -1.0181 - 0.6313*senior - 0.5206*rooms$$
$$+ 0.3257*sex - 0.1777*persons - 0.4815*edug - 2.6337*hinch + 2.5541 * morth - 0.1605*vehicl$$

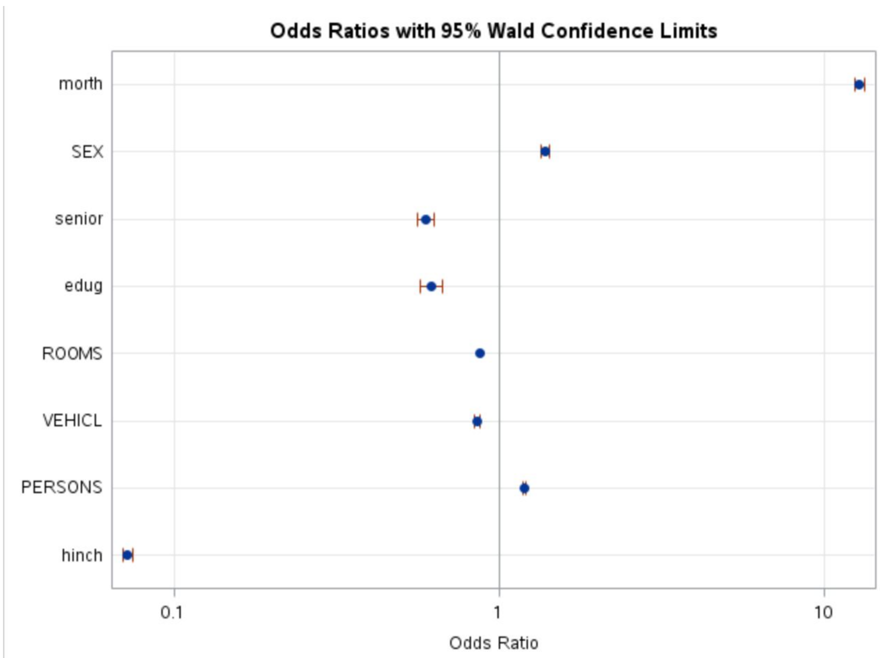
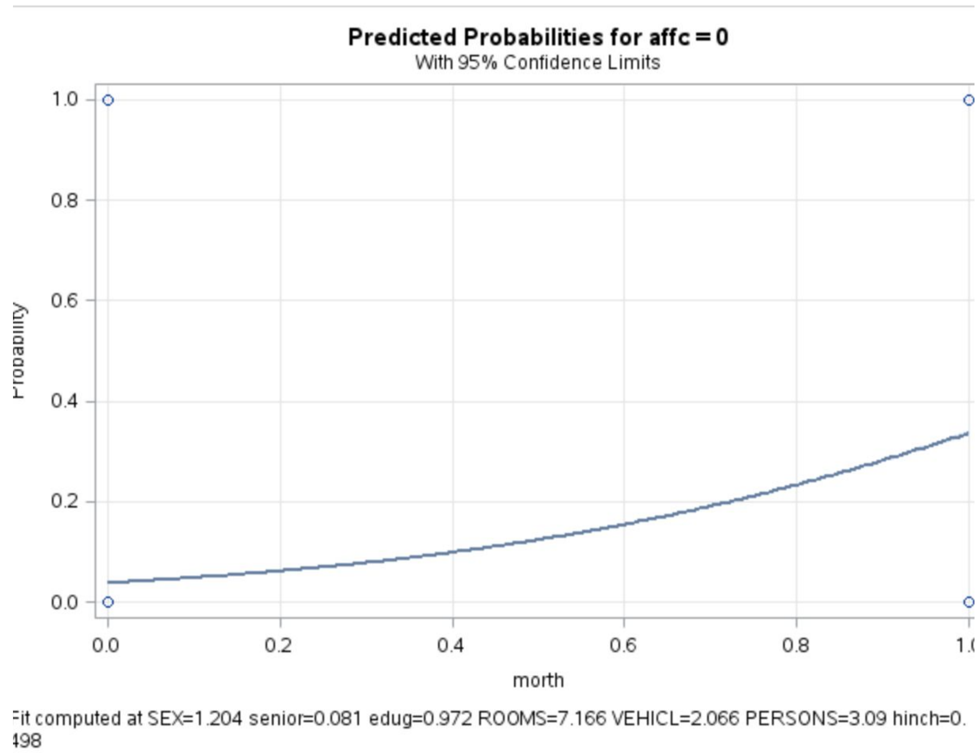


Figure 17. Odds ratios graph (Logbase=10) of affc for high income owners in NJ

This odds ratio graph in Figure 17 indicates that morth significantly affects the response variable, whose point estimate is over 10. As in Model 2 and 3, hinch has a strong negative effect on the response variable.



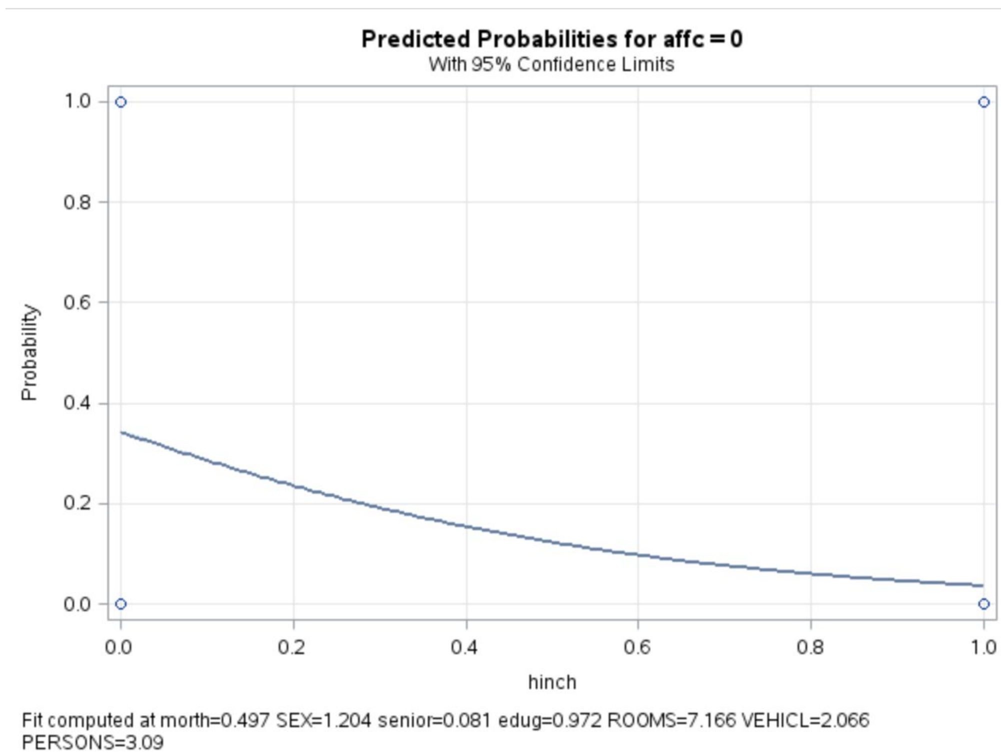


Figure 18 . The relationship of variable and response for high income owners in NJ; (a) is of predictor morth and (b) is of predictor hinch

The predicted variable is morth = 0/1(lmort: log(mortgage) is less than 14.57 or greater than 14.57 where 14.57 is the median lmort). Figure 18.a indicates that when mortgage paid is greater than the average mortgage, the probability of unaffordable units will significantly increase to about 1. With the increasing household income, the probability of affc=0 is decreasing, as shown in Figure 18.b.

f. Model 4- Homeowners with Low household Income in NJ

Using proc logistic in SAS, we found that the likelihood ratio chi-square is 17877.49 with p-value less than 0.0001. The p-values of likelihood estimate for all predictors are less than 0.0001. Therefore, Model 4 is a satisfactory model. It is shown as below:

$$log(odds) = log(\frac{P}{(1-P)}) = 0.1418 - 0.3876*senior + 0.0492*rooms$$
$$+ 0.3251*sex - 0.2780*edug - 2.8897*hinch + 4.3962 * morth - 0.1294*vehicl$$

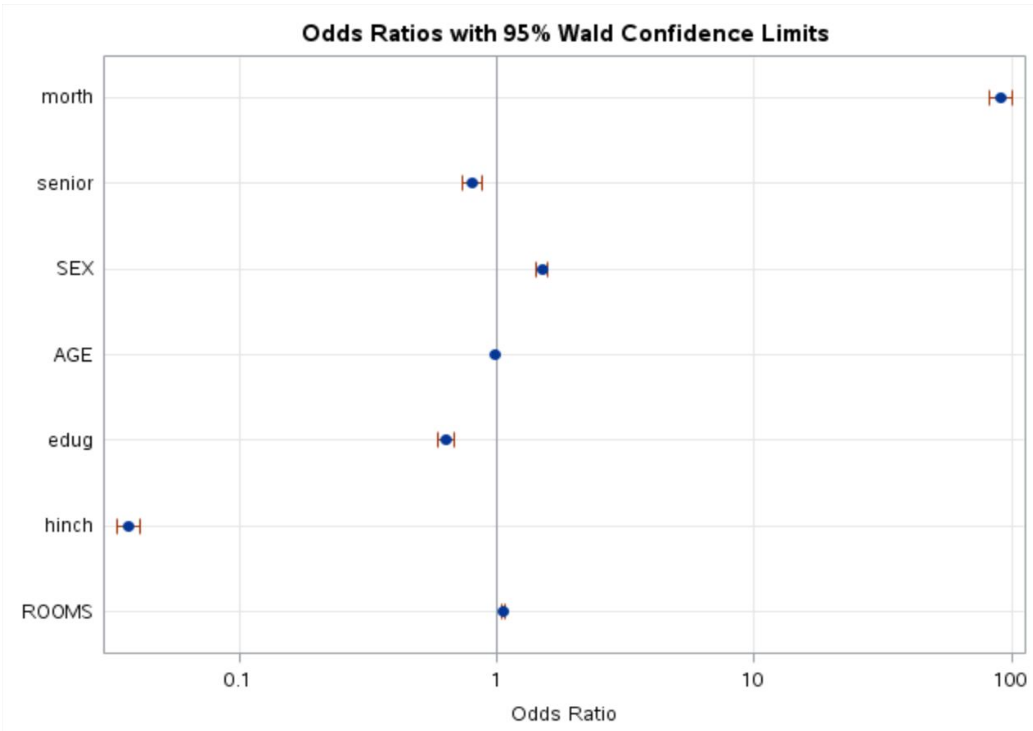


Figure 19. Odds ratios graph (Logbase=10) of affc for low income owner in NJ

The odds ratio graph in Figure 19 indicates that the most significant variable in the model is morth, compared to the other predictors.

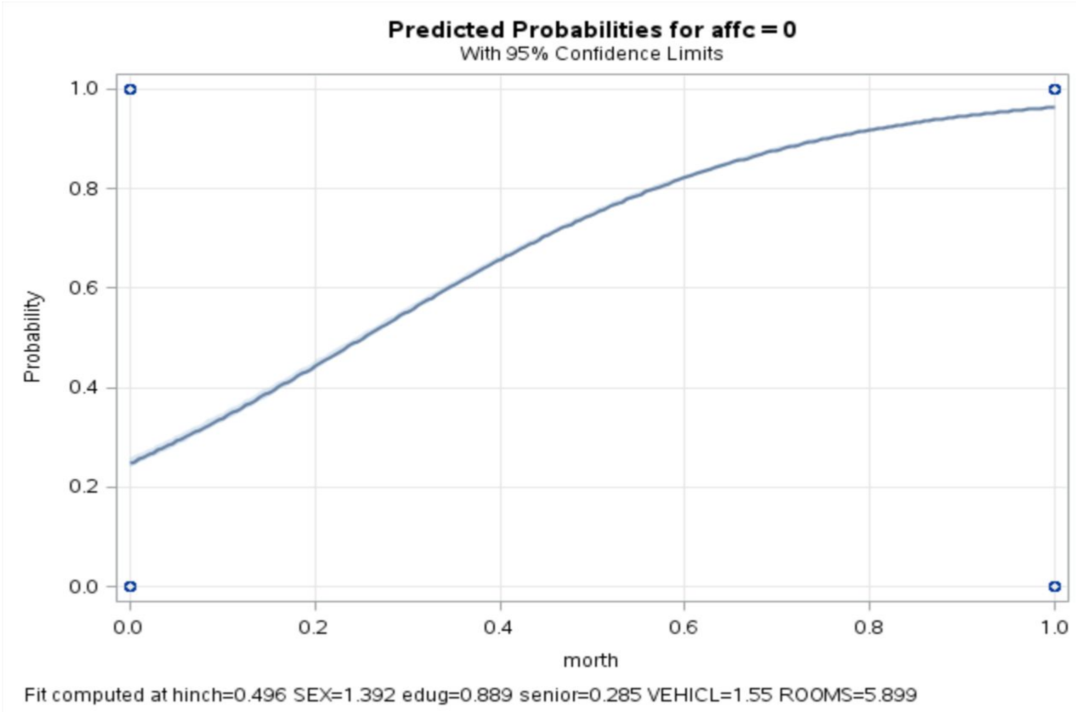


Figure 20. The relationship of morth predictor and response for low income owner in NJ

In Model 4, the response variable is morth = 0/1(lmort: log(mortgage) is less than 13.98 or greater than 13.98 where 13.98 is the median lmort). In Figure 19, morth is the predictor with the largest value. It indicates that when mortgage is higher than average mortgage, the probability of unaffordable units will significantly increase from 0.2 to about 1.

II. Model Fitting for AL Data

Just like the NJ data, all predictors are selected by using lasso in R and then the significant ones are chosen to fit the logistic model in SAS for AL data. The corresponding results are shown in Table 4.

Table 4. Logistic model fitting in SAS for AL data

predictor	Model 5 Rental,high income	Model 6 Rental,low income	Model 7 Owner,high income	Model 8 Owner,low income
hinch	-3.1051 (p <.0001)	-1.4428 (p<.0001)	-3.0388 (p<.0001)	-0.7123 (p<0.0001)
sex	0.1578 (p <.0001)	N/A	0.3488 (p<.0001)	0.2557 (p<0.0001)
rooms	-0.1114 (p <.0001)	-0.8192 (p<.0001)	-0.1017 (p<.0001)	-0.0598 (p<0.0001)

edug	0.4004 (p <.0001)	-0.7089 (p<.0001)	-0.1633 (p<.0001)	N/A
vehi	0.2151 (p <.0001)	-0.3653 (p<.0001)	-0.1047 (p<0.0001)	N/A
renth	3.1059 (p <.0001)	.9960 (p<.0001)	N/A	N/A
morth	N/A	N/A	2.8850 (p<.0001)	3.9844 (p<0.0001)
persons	N/A	.9300 (p<.0001)	0.1833 (p<.0001)	0.1567 (p<.0001)
bldgsz	N/A	N/A	N/A	N/A
senior	N/A	.3971 (p<.0001)	0.1481 (p<.0001)	-0.5605 (p<0.0001)

III. Model Fitting for FL Data

Similar to the model development procedure for NJ and AL data, the logistic models for affc are developed using FL data after selecting the predictors by lasso in R. The corresponding results are shown in Table 5.

Table 5. Logistic model fitting in SAS for FL data

predictor	Model 9 Rental,high income	Model 10 Rental,low income	Model 11 Owner,high income	Model 12 Owner,low income
hinch	-1.8516 (p<.001)	-1.6865 (p<.0001)	-2.7404 (p<.0001)	-2.3435 (p<.0001)
sex	-0.1205 (p<.0001)	N/A	0.2716 (p<.0001)	0.3184 (p<.0001)
rooms	-0.1947 (p<.0001)	-0.6124 (p<.0001)	0.00449 (p<.0001)	-0.1982 (p<.0001)
edug	0.1140 (p<.0001)	-0.2519 (p<.0001)	N/A	-0.3179 (p<.0001)
vehi	0.1603	-0.2217	0.0375	-0.1442

	(p<.0001)	(p<.0001)	(p<.0001)	(p<0.0001)
renth	2.7105 (p<.0001)	1.6632 (p<.0001)	N/A	N/A
morth	N/A	N/A	4.8046 (p<.0001)	2.7171 (p<0.0001)
persons	N/A	0.7985 (p<.0001)	N/A	0.2982 (p<0.0001)
bldgsz	N/A	N/A	N/A	N/A
senior	N/A	1.2101 (p<.0001)	-0.4702 (p<.0001)	-0.1704 (p<0.0001)

IV. NJ, AL and FL Comparison

Across the three states, there is much difference between ownership and rentals, especially when we take into account the levels of income.

1. An important predictor in both NJ and FL is the rent in terms of renters with high income. The odds of unaffordable and crowd units increase with increasing average rent. This variable plays an even bigger role in AL, where renters with higher income have a much bigger negative impact on affordable and uncrowded housing.
2. For renters with low income, gender has an influence in affordable and uncrowded units. It seems that for females, it will be more difficult to afford uncrowded units in Alabama, but in New Jersey and Florida, females find it easier to rent an affordable and crowd unit.
3. For owners regardless of household income, an important variable of increasing influence on unaffordable and crowded units will always be morth which relates to mortgage. An increasing influence of affordable and crowd units is still hinch variable across the three states, which means that when household income grows, units become more affordable and uncrowded. The only difference is the degree of influence (small or large) household income has on each state.
4. Below is a table of the frequencies of affordable and uncrowded units across the states, taking into account ownership status and income levels. Looking at this data on a state level, the poorest state (AL), has the highest ratios (affc =1/affc =0) regardless of proprietorship status and income level. This indicates that in general, AL has a lot of affordable and uncrowded units. The richest state, NJ, has the least amount of affordable and uncrowded (affc) units. Breaking down Table 6 by income level, people with low income have higher ratios of affc units - this means they are much more likely to live in affc units because there affc units are available and/or readily accessible.

Table 6. Affc frequency in NJ, AL and FL

state	affc	Rental High in	Rental Low in	Owner High in	Owner Low in	Total
NJ	0	1022	320	1378	1830	4550
	1	478	1158	708	7788	10132
	ratio	0.47	**3.62**	0.51	**4.26**	2.23
AL	0	901	132	2131	1347	4511
	1	679	1096	1870	10050	13695
	ratio	0.75	**8.30**	0.88	**7.46**	3.04
FL	0	1022	320	6130	7007	14479
	1	478	1158	4147	34151	39934
	ratio	0.47	3.62	0.68	4.87	2.76

V. Evaluate Seniors as a Predictor in the Logistic Model Development for Affordable Housing in NJ, AL and FL

Table 7. The coefficients of senior in all logistic models in this project

Data set	NJ	AL	FL
High household income rentals	1.4259 (p<.0001)	N/A	N/A
Low household income rentals	N/A	0.3971 (p<.0001)	1.2101 (p<.0001)
High household income owners	-0.6313 (p<.0001)	0.1481 (p<.0001)	-0.4702 (p<.0001)
Low household income owners	-0.3876 (p<.0001)	-0.5605 (p<.0001)	-0.1704 (p<.0001)

Based on the coefficients of seniors in all the fitted models, we can get the following conclusion.

1. Senior is a significant predictor for owner housing in all three states. The negative coefficients indicate that the larger senior frequency is, the more affordable housing becomes in all three states.
2. For renters with high income, the probability of response is related to seniors in NJ; it indicates that as the renter turns 65, the probability of unaffordable and crowd the units increases in NJ. Therefore, NJ may not be a good option for the geriatric. In FL and AL, age is not a significant factor for renters with a high income. In general, seniors who are owners are more likely to have affordable and uncrowded units except in Alabama where having a high income negatively affects the odds of affc units.

Conclusion

Based on the data analysis, we made the following conclusions:

1. All three states have similar income distribution and household distribution. For rent distribution, Alabama has the lowest rent value and New Jersey has the highest overall rent value.
2. In regard to basic statistics of the three states, New Jersey has the highest Median income and Alabama has the lowest one. Florida has the lowest proportion of low-income families whereas Alabama has the highest proportion. Furthermore, both New Jersey and Florida have similar proportion of low-Income units that are affordable and uncrowded but the state of Alabama has the highest proportion among three states. Finally Alabama has the highest proportion of units inhabited by low income families that are affordable and uncrowded; meanwhile New Jersey has the lowest proportion.
3. Through developing logistic models for affc, it is found that several predictors are very important, including renth for renters, morth for owners and hinch for both renters and owners. Besides those similarities, different various predictors are selected in different models, controlling for household income and ownership status in different states.

Appendix

Part I:

Table A.1 Basic statistic summary for lhinc, grent, and valuen

	lhinc			grent			valuen		
state	1	12	34	1	12	34	1	12	34
N	82094	306834	149590	18817	83843	46512	63277	222991	103078
Mean	10.32	10.50	10.83	4.56E+02	6.72E+02	7.85E+02	9.64E+04	1.26E+05	2.10E+05
Std Deviation	1.01	0.95	0.97	2.46E+02	3.35E+02	3.81E+02	9.92E+04	1.27E+05	1.53E+05
Skewness	-1.00	-0.77	-0.89	1.82	1.83	1.45	4.31	3.78	2.45
Uncorrected SS	8.82E+06	3.41E+07	1.77E+07	5.05E+09	4.73E+10	3.54E+10	1.21E+15	7.16E+15	6.93E+15
Coeff Variation	9.80	9.08	8.95	53.92	49.92	48.59	102.91	100.29	72.91
Sum Weights	8.21E+04	3.07E+05	1.50E+05	1.88E+04	8.38E+04	4.65E+04	6.33E+04	2.23E+05	1.03E+05
Sum Observations	8.47E+05	3.22E+06	1.62E+06	8.58E+06	5.63E+07	3.65E+07	6.10E+09	2.82E+10	2.16E+10
Variance	1.02	0.91	0.94	6.05E+04	1.12E+05	1.45E+05	9.83E+09	1.61E+10	2.33E+10
Kurtosis	3.28	3.26	2.91	8.52	6.47	4.41	28.76	18.97	7.89
Corrected SS	8.38E+04	2.79E+05	1.40E+05	1.14E+09	9.43E+09	6.76E+09	6.22E+14	3.59E+15	2.41E+15
Std Error Mean	0.00	0.00	0.00	1.79	1.16	1.77	394.21	268.65	475.79
Mean	10.32	10.50	10.83	456	672	785	9.64E+04	1.26E+05	2.10E+05
Median	10.45	10.58	10.94	431	630	750	7.50E+04	9.50E+04	1.62E+05
Mode	10.31	10.31	10.60	400	500	700	1.12E+05	1.12E+05	1.37E+05
Std Deviation	1.01	0.95	0.97	246	335	381	9.92E+04	1.27E+05	1.53E+05
Variance	1.02	0.91	0.94	60490	112442	145374	9.83E+09	1.61E+10	2.33E+10
Range	13.70	13.87	12.53	2484	3113	3196	9.95E+05	9.95E+05	9.95E+05
Interquartile Range	1.21	1.11	1.14	271	328	350	6.75E+04	7.25E+04	1.13E+05

Part II: Summary statistics for affordable house in NJ, AL, and FL

Table A.2.1 NJ

Obs	PUMA1	sdenom	sdenom2	snum	senum	scnum	slinc	stot	rat	rat2	rat3
1	34011	32369	8284	1955	2163	7850	100171	251323	0.06	0.24	0.40
2	34012	21494	5411	1076	1231	5081	89505	224990	0.05	0.20	0.40
3	34020	33584	9204	3296	3341	9043	78432	195646	0.10	0.36	0.40
4	34030	31077	8640	2441	2597	8347	89386	219297	0.08	0.28	0.41
5	34041	20483	5497	1427	1610	4969	76465	200330	0.07	0.26	0.38
6	34042	23902	5770	1821	2090	5303	71021	182146	0.08	0.32	0.39
7	34050	30115	7150	2294	2500	6845	57851	150434	0.08	0.32	0.38
8	34060	13966	6141	1603	1824	5450	115616	274655	0.11	0.26	0.42
9	34070	14070	3584	528	672	3103	82258	222219	0.04	0.15	0.37
10	34080	7338	3300	869	1095	2843	73451	181307	0.12	0.26	0.41
11	34090	33254	9081	3326	3386	8944	96992	254569	0.10	0.37	0.38
12	34101	5739	1918	365	440	1697	61938	160800	0.06	0.19	0.39
13	34102	12737	3769	899	1021	3391	65535	162536	0.07	0.24	0.40
14	34110	8833	2783	502	749	2363	64798	160543	0.06	0.18	0.40
15	34120	25565	6094	2046	2099	5974	55894	150967	0.08	0.34	0.37

Table A.2.2 AL

Obs	PUMA1	sdenom	sdenom2	snum	senum	scnum	slinc	stot	rat	rat2	rat3
1	1100	67696	20063	9283	9717	19139	98918	254062	0.14	0.46	0.39
2	1200	38780	12446	5551	5864	11959	75351	185001	0.14	0.45	0.41
3	1300	50947	18106	6216	6911	16843	88901	210080	0.12	0.34	0.42
4	1400	52464	18618	5894	6467	17581	88505	206914	0.11	0.32	0.43
5	1500	39242	11537	4271	4697	10743	63322	166826	0.11	0.37	0.38
6	1600	60432	19791	7746	8405	18386	104015	253944	0.13	0.39	0.41
7	1701	51050	14848	6896	7144	14387	61583	157725	0.14	0.46	0.39

8	1702	27962	7321	2311	2477	6919	87606	213755	0.08	0.32	0.41

Table A.2.3 FL

Obs	PUMA1	sdenom	sdenom2	snum	senum	scnum	slinc	stot	rat	rat2	rat3
1	12010	37921	10358	3500	3697	9913	57972	149526	0.09	0.34	0.39
2	12020	43419	12629	5108	5349	12240	62537	165328	0.12	0.40	0.38
3	12030	42853	15318	4694	5115	14387	63913	147565	0.11	0.31	0.43
4	12040	45269	16105	5004	5437	15080	68720	162954	0.11	0.31	0.42
5	12051	54824	13005	3502	3969	12057	81026	216451	0.06	0.27	0.37
6	12052	48388	12817	4583	4861	12127	74766	197348	0.09	0.36	0.38
7	12060	49455	13254	4480	4650	12545	83589	228012	0.09	0.34	0.37
8	12070	56779	14582	5941	6133	13948	61444	174962	0.10	0.41	0.35
9	12081	46124	10883	4458	4556	10624	72900	198349	0.10	0.41	0.37
10	12082	27798	6377	1621	1736	6204	73661	193735	0.06	0.25	0.38
11	12083	12624	3175	1073	1120	3075	82914	210040	0.09	0.34	0.39
12	12084	34704	8307	1784	2045	7406	83181	216290	0.05	0.21	0.38
13	12085	48161	13394	3871	4398	12235	62814	162707	0.08	0.29	0.39
14	12091	65702	18113	6506	7079	16981	79472	222538	0.10	0.36	0.36
15	12092	57578	17409	3279	3813	15763	80874	220818	0.06	0.19	0.37
16	12093	62663	17695	3705	4387	15269	59212	168020	0.06	0.21	0.35
17	12100	45160	12268	4367	4657	11617	73214	193072	0.10	0.36	0.38
18	12110	43038	11764	4458	4853	11195	66593	181679	0.10	0.38	0.37
19	12120	48779	14287	4892	5291	13555	69982	188752	0.10	0.34	0.37
20	12130	54781	14742	4994	5919	13003	60111	168333	0.09	0.34	0.36
21	12140	54266	13193	4216	4488	12395	93323	254745	0.08	0.32	0.37
22	12150	65483	16505	5926	6393	15653	86751	245725	0.09	0.36	0.35

23	12161	33566	8442	2057	2447	7401	80295	214109	0.06	0.24	0.38
24	12162	71716	18771	5620	6067	17555	94800	247185	0.08	0.30	0.38
25	12171	36495	8961	1234	1380	8153	71794	181189	0.03	0.14	0.40
26	12172	24724	6888	1208	1453	6109	84993	224881	0.05	0.18	0.38
27	12173	66103	15299	2725	2946	13585	90642	227384	0.04	0.18	0.40
28	12181	27870	7348	496	822	5316	48965	140436	0.02	0.07	0.35
29	12182	8222	3138	501	643	2278	63204	169695	0.06	0.16	0.37
30	12183	20952	6312	743	1075	4192	70325	174599	0.04	0.12	0.40
31	12184	17930	6262	737	1050	4713	58124	150191	0.04	0.12	0.39
32	12185	37714	12567	2258	3322	9593	54057	142992	0.06	0.18	0.38

Part III

a. SAS code

```
/* Part I: State code: 34 - NJ; 1 - als; 12 - flo; */
/*compare household income;*/

data state; set sasuser.G7ALAFLO sasuser.Gnj;
run;
proc sort data=state;
by state; run;
proc univariate freq data=state;
var lhinc;
by state;
output out=a  PCTLPRE=p
PCTLPTS=1 2 3 4 5 7 10 20 30 40 50 60 70 80 90 93 95 96 97 98 99;
run;
proc transpose data=a out=b;
run;
data b; set b ; if _N_> 1;run;quit;

/* Use plot on dataset b to compare the income distributions. */
/* 1. Percentile plots*/;
proc sgplot data=b;
series x=_NAME_ y=col1;
series x=_NAME_ y=col2;
series x=_NAME_ y=col3;
XAXIS LABEL = 'Percentiles';
YAXIS LABEL = 'lhinc';
run;
quit;

/* 2. Percentile-percentile plots*/;
proc sgplot data=b;
series x=col1 y=col2 / legendlabel='1vs12';
series x=col1 y=col3 / legendlabel='1vs34';
series x=col2 y=col3 / legendlabel='12vs34';
XAXIS LABEL = 'log household income at 1-99% Percentiles';
YAXIS LABEL = 'log household income at 1-99% Percentiles';
run;
quit;

/*compare rent and housing values need to repeat the same code but only rentals or owner occupied;*/
/*compare rentals;*/
data state2; set state;
if smi = 0;  /* for rentals  */
        /*  for owners use if smi = 1;  */
run;
proc sort data=state2;
by state; run;
proc univariate freq data=state2;
var grent;
```

```
by state;
output out=a  PCTLPRE=p
PCTLPTS=1 2 3 4 5 7 10 20 30 40 50 60 70 80 90 93 95 96 97 98 99;
run;
proc transpose data=a out=b;
run;
data b; set b ; if _N_> 1;run;

/* Use lineplot on datasetb to compare the distributions. */
/* 1. Percentile plots*/;
proc sgplot data=b;
series x=_NAME_ y=col1;
series x=_NAME_ y=col2;
series x=_NAME_ y=col3;
XAXIS LABEL = 'Percentiles';
YAXIS LABEL = 'grent (rent per month)';
run;
quit;

/* 2. Percentile-percentile plots*/;
proc sgplot data=b;
series x=col1 y=col2 / legendlabel='1vs12';
series x=col1 y=col3 / legendlabel='1vs34';
series x=col2 y=col3 / legendlabel='12vs34';
XAXIS LABEL = 'grent distribution at 1-99% percentiles';
YAXIS LABEL = 'grent distribution at 1-99% percentiles';
run;
quit;

/*compare housing value by owners;*/
/*transform the housing value into numerical data*/
data sasuser.val;
input value r1 r2;
valuen = (r1+r2)/2.0;
keep value valuen;
cards;
01 0 10000
02 10000 14999
03 15000 19999
04 20000 24999
05 25000 29999
06 30000 34999
07 35000 39999
08 40000 49999
09 50000 59999
10 60000 69999
11 70000 79999
12 80000 89999
13 90000 99999
14 100000 124999
```

```
15 125000 149999
16 150000 174999
17 175000 199999
18 200000 249999
19 250000 299999
20 300000 399999
21 400000 499999
22 500000 749999
23 750000 999999
24 1000000 1000000
;
run;
proc sort data=sasuser.val;
by value;
run;
data state3; set state;
if smi = 1;  /* for rentals  */
           /*  for owners use if smi = 1;  */
run;
proc sort data=state3;
by value;
data state4; merge state3 sasuser.val;
by value;
run;
proc sort data=state4;
by state; run;
proc univariate freq data=state4;
var valuen;
by state;
output out=a  PCTLPRE=p
PCTLPTS=1 2 3 4 5 7 10 20 30 40 50 60 70 80 90 93 95 96 97 98 99;
run;
proc transpose data=a out=b;
run;
data b; set b; if _N_> 1;run;

/* Use lineplot on datasetb to compare valuen distributions. */
/* 1. Percentile plots*/;
proc sgplot data=b;
series x=_NAME_ y=col1;
series x=_NAME_ y=col2;
series x=_NAME_ y=col3;
XAXIS LABEL = 'Percentiles';
YAXIS LABEL = 'valuen(transformed house value)';
run;
quit;

/* 2. Percentile-percentile plots*/;
proc sgplot data=b;
series x=col1 y=col2 / legendlabel='1vs12';
```

```
series x=col1 y=col3 / legendlabel='1vs34';
series x=col2 y=col3 / legendlabel='12vs34';
XAXIS LABEL = 'valuen at 1-99% percentiles';
YAXIS LABEL = 'valuen at 1-99% percentiles';
run;
quit;


/*Part II: *compare state basic statistics;*/
/*a. median income
  b. proportion of low income families, (rat3)
  c. Among the new units, what is the proportion of units that are low income, and are affordable and are not crowded.  (rat)
  d. -      Among all new units that are occupied by low income families, what is the proportion of units that are affordable and are not crowded. (rat2);*/
/*a;*/
data state; set sasuser.G7ALAFLO sasuser.Gnj;
run;

proc means data = state median;
var finc hinc;
class state;
run;
quit;

/*b;*/
/*1.do basic data analysis for NJ*/
/*filter NJ data with household 1,2 and 3 families, do basic statistics*/
data stateNJ; set state;
if state=34;
run;
proc sort data=stateNJ; by puma1;
data famincNJ; set stateNJ;
if hht=1 or hht=2 or hht=3;
run;

proc means median data=famincNJ noprint;
 var finc;
 weight hweight;
 by puma1;
 output out=c median=med;
run;

data c;
  set c;
  keep puma1 med;
  run;
proc sort data=stateNJ;
by puma1;
run;
```

```
data stateNJ;
  merge stateNJ c;
  by puma1;

/*add proper new variables into the data set and filter the data again*/

data stateNJ; set stateNJ;
m4 = med*0.8;
/* if m4 > 50200 then m4 = 50200; */
hlinc = 0;
 units=1;
 if persons <= 4 and hinc < m4*(1-(4-persons)*0.1) then hlinc = 1;
 if persons > 4 and hinc < m4*(1+(persons-4)*0.08) then hlinc = 1;
/***************
* This calculates the percentage of households
* below the 80pct of median
******************/
/* Crowded */
uncrowd = 0;
if persons <= rooms then uncrowd = 1;
/* new construction */
new = 0;
if yrbuilt=1 or yrbuilt=2 or yrbuilt=3 then new =1;
/* Affordable */
Aff = 0;
if smocapi > 0 then grapi=.;
if grapi > 0 then smocapi=.;
if smocapi ne . and  smocapi <= 30 then Aff=1;
if grapi ne . and grapi <= 30 then Aff=1;
xnum = Aff*uncrowd*new*hlinc;
enum = Aff*new*hlinc;
cnum = uncrowd*new*hlinc;
denom = new;
denom2 = new*hlinc;
tot = 1;
run;

proc means data=stateNJ noprint;
var denom denom2 xnum enum cnum hlinc tot;
by puma1;
weight hweight;
output out = res sum= sdenom sdenom2 snum senum scnum slinc stot;
run;

data res; set res;
rat = snum/sdenom;
rat2 = snum/sdenom2;
rat3 = slinc/stot;
state = 'NJ';
keep puma1 state snum senum scnum sdenom sdenom2 slinc stot rat rat2 rat3;
```

```
run;
proc print;
run;


/*2.do basic data analysis for Ala*/
data stateALA; set state;
if state=1;
run;
proc sort data=stateALA; by puma1;
data famincALA; set stateALA;
if hht=1 or hht=2 or hht=3;
run;

proc means median data=famincALA noprint;
 var finc;
 weight hweight;
 by puma1;
 output out=d median=med;
run;

data d;
  set d;
  keep puma1 med;
  run;
proc sort data=stateALA;
by puma1;
run;

data stateALA;
  merge stateALA d;
  by puma1;
/*add proper new variables into the data set and filter the data again*/

data stateALA; set stateALA;
m4 = med*0.8;
/* if m4 > 50200 then m4 = 50200; */
hlinc = 0;
 units=1;
 if persons <= 4 and hinc < m4*(1-(4-persons)*0.1) then hlinc = 1;
 if persons > 4 and hinc < m4*(1+(persons-4)*0.08) then hlinc = 1;
/***************
* This calculates the percentage of households
* below the 80pct of median
*****************/
/* Crowded */
uncrowd = 0;
if persons <= rooms then uncrowd = 1;
/* new construction */
new = 0;
```

```
if yrbuilt=1 or yrbuilt=2 or yrbuilt=3 then new =1;
/* Affordable */
Aff = 0;
if smocapi > 0 then grapi=.;
if grapi > 0 then smocapi=.;
if smocapi ne . and  smocapi <= 30 then Aff=1;
if grapi ne . and grapi <= 30 then Aff=1;
xnum = Aff*uncrowd*new*hlinc;
enum = Aff*new*hlinc;
cnum = uncrowd*new*hlinc;
denom = new;
denom2 = new*hlinc;
tot = 1;
run;

proc means data=stateALA noprint;
var denom denom2 xnum enum cnum hlinc tot;
by puma1;
weight hweight;
output out = resA sum= sdenom sdenom2 snum senum scnum slinc stot;
run;

data resA; set resA;
rat = snum/sdenom;
rat2 = snum/sdenom2;
rat3 = slinc/stot;
state = 'ALA';
keep puma1 state snum senum scnum sdenom sdenom2 slinc stot rat rat2 rat3;
run;
proc print;
run;


/*3.do basic data analysis for flo*/
data stateFLO; set state;
if state=12;
run;
proc sort data=stateFLO; by puma1;
data famincFLO; set stateFLO;
if hht=1 or hht=2 or hht=3;
run;

proc means median data=famincFLO noprint;
 var finc;
 weight hweight;
 by puma1;
 output out=e median=med;
run;

data e;
```

```
  set e;
  keep puma1 med;
  run;
proc sort data=stateFLO;
by puma1;
run;

data stateFLO;
  merge stateFLO e;
  by puma1;
/*add proper new variables into the data set and filter the data again*/

data stateFLO; set stateFLO;
m4 = med*0.8;
/*hlinc = 1, lowincome household; hlinc = 0, not lowincome household*/
hlinc = 0;
 units=1;
 if persons <= 4 and hinc < m4*(1-(4-persons)*0.1) then hlinc = 1;
 if persons > 4 and hinc < m4*(1+(persons-4)*0.08) then hlinc = 1;
/* Crowded */
uncrowd = 0;
if persons <= rooms then uncrowd = 1;
/* new construction */
new = 0;
if yrbuilt=1 or yrbuilt=2 or yrbuilt=3 then new =1;
/* Affordable */
Aff = 0;
if smocapi > 0 then grapi=.;
if grapi > 0 then smocapi=.;
if smocapi ne . and  smocapi <= 30 then Aff=1;
if grapi ne . and grapi <= 30 then Aff=1;
xnum = Aff*uncrowd*new*hlinc;
enum = Aff*new*hlinc;
cnum = uncrowd*new*hlinc;
denom = new;
denom2 = new*hlinc;
tot = 1;
run;

proc means data=stateFLO noprint;
var denom denom2 xnum enum cnum hlinc tot;
by puma1;
weight hweight;
output out = resF sum= sdenom sdenom2 snum senum scnum slinc stot;
run;

data resF; set resF;
rat = snum/sdenom;
rat2 = snum/sdenom2;
rat3 = slinc/stot;
```

```
state = 'FLO';
keep puma1 state snum senum scnum sdenom sdenom2 slinc stot rat rat2 rat3;
run;
proc print;
run;
data ress; set res resa resf;
merge res resa resf;
keep rat rat2 rat3 state;
by state;
run;quit;
/*one-way ANOVA test for the rat data*/
proc glm data=ress plots=diagnostics;
class state;
model rat=state;
means state / hovtest=levene;
title"one-way ANOVA with state as predictor";
run;
quit;
proc glm data=ress plots=diagnostics;
class state;
model rat2=state;
means state / hovtest=levene;
title"one-way ANOVA with state as predictor";
run;
quit;
proc glm data=ress plots=diagnostics;
class state;
model rat3=state;
means state / hovtest=levene;
title"one-way ANOVA with state as predictor";
run;
quit;

/*two-way ANOVA using post-hot tests to identify which pair of groups are significant different*/
ods graphics;
ods select lsmeans diff diffplot controlplot;
proc glm data = ress
        plots(only)= (diffplot(center) controlplot);
class state;
model rat = state;
lsmeans state / pdiff=all
             adjust=tukey;
title "Post-Hoc (multiple comparison test)Analysis of ANOVA - state as predictor";
run;
proc glm data = ress
       plots(only)= (diffplot(center) controlplot);
class state;
model rat2 = state;
lsmeans state / pdiff=all
             adjust=tukey;
```

```
run;
proc glm data = ress
        plots(only)= (diffplot(center) controlplot);
class state;
model rat3 = state;
lsmeans state / pdiff=all
              adjust=tukey;
run;quit;


/*Part III: We will find out a model for predicting the proportion of low income housing
that is affordable and uncrowded controlling for other variables and which
variables are important predictors.*/
/* state = 34  for NJ */
data state;
set sasuser.g7alaflo sasuser.gnj;
run;

proc means data = state median;
var finc hinc;
class state;
run;
quit;

/* get rental table in NJ */
data stateNJ; set state; if state=34; run;
proc sort data=stateNJ; by puma1;
data famincNJ; set stateNJ;
if hht=1 or hht=2 or hht=3;
run;
proc means median data=famincNJ noprint;
 var finc;
 weight hweight;
 by puma1;
 output out=c median=med;
run;

data c;
  set c;
  keep puma1 med;
  run;
proc sort data=stateNJ;
by puma1;
run;

data stateNJ;
  merge stateNJ c;
  by puma1;

data stateNJ; set stateNJ;
```

```
m4 = med*0.8;
hlinc = 0;
 units=1;
 if persons <= 4 and hinc < m4*(1-(4-persons)*0.1) then hlinc = 1;
 if persons > 4 and hinc < m4*(1+(persons-4)*0.08) then hlinc = 1;
/* Crowded */
uncrowd = 0;
if persons <= rooms then uncrowd = 1;
/* new construction */
new = 0;
if yrbuilt=1 or yrbuilt=2 or yrbuilt=3 then new =1;
/* Affordable */
Aff = 0;
if smocapi > 0 then grapi=.;
if grapi > 0 then smocapi=.;
if smocapi ne . and  smocapi <= 30 then Aff=1;
if grapi ne . and grapi <= 30 then Aff=1;

/*********************************************************************************
**************************************************************************/
/* rental low income hlinc=1*/
data rentalNJ; set stateNJ;
if grapi>0 and grapi <= 99 and hlinc=1;
lhinc = log(1+hinc);
gra = log( (grapi+2)/(102-grapi));
edug = 0;
if educ >7 then edug = 1;
  if new = 1;
run;

/* affc = aff &uncrowd */
data rentalNJ; set rentalNJ;
affc=0;
if aff=1 and uncrowd = 1 then affc=1;
run;

/*compute median of variable*/
proc univariate data = rentalNJ;
var grent lhinc;
run;quit;

data rentalNJ1; set rentalNJ;
renth = 0; if grent > 710 then renth =1;
hinch = 0; if lhinc > 10.04 then hinch =1;
senior = 0; if age>=65 then senior = 1;
/*Aff = 0;if grapi <= 30 then Aff=1;*/
run;

 proc logistic data= rentalNJ1;
 weight hweight;
```

```
 model Affc = sex rooms renth EDUG hinch vehicl;
 run;

ods graphic on;
proc logistic data= rentalNJ1 plots(only)= oddsratio(logbase=10);;
 weight hweight;
 model Affc = sex rooms renth EDUG hinch vehicl;
 effectplot fit(x=renth);
 effectplot fit(x=hinch);
 run;
ods graphics off;

/**********************************************/
/*rental high income hlinc=0*/

data rentalNJ; set stateNJ;
if grapi>0 and grapi <= 99 and hlinc=0;
lhinc = log(1+hinc);
gra = log( (grapi+2)/(102-grapi));
edug = 0;
if educ >7 then edug = 1;
  if new = 1;
run;

data rentalNJ; set rentalNJ;
affc=0;
if aff=1 and uncrowd = 1 then affc=1;
run;

proc univariate data = rentalNJ;
var grent lhinc;
run;quit;

data rentalNJ1; set rentalNJ;
renth = 0; if grent > 1100 then renth =1;
hinch = 0; if lhinc > 11.23 then hinch =1;
senior = 0; if age>=65 then senior = 1;
Aff = 0;if grapi <= 30 then Aff=1;
run;


 proc logistic data= rentalNJ1;
 weight hweight;
 model Affc = senior renth EDUG sex
             rooms vehicl hinch;
 run;

ods graphic on;
proc logistic data= rentalNJ1 plots(only)= oddsratio(logbase=E);;
 weight hweight;
```

```
model Affc = senior renth EDUG sex
              rooms vehicl hinch;
 effectplot fit(x=renth);
 effectplot fit(x=senior);
 run;
ods graphics off;



/******************************************************************************
*******************************************************************************/
data state;
set sasuser.g7alaflo sasuser.gnj;
run;

data stateNJ; set state; if state=34; run;
proc sort data=stateNJ; by puma1;
data famincNJ; set stateNJ;
if hht=1 or hht=2 or hht=3;
run;
proc means median data=famincNJ noprint;
 var finc;
 weight hweight;
 by puma1;
 output out=c median=med;
run;

data c;
  set c;
  keep puma1 med;
  run;
proc sort data=stateNJ;
by puma1;
run;

data stateNJ;
  merge stateNJ c;
  by puma1;

data stateNJ; set stateNJ;
m4 = med*0.8;
/* if m4 > 50200 then m4 = 50200; */
hlinc = 0;
 units=1;
 if persons <= 4 and hinc < m4*(1-(4-persons)*0.1) then hlinc = 1;
 if persons > 4 and hinc < m4*(1+(persons-4)*0.08) then hlinc = 1;
/* Crowded */
uncrowd = 0;
if persons <= rooms then uncrowd = 1;
/* new construction */
```

```
new = 0;
if yrbuilt=1 or yrbuilt=2 or yrbuilt=3 then new =1;
/*Affordable*/
Aff = 0;
if smocapi > 0 then grapi=.;
if grapi > 0 then smocapi=.;
if smocapi ne . and  smocapi <= 30 then Aff=1;
if grapi ne . and grapi <= 30 then Aff=1;

/*******************************************************************************
************************************************************************/
/* owner low income*/
data ownerNJ; set stateNJ;
if smocapi>0 and smocapi <= 99 and hlinc=1;
lhinc = log(1+hinc);
sm = log( (smocapi+2)/(102-smocapi));
edug = 0;
if educ >7 then edug = 1;
  if new = 1;
run;

data ownerNJ;
set ownerNJ;
lmort = log(smocapi*hinc);
run;

data ownerNJ; set ownerNJ;
affc=0;
if aff=1 and uncrowd = 1 then affc=1;
run;

proc univariate data = ownerNJ;
var lmort lhinc;
run;quit;

data ownerNJ1; set ownerNJ;
morth = 0; if lmort > 13.98 then morth =1;
hinch = 0; if lhinc > 10.42 then hinch =1;
senior = 0; if age>=65 then senior = 1;
/*Aff = 0;if smocapi <= 30 then Aff=1;*/
run;


 proc logistic data= ownerNJ1;
 weight hweight;
 model Affc = morth hinch sex senior rooms vehi edug;
 run;

ods graphic on;
proc logistic data= ownerNJ1 plots(only)= oddsratio(logbase=10);;
```

```
 weight hweight;
 model Affc = morth hinch sex senior rooms vehi edug;
 effectplot fit(x=morth);
 run;
ods graphics off;

/**********************************************/
/*owner high income*/

data ownerNJ; set stateNJ;
if smocapi>0 and smocapi <= 99 and hlinc=0;
lhinc = log(1+hinc);
sm = log( (smocapi+2)/(102-smocapi));
edug = 0;
if educ >7 then edug = 1;
  if new = 1;
run;

data ownerNJ;
set ownerNJ;
lmort = log(smocapi*hinc);
run;

data ownerNJ; set ownerNJ;
affc=0;
if aff=1 and uncrowd = 1 then affc=1;
run;

proc univariate data = ownerNJ;
var lmort lhinc;
run;quit;



data ownerNJ1; set ownerNJ;
morth = 0; if lmort > 14.57 then morth =1;
hinch = 0; if lhinc > 11.51 then hinch =1;
senior = 0; if age>=65 then senior = 1;
Aff = 0;if smocapi <= 30 then Aff=1;
run;

 proc logistic data= ownerNJ1;
 weight hweight;
 model Affc = morth sex senior EDUG rooms vehicl persons hinch;
 run;

ods graphic on;
proc logistic data= ownerNJ1 plots(only)= oddsratio(logbase=10);;
 weight hweight;
 model Affc = morth sex senior EDUG rooms vehicl persons hinch;
```

```
 effectplot fit(x=morth);
 effectplot fit(x=hinch);
 run;
ods graphics off;
/**********************************************************************************************************
****************************************************/

/* state=1 for AL*/

data state;
set sasuser.g7alaflo sasuser.gnj;
run;

proc means data = state median;
var finc hinc;
class state;
run;
quit;

/* get rental table in AL */
data stateNJ; set state; if state=1; run;
proc sort data=stateNJ; by puma1;
data famincNJ; set stateNJ;
if hht=1 or hht=2 or hht=3;
run;
proc means median data=famincNJ noprint;
 var finc;
 weight hweight;
 by puma1;
 output out=c median=med;
run;

data c;
  set c;
  keep puma1 med;
  run;
proc sort data=stateNJ;
by puma1;
run;

data stateNJ;
  merge stateNJ c;
  by puma1;

data stateNJ; set stateNJ;
m4 = med*0.8;
hlinc = 0;
 units=1;
 if persons <= 4 and hinc < m4*(1-(4-persons)*0.1) then hlinc = 1;
 if persons > 4 and hinc < m4*(1+(persons-4)*0.08) then hlinc = 1;
```

```
/* Crowded */
uncrowd = 0;
if persons <= rooms then uncrowd = 1;
/* new construction */
new = 0;
if yrbuilt=1 or yrbuilt=2 or yrbuilt=3 then new =1;
/* Affordable */
Aff = 0;
if smocapi > 0 then grapi=.;
if grapi > 0 then smocapi=.;
if smocapi ne . and  smocapi <= 30 then Aff=1;
if grapi ne . and grapi <= 30 then Aff=1;

/*******************************************************************************************
********************************************************************************/
/* rental low income hlinc=1*/
data rentalNJ; set stateNJ;
if grapi>0 and grapi <= 99 and hlinc=1;
lhinc = log(1+hinc);
gra = log( (grapi+2)/(102-grapi));
edug = 0;
if educ >7 then edug = 1;
  if new = 1;
run;

/* affc = aff &uncrowd */
data rentalNJ; set rentalNJ;
affc=0;
if aff=1 and uncrowd = 1 then affc=1;
run;

/*compute median of variable*/
proc univariate data = rentalNJ;
var grent lhinc;
run;quit;

data rentalNJ1; set rentalNJ;
renth = 0; if grent > 400 then renth =1;
hinch = 0; if lhinc > 9.59 then hinch =1;
senior = 0; if age>=65 then senior = 1;
/*Aff = 0;if grapi <= 30 then Aff=1;*/
run;

 proc logistic data= rentalNJ1;
 weight hweight;
 model Affc = sex rooms renth EDUG hinch vehicl;
 run;

ods graphic on;
proc logistic data= rentalNJ1 plots(only)= oddsratio(logbase=10);;
```

```
 weight hweight;
 model Affc = sex rooms renth EDUG hinch vehicl;
 effectplot fit(x=renth);
 effectplot fit(x=hinch);
 run;
ods graphics off;

/**********************************************/
/*rental high income hlinc=0*/

data rentalNJ; set stateNJ;
if grapi>0 and grapi <= 99 and hlinc=0;
lhinc = log(1+hinc);
gra = log( (grapi+2)/(102-grapi));
edug = 0;
if educ >7 then edug = 1;
  if new = 1;
run;

data rentalNJ; set rentalNJ;
affc=0;
if aff=1 and uncrowd = 1 then affc=1;
run;

proc univariate data = rentalNJ;
var grent lhinc;
run;quit;

data rentalNJ1; set rentalNJ;
renth = 0; if grent > 615.5 then renth =1;
hinch = 0; if lhinc > 10.76 then hinch =1;
senior = 0; if age>=65 then senior = 1;
Aff = 0;if grapi <= 30 then Aff=1;
run;


 proc logistic data= rentalNJ1;
 weight hweight;
 model Affc = senior renth EDUG sex
              rooms vehicl hinch;
 run;

ods graphic on;
proc logistic data= rentalNJ1 plots(only)= oddsratio(logbase=E);;
 weight hweight;
model Affc = senior renth EDUG sex
              rooms vehicl hinch;
 effectplot fit(x=renth);
 effectplot fit(x=senior);
 run;
```

```
ods graphics off;



/**************************************************************************************
***************************************************************************************/

data state;
set sasuser.g7alaflo sasuser.gnj;
run;

data stateNJ; set state; if state=1; run;
proc sort data=stateNJ; by puma1;
data famincNJ; set stateNJ;
if hht=1 or hht=2 or hht=3;
run;
proc means median data=famincNJ noprint;
 var finc;
 weight hweight;
 by puma1;
 output out=c median=med;
run;

data c;
  set c;
  keep puma1 med;
  run;
proc sort data=stateNJ;
by puma1;
run;

data stateNJ;
  merge stateNJ c;
  by puma1;

data stateNJ; set stateNJ;
m4 = med*0.8;
/* if m4 > 50200 then m4 = 50200; */
hlinc = 0;
 units=1;
 if persons <= 4 and hinc < m4*(1-(4-persons)*0.1) then hlinc = 1;
 if persons > 4 and hinc < m4*(1+(persons-4)*0.08) then hlinc = 1;
/* Crowded */
uncrowd = 0;
if persons <= rooms then uncrowd = 1;
/* new construction */
new = 0;
if yrbuilt=1 or yrbuilt=2 or yrbuilt=3 then new =1;
/*Affordable*/
Aff = 0;
```

```
if smocapi > 0 then grapi=.;
if grapi > 0 then smocapi=.;
if smocapi ne . and  smocapi <= 30 then Aff=1;
if grapi ne . and grapi <= 30 then Aff=1;

/*********************************************************************************
*************************************************************************/
/* owner low income*/
data ownerNJ; set stateNJ;
if smocapi>0 and smocapi <= 99 and hlinc=1;
lhinc = log(1+hinc);
sm = log( (smocapi+2)/(102-smocapi));
edug = 0;
if educ >7 then edug = 1;
  if new = 1;
run;

data ownerNJ;
set ownerNJ;
lmort = log(smocapi*hinc);
run;

data ownerNJ; set ownerNJ;
affc=0;
if aff=1 and uncrowd = 1 then affc=1;
run;

proc univariate data = ownerNJ;
var lmort lhinc;
run;quit;

data ownerNJ1; set ownerNJ;
morth = 0; if lmort > 13.24 then morth =1;
hinch = 0; if lhinc > 9.88 then hinch =1;
senior = 0; if age>=65 then senior = 1;
/*Aff = 0;if smocapi <= 30 then Aff=1;*/
run;


 proc logistic data= ownerNJ1;
 weight hweight;
 model Affc = morth hinch sex senior rooms vehi edug;
 run;

ods graphic on;
proc logistic data= ownerNJ1 plots(only)= oddsratio(logbase=10);;
 weight hweight;
 model Affc = morth hinch sex senior rooms vehi edug;
 effectplot fit(x=morth);
 run;
```

```
ods graphics off;

/**********************************************/
/*owner high income*/

data ownerNJ; set stateNJ;
if smocapi>0 and smocapi <= 99 and hlinc=0;
lhinc = log(1+hinc);
sm = log( (smocapi+2)/(102-smocapi));
edug = 0;
if educ >7 then edug = 1;
  if new = 1;
run;

data ownerNJ;
set ownerNJ;
lmort = log(smocapi*hinc);
run;

data ownerNJ; set ownerNJ;
affc=0;
if aff=1 and uncrowd = 1 then affc=1;
run;

proc univariate data = ownerNJ;
var lmort lhinc;
run;quit;



data ownerNJ1; set ownerNJ;
morth = 0; if lmort > 13.81 then morth =1;
hinch = 0; if lhinc > 10.97 then hinch =1;
senior = 0; if age>=65 then senior = 1;
Aff = 0;if smocapi <= 30 then Aff=1;
run;

 proc logistic data= ownerNJ1;
 weight hweight;
 model Affc = morth sex senior EDUG rooms vehicl persons hinch;
 run;

ods graphic on;
proc logistic data= ownerNJ1 plots(only)= oddsratio(logbase=10);;
 weight hweight;
 model Affc = morth sex senior EDUG rooms vehicl persons hinch;
 effectplot fit(x=morth);
 effectplot fit(x=hinch);
 run;
ods graphics off;
```

```
/***************************************************************************************************
***************************************************/

data state;
set sasuser.g7alaflo sasuser.gnj;
run;

proc means data = state median;
var finc hinc;
class state;
run;
quit;

/* get rental table in FL */
data stateNJ; set state; if state=12; run;
proc sort data=stateNJ; by puma1;
data famincNJ; set stateNJ;
if hht=1 or hht=2 or hht=3;
run;
proc means median data=famincNJ noprint;
 var finc;
 weight hweight;
 by puma1;
 output out=c median=med;
run;

data c;
  set c;
  keep puma1 med;
  run;
proc sort data=stateNJ;
by puma1;
run;

data stateNJ;
  merge stateNJ c;
  by puma1;

data stateNJ; set stateNJ;
m4 = med*0.8;
hlinc = 0;
 units=1;
 if persons <= 4 and hinc < m4*(1-(4-persons)*0.1) then hlinc = 1;
 if persons > 4 and hinc < m4*(1+(persons-4)*0.08) then hlinc = 1;
/* Crowded */
uncrowd = 0;
if persons <= rooms then uncrowd = 1;
/* new construction */
new = 0;
```

```
if yrbuilt=1 or yrbuilt=2 or yrbuilt=3 then new =1;
/* Affordable */
Aff = 0;
if smocapi > 0 then grapi=.;
if grapi > 0 then smocapi=.;
if smocapi ne . and  smocapi <= 30 then Aff=1;
if grapi ne . and grapi <= 30 then Aff=1;

/*********************************************************************************
************************************************************************/
/* rental low income hlinc=1*/
data rentalNJ; set stateNJ;
if grapi>0 and grapi <= 99 and hlinc=1;
lhinc = log(1+hinc);
gra = log( (grapi+2)/(102-grapi));
edug = 0;
if educ >7 then edug = 1;
  if new = 1;
run;

/* affc = aff &uncrowd */
data rentalNJ; set rentalNJ;
affc=0;
if aff=1 and uncrowd = 1 then affc=1;
run;

/*compute median of variable*/
proc univariate data = rentalNJ;
var grent lhinc;
run;quit;

data rentalNJ1; set rentalNJ;
renth = 0; if grent > 638 then renth =1;
hinch = 0; if lhinc > 9.85 then hinch =1;
senior = 0; if age>=65 then senior = 1;
/*Aff = 0;if grapi <= 30 then Aff=1;*/
run;

 proc logistic data= rentalNJ1;
 weight hweight;
 model Affc = sex rooms renth EDUG hinch vehicl;
 run;

ods graphic on;
proc logistic data= rentalNJ1 plots(only)= oddsratio(logbase=10);;
 weight hweight;
 model Affc = sex rooms renth EDUG hinch vehicl;
 effectplot fit(x=renth);
 effectplot fit(x=hinch);
 run;
```

```
ods graphics off;

/*********************************************/
/*rental high income hlinc=0*/

data rentalNJ; set stateNJ;
if grapi>0 and grapi <= 99 and hlinc=0;
lhinc = log(1+hinc);
gra = log( (grapi+2)/(102-grapi));
edug = 0;
if educ >7 then edug = 1;
  if new = 1;
run;

data rentalNJ; set rentalNJ;
affc=0;
if aff=1 and uncrowd = 1 then affc=1;
run;

proc univariate data = rentalNJ;
var grent lhinc;
run;quit;

data rentalNJ1; set rentalNJ;
renth = 0; if grent > 859.5 then renth =1;
hinch = 0; if lhinc > 10.86 then hinch =1;
senior = 0; if age>=65 then senior = 1;
Aff = 0;if grapi <= 30 then Aff=1;
run;


 proc logistic data= rentalNJ1;
 weight hweight;
 model Affc = senior renth EDUG sex
              rooms vehicl hinch;
 run;

ods graphic on;
proc logistic data= rentalNJ1 plots(only)= oddsratio(logbase=E);;
 weight hweight;
model Affc = senior renth EDUG sex
              rooms vehicl hinch;
 effectplot fit(x=renth);
 effectplot fit(x=senior);
 run;
ods graphics off;


/*************************************************************************
```

```
***************************************************************************/

data state;
set sasuser.g7alaflo sasuser.gnj;
run;

data stateNJ; set state; if state=12; run;
proc sort data=stateNJ; by puma1;
data famincNJ; set stateNJ;
if hht=1 or hht=2 or hht=3;
run;
proc means median data=famincNJ noprint;
 var finc;
 weight hweight;
 by puma1;
 output out=c median=med;
run;

data c;
  set c;
  keep puma1 med;
  run;
proc sort data=stateNJ;
by puma1;
run;

data stateNJ;
  merge stateNJ c;
  by puma1;

data stateNJ; set stateNJ;
m4 = med*0.8;
/* if m4 > 50200 then m4 = 50200; */
hlinc = 0;
 units=1;
 if persons <= 4 and hinc < m4*(1-(4-persons)*0.1) then hlinc = 1;
 if persons > 4 and hinc < m4*(1+(persons-4)*0.08) then hlinc = 1;
/* Crowded */
uncrowd = 0;
if persons <= rooms then uncrowd = 1;
/* new construction */
new = 0;
if yrbuilt=1 or yrbuilt=2 or yrbuilt=3 then new =1;
/*Affordable*/
Aff = 0;
if smocapi > 0 then grapi=.;
if grapi > 0 then smocapi=.;
if smocapi ne . and  smocapi <= 30 then Aff=1;
if grapi ne . and grapi <= 30 then Aff=1;
```

```
/*********************************************************************************
************************************************************************/
/* owner low income*/
data ownerNJ; set stateNJ;
if smocapi>0 and smocapi <= 99 and hlinc=1;
lhinc = log(1+hinc);
sm = log( (smocapi+2)/(102-smocapi));
edug = 0;
if educ >7 then edug = 1;
  if new = 1;
run;

data ownerNJ;
set ownerNJ;
lmort = log(smocapi*hinc);
run;

data ownerNJ; set ownerNJ;
affc=0;
if aff=1 and uncrowd = 1 then affc=1;
run;

proc univariate data = ownerNJ;
var lmort lhinc;
run;quit;

data ownerNJ1; set ownerNJ;
morth = 0; if lmort > 13.49 then morth =1;
hinch = 0; if lhinc > 10 then hinch =1;
senior = 0; if age>=65 then senior = 1;
/*Aff = 0;if smocapi <= 30 then Aff=1;*/
run;


 proc logistic data= ownerNJ1;
 weight hweight;
 model Affc = morth hinch sex senior rooms vehi edug;
 run;

ods graphic on;
proc logistic data= ownerNJ1 plots(only)= oddsratio(logbase=10);;
 weight hweight;
 model Affc = morth hinch sex senior rooms vehi edug;
 effectplot fit(x=morth);
 run;
ods graphics off;

/***********************************************/
/*owner high income*/
```

```
data ownerNJ; set stateNJ;
if smocapi>0 and smocapi <= 99 and hlinc=0;
lhinc = log(1+hinc);
sm = log( (smocapi+2)/(102-smocapi));
edug = 0;
if educ >7 then edug = 1;
  if new = 1;
run;

data ownerNJ;
set ownerNJ;
lmort = log(smocapi*hinc);
run;

data ownerNJ; set ownerNJ;
affc=0;
if aff=1 and uncrowd = 1 then affc=1;
run;

proc univariate data = ownerNJ;
var lmort lhinc;
run;quit;



data ownerNJ1; set ownerNJ;
morth = 0; if lmort > 14.06 then morth =1;
hinch = 0; if lhinc > 11.11 then hinch =1;
senior = 0; if age>=65 then senior = 1;
Aff = 0;if smocapi <= 30 then Aff=1;
run;

 proc logistic data= ownerNJ1;
 weight hweight;
 model Affc = morth sex senior EDUG rooms vehicl persons hinch;
 run;

ods graphic on;
proc logistic data= ownerNJ1 plots(only)= oddsratio(logbase=10);;
 weight hweight;
 model Affc = morth sex senior EDUG rooms vehicl persons hinch;
 effectplot fit(x=morth);
 effectplot fit(x=hinch);
 run;
ods graphics off;
```

B. R code for variable selection using LASSO

import all 4*3 data sets into R by using "import dataset" option
##NJ
###1.1rent with low income

```{r}
library(Hmisc)
NJRM <- c("senior","renth","edug","affc","hinch","SEX","BLDGSZ","PERSONS","ROOMS","VEHICL")
NJRMNl <- rentalnjml[NJRM]
maxnjrl <- as.matrix(NJRMNl)
rcorr(maxnjrl[,-4], type = c("pearson","spearman"))
lm <- lsfit(maxnjrl[,-4],maxnjrl[,4])
```
```{r}
library(glmnet)
library(dygraphs)
library(coefplot)
library(ggplot2)
fit <- glmnet(maxnjrl[,-4],maxnjrl[,4], family = "binomial")
plot(fit, xvar = "lambda", label = TRUE)
coefpath(fit)
```
```{r}
rcorr(maxnjrl[,-c(1,4,6)], type = c("pearson","spearman"))
lm2 <- lsfit(maxnjrl[,-c(1,4,6)],maxnjrl[,4])
summary(lm2)
```
###1.2 rent with high income
```{r}
library(Hmisc)
NJRM <- c("senior","renth","edug","affc","hinch","SEX","BLDGSZ","PERSONS","ROOMS","VEHICL")
NJRMNh <- rentalnjmh[NJRM]
maxnjrh <- as.matrix(NJRMNh)
rcorr(maxnjrh[,-4], type = c("pearson","spearman"))
lm3 <- lsfit(maxnjrh[,-4],maxnjrh[,4])
```
```{r}
library(glmnet)
library(dygraphs)
library(coefplot)
library(ggplot2)
fit2 <- glmnet(maxnjrh[,-4],maxnjrh[,4], family = "binomial")
plot(fit2, xvar = "lambda", label = TRUE)
coefpath(fit2)
```

```{r}
rcorr(maxnjrh[,-c(4,6,7,8)], type = c("pearson","spearman"))
```

~~~
lm4 <- lsfit(maxnjrh[,-c(4,6,7,8)],maxnjrh[,4])
summary(lm4)
```

###owner with low income
```{r}
NJOM <- c("senior","morth","edug","affc","hinch","SEX","BLDGSZ","PERSONS","ROOMS","VEHICL")
NJOMNl <- ownernjml[NJOM]
maxnjol <- as.matrix(NJOMNl)
```

```{r}
library(glmnet)
library(dygraphs)
library(coefplot)
library(ggplot2)
fit3 <- glmnet(maxnjol[,-4],maxnjol[,4], family = "binomial")
plot(fit3, xvar = "lambda", label = TRUE)
coefpath(fit3)
```
```{r}
rcorr(maxnjol[,-c(1,4,6)], type = c("pearson","spearman"))
lm5 <- lsfit(maxnjol[,-c(1,4,6)],maxnjol[,4])
summary(lm5)
```

###owner with high income

```{r}
NJOM <- c("senior","morth","edug","affc","hinch","SEX","BLDGSZ","PERSONS","ROOMS","VEHICL")
NJOMNh <- ownernjmh[NJOM]
maxnjoh <- as.matrix(NJOMNh)
```

```{r}
library(glmnet)
library(dygraphs)
library(coefplot)
library(ggplot2)
fit4 <- glmnet(maxnjoh[,-4],maxnjoh[,4], family = "binomial")
plot(fit4, xvar = "lambda", label = TRUE)
coefpath(fit4)
```

```{r}
rcorr(maxnjoh[,-c(4,6,8)], type = c("pearson","spearman"))
lm6 <- lsfit(maxnjol[,-c(4,6,8)],maxnjol[,4])
summary(lm6)
```
##the same operation for AL and FL data
~~~