

Analysis of the Behavior of Electricity Price in Oregon Households

Sam Oliszewski

Oregon State University

### Abstract

This paper discusses an analysis of the electricity expenses from the American Community Survey for households in Oregon (ACS) during the year 2015. The analysis has both an explanatory and predictive question of interest it aims to answer. First, the explanatory question “Is there a difference in electricity expenses for people living in houses versus apartments?” is investigated. Next, the predictive question “Can a model be created to predict electricity costs for a household in Oregon?” is explored. To begin, an exploratory analysis was performed to provide insight on the data being studied. Next, the approaches for model determination in the explanatory and predictive settings are described, as well as limitations of the selected models. Finally, conclusions of the analysis are reported, and caveats of the analysis are presented to aid in the overall interpretation of the study results. It was determined that there is no significant difference in electricity expense for people living in houses versus apartments among Oregon households. Additionally, a model was presented that could be used for predicting the electricity expense of an Oregon household, but the predictive power is less than ideal. It is recommended that future research be conducted to identify additional characteristics of households that may influence the electricity expense and help make predictions more accurate.

*Keywords:* multiple linear regression, Oregon household electricity, predictive modeling

### Analysis of the Behavior of Electricity Price in Oregon Households

The American Community Survey (ACS) was conducted for households in Oregon during the year 2015. Two questions of interest are raised about this data— one explanatory and one predictive in nature. This analysis will first address the explanatory question “Is there a difference in electricity expenses for people living in houses versus apartments?” Next, the predictive question “Can a model be created to predict electricity costs for a household in Oregon?” is explored. The paper will begin by discussing the exploratory analysis that was performed to provide insight on the data being studied. Next, the approaches for model determination in the explanatory and predictive settings are described, in addition to discussion about the limitations of the selected models. Finally, conclusions drawn from the analysis are reported, and caveats of the analysis are offered to aid in the overall interpretation of the study results.

### Exploratory Analysis

This section discusses the exploratory analysis performed to investigate the ACS dataset.

#### Assessing the Available Variables

**Original Variables.** The raw dataset contained fifteen predictor variables. The original variables included in the ACS dataset used for this analysis are: serial number (SERIALNO), type of unit (TYPE), number of people in the household (NP), lot size in acres (ACR), bedrooms in household (BDSP), units in structure (BLD), fuel cost (FULP), gas cost (GASP), house heating fuel type (HFL), number of rooms in household (RMSP), tenure (TEN), property value (VALP), year structure was built (YBL), presence of under age 18 persons (R18), and presence of over age 60 persons (R60). The response variable in this study is the price of electricity per

household (ELEP). A sample of the data, as well as descriptions of possible values for each variable, can be referenced in the Appendix.

**Removing Unnecessary Predictors.** The dataset originally contained fifteen predictor variables and the modified set of predictors used in this analysis contained thirteen. The variables deemed unnecessary for the study were SERIALNO and TYPE. The SERIALNO variable was excluded from the data set because the value of this predictor is an identifier of the observation and has no relationship with the response variable ELEP. The TYPE variable was excluded from the data set because the value is identical for all the observations in the data set and therefore would not contribute to the change in the response variable ELEP. The predictor variables included in the reduced data set used for this analysis are: NP, ACR, BDSP, BLD, FULP, GASP, HFL, RMSP, TEN, VALP, YBL, R18, and R60. The response variable in this study is ELEP.

### **Addressing Missing Values**

**Explanation of Problem.** Since there are several missing values in the data set, the number of observations able to be studied is reduced when left in its raw state. Leaving the data untouched would limit the ability to draw meaningful conclusions from the data.

**Solution to the Missing Value Problem.** To address the issue of missing values, imputation was performed to increase the number of observations able to be studied. There was a mix of integer and factor variables in the dataset and two different methods of imputation were performed as a result— median and mode imputation. For integer variables, median imputation was used because the median was a good measure of the center of the data that preserves the integer value in the result. Since integers are ordered, a median is a reasonable statistic to use for imputation. For the factor values, mode imputation was used. This method was preferred due to many of the levels being unordered.

## **Modifying Variable Structure for Different Goals**

**BLD in the Explanatory Problem Setting.** Since the goal in the explanatory problem is to assess the difference in electricity expenses for people living in houses versus apartments, a variable needed to be chosen to classify a household as a house or an apartment. Two possible variables were available—TEN and BLD. Ultimately, the BLD variable was determined to be better for this classification because the responses for this variable were more explicit about the house versus apartment classification. However, the structure of the variable needed to be altered to fit the needs of the question. Therefore, the BLD variable was restructured from the original ten factors into two factor levels: “House” and “Apartment”. The level House was taken from the original levels “One-family house detached” and “One-family house attached”. The level Apartment was taken from the original levels “2 Apartments”, “3-4 Apartments”, “5-9 Apartments”, “10-19 Apartments”, “20-49 Apartments” and “50 or more apartments”. After the variable was restructured, the observations in the dataset containing BLD values “Mobile home or trailer” and “Boat, RV, van, etc.” were removed because they were not relevant to the study for the context of this question of interest.

**BLD in the Prediction Problem Setting.** In the prediction problem for this analysis, there is no need to use the restructured BLD variable as discussed previously. It is preferred to consider the original factor levels and their associated observations because the question does not specify that there should be a grouping of BLD type. Further, the specificity of BLD type could prove influential on the prediction and should thus be considered for this problem. Therefore, the BLD variable in the prediction problem will include the original ten factor levels “Mobile home or trailer”, “One-family house detached”, “One-family house attached”, “2 Apartments”, “3-4 Apartments”, “5-9 Apartments”, “10-19 Apartments”, “20-49 Apartments”, “50 or more

apartments”, and “Boat, RV, van, etc.”. All the original observations are included in this problem.

## Visualizing Relationships in Data

**Figures of Predictors Against Response.** The initial exploration of the data included plotting each predictor value against the response to determine any apparent relationships within the dataset. These relationships are depicted after imputation occurred. The resulting plots are shown below:

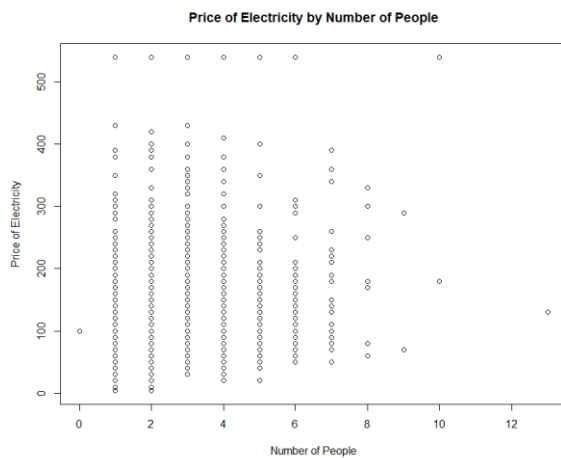


Figure 1. ELEP by NP

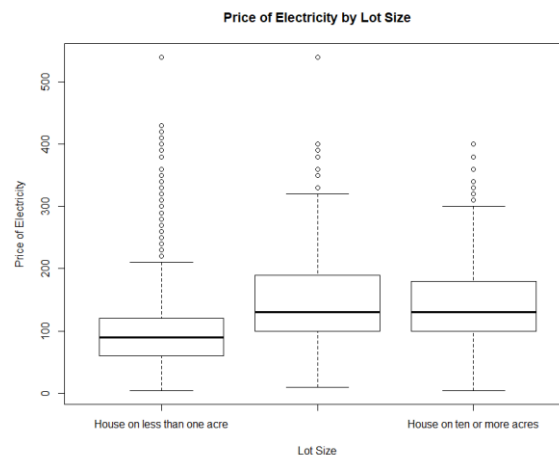


Figure 2. ELEP by ACR

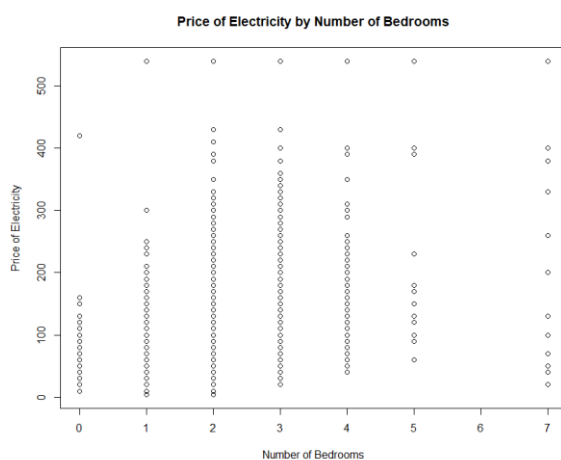


Figure 3. ELEP by BDSP

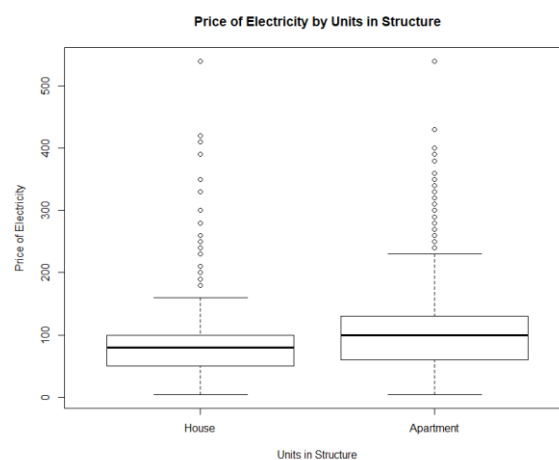


Figure 4. ELEP by BLD

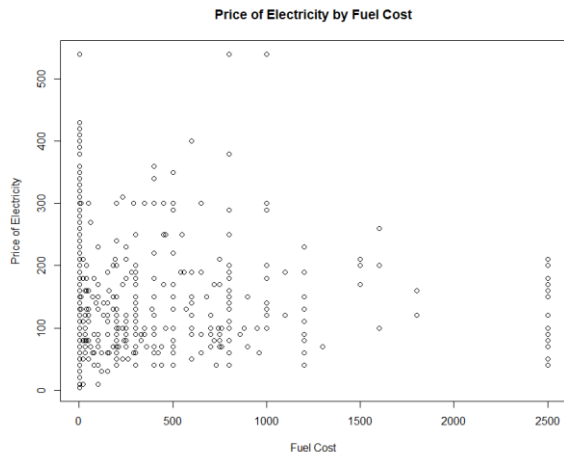


Figure 5. ELEP by FULP

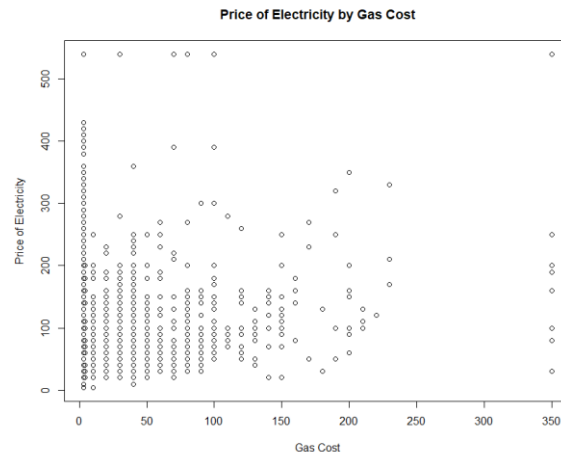


Figure 6. ELEP by GASP

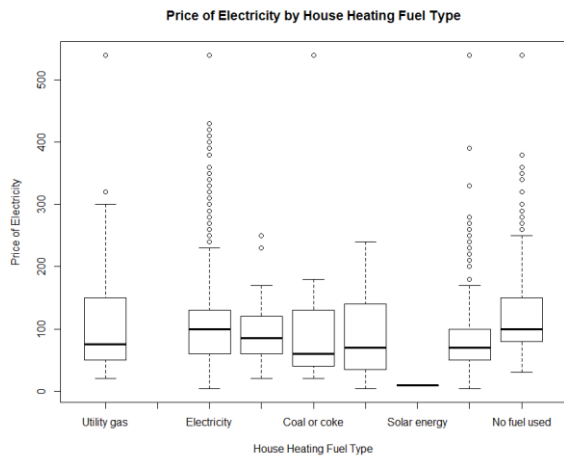


Figure 7. ELEP by HFL

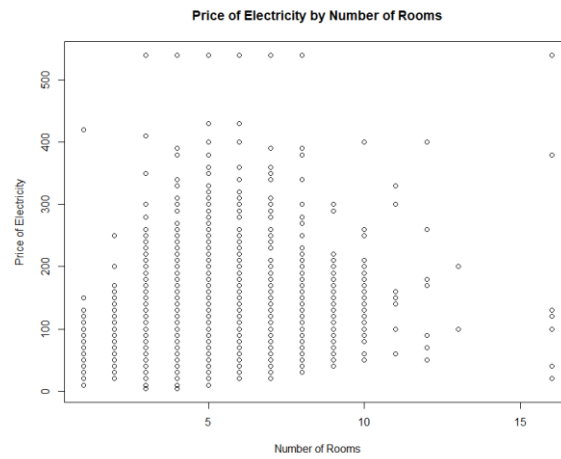


Figure 8. ELEP by RMSP

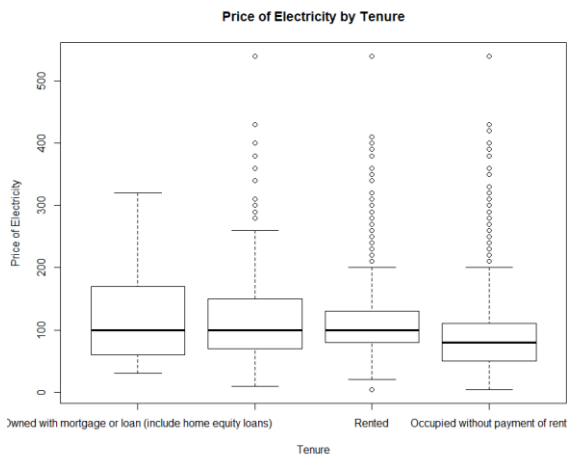


Figure 9. ELEP by TEN

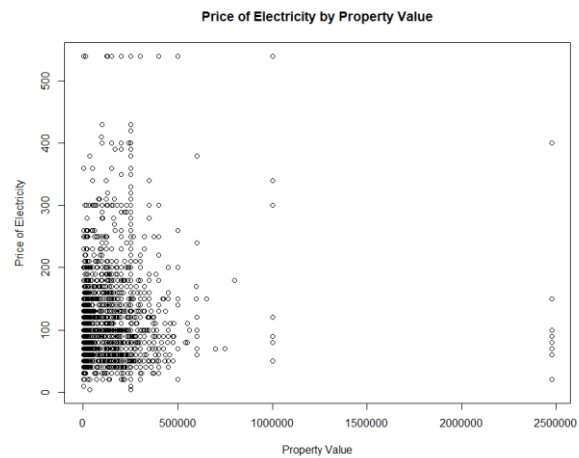


Figure 10. ELEP by VALP

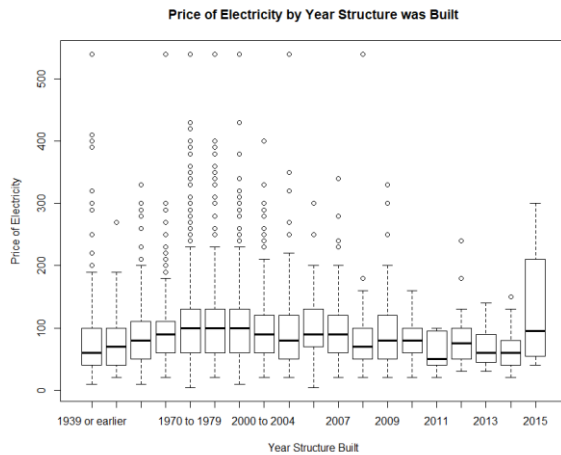


Figure 11. ELEP by YBL

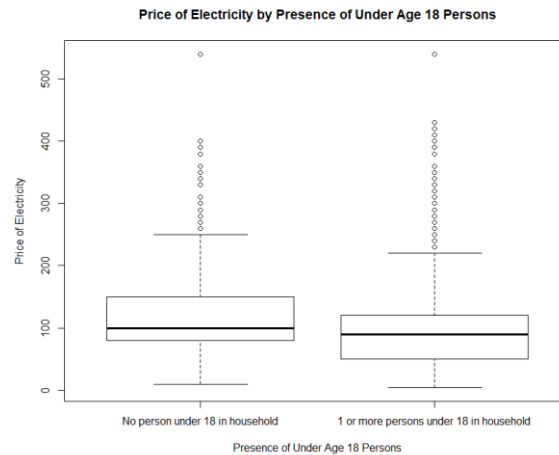


Figure 12. ELEP by R18

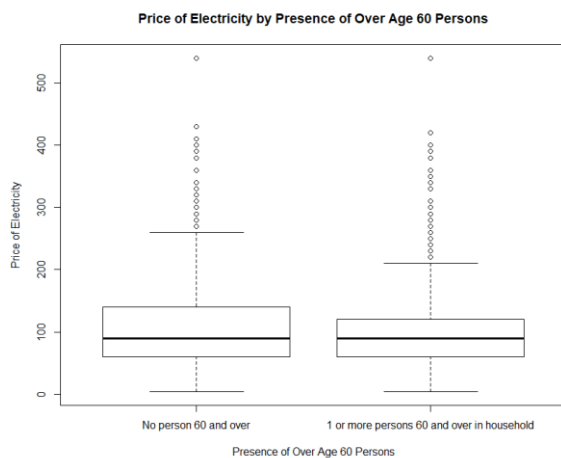


Figure 13. ELEP by R60

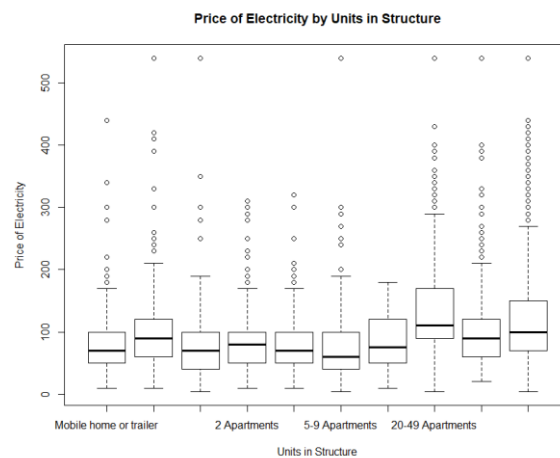


Figure 14. ELEP by Original BLD Factors

**Conclusions.** After examining these plots, there is no strong evidence of impact on the response variable for different values of the predictor for the predictor variables R60, TEN or VALP (most of the observations are clustered). It is seen that there is some evidence of impact on the response variable for different values of the predictor for the predictor variables R18 (slightly higher for no 18 year old persons), YBL (clear outliers for 2008 and earlier years and 2015 is much higher), RMSP (consistency for number of rooms 4 and greater and lower prices for 3 and fewer rooms), HFL (fairly consistent medians across fuel types, but more high values for electricity), GASP (more high values when gas price is almost zero), FULP (more high



values when gas price is almost zero), BLD (in the original factoring the 20-49 apartments level has a median higher than most other structure types Q3 value, and in the restructured BLD the apartment level is higher than house), BDSP (consistency with 2 or more bedrooms and lower electricity price with 0 and 1 bedrooms), ACR (lower electricity price with house on less than one acre, but more outliers), and NP (consistent electricity prices with 7 and fewer people, but lower values with 8 or more people per household).

### **Explanatory Problem**

#### **Overview**

The goal in this section is to determine whether people living in apartments pay less on electricity than those living in houses and by how much. To make this determination, first a multiple linear regression (MLR) model must be constructed to reflect the dataset. This process will involve fitting both a model without interactions and a model with interactions. Both of the fitted models will include all the predictors in the dataset which means that the regression line will have its slope altered based on the values of each term because each predictor has an associated coefficient (beta) that represents the estimated amount by which the mean value of the response variable (ELEP) changes for a unit change in the predictor when all other predictors are held fixed. Once two models are fit, the preferred model will be determined, and a summary of the model will be reported that will identify the difference in electricity expenses between people living in apartments and people living in houses.

## Methods

**Fitting a Model without Interactions.** An MLR model including all thirteen predictors of the following form was fit:

$$\begin{aligned} ELEP = & \beta_0 + \beta_1 NP + \beta_2 ACR + \beta_3 BDSP + \beta_4 BLD + \beta_5 FULP + \beta_6 GASP + \beta_7 HFL \\ & + \beta_8 RMSP + \beta_9 TEN + \beta_{10} VALP + \beta_{11} YBL + \beta_{12} R18 + \beta_{13} R60 \end{aligned}$$

To assess the fit of this model, first a residuals versus fitted plot was generated. This plot is shown below:

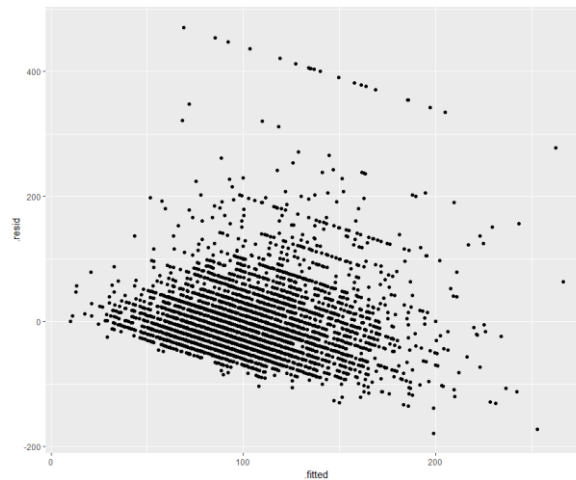


Figure 15. Residual versus Fitted – Model without Interactions

Of note in Figure 15 is the apparent parallel lines created by the plotted residuals. This is likely a reflection of the discrete nature of the response variable *ELEP*, forcing each value to take on integer values rather than continuous ones. This is also a possible cause of the suspected heteroscedastic behavior causing the residuals to take a shape. With this understanding of the response variable, it is reasonable to assume that the model satisfies the constant variance assumption for MLR. Further, the assumption of linearity is also satisfied because the plotted points would generally be random if the response variable were not discrete and thus MLR is still reasonable to pursue. The sample size is sufficiently large to assume that the assumption of normality for each *Y* around its mean is satisfied and the assumption of independence is also met

because the households studied were selected at random and there is no reason to believe that one household's electricity price should be related to another's.

To further assess whether MLR is appropriate to perform, residuals versus explanatory variable plots for each of the thirteen explanatory variables were created. These plots are shown below:

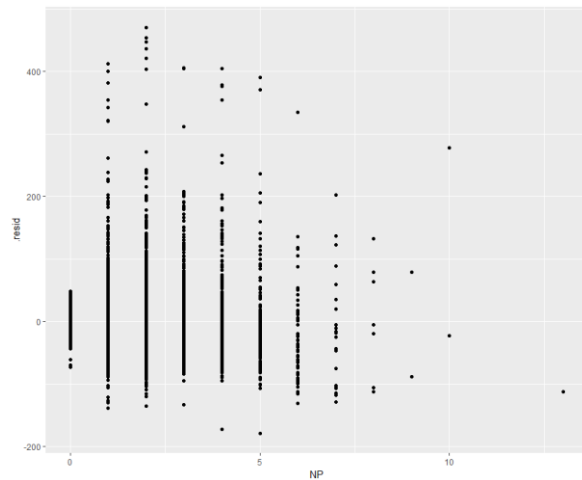


Figure 16. Residuals versus NP

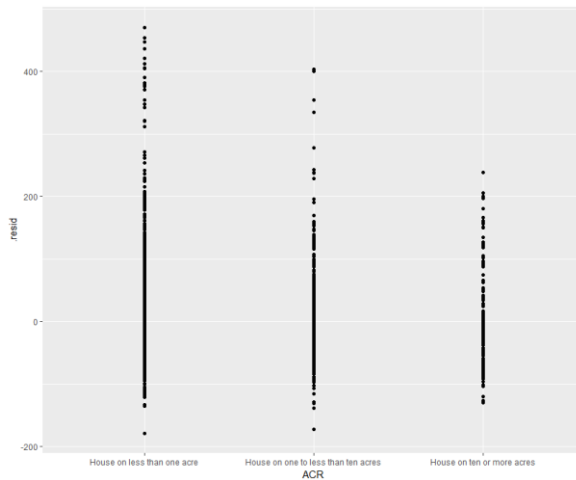


Figure 17. Residuals versus ACR

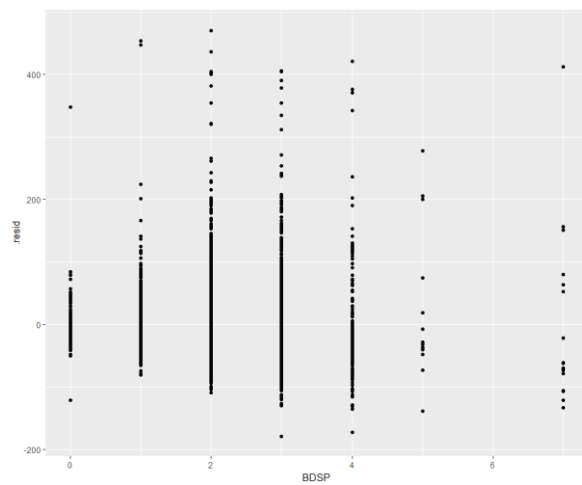


Figure 18. Residuals versus BDSP

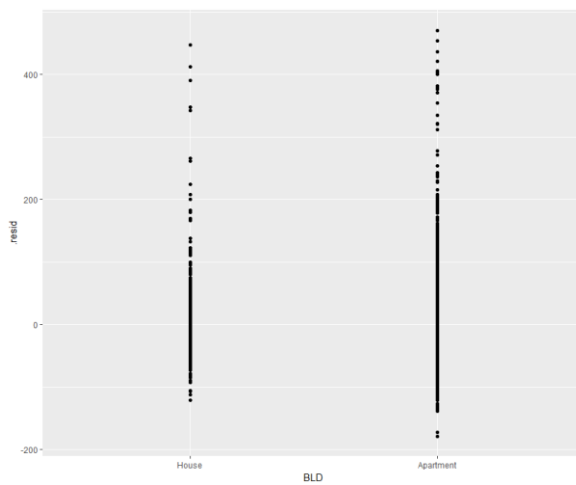


Figure 19. Residuals versus BLD

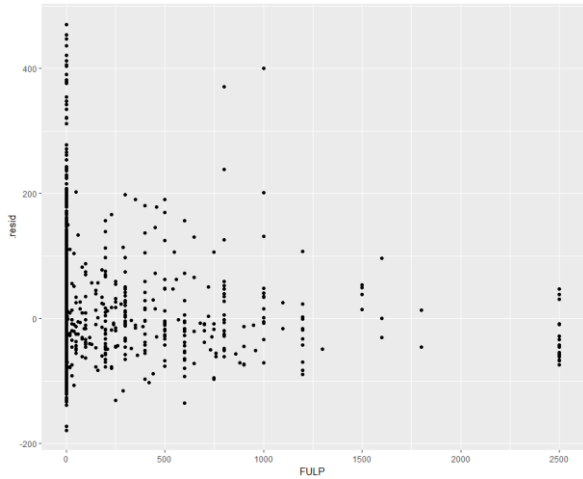


Figure 20. Residuals versus FULP

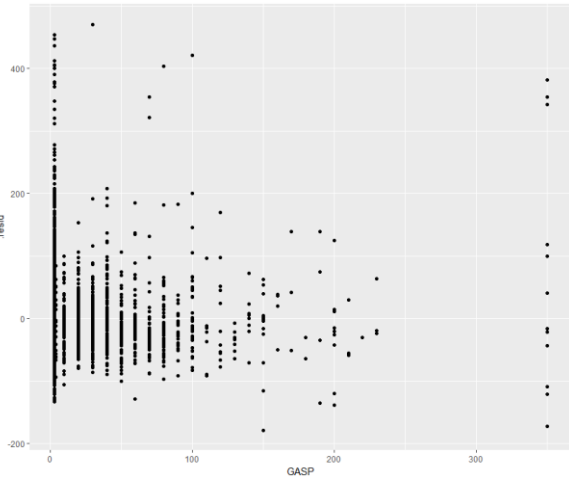


Figure 21. Residuals versus GASP

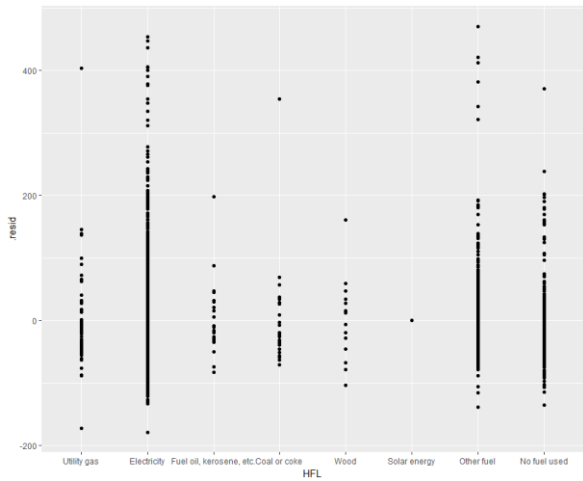


Figure 22. Residuals versus HFL

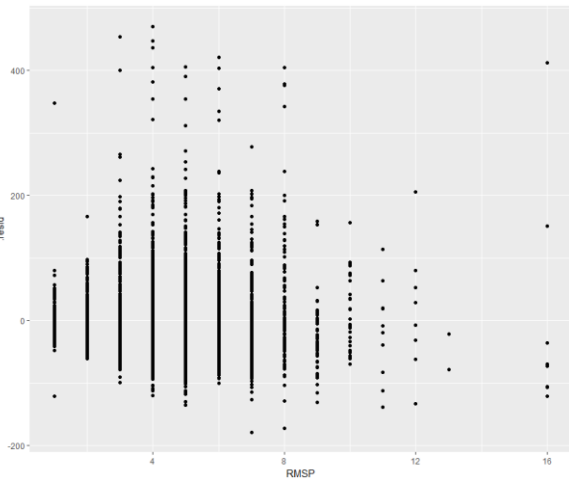


Figure 23. Residuals versus RMSP

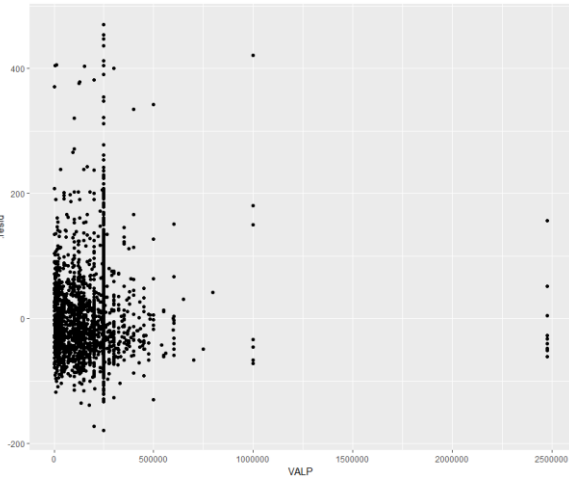
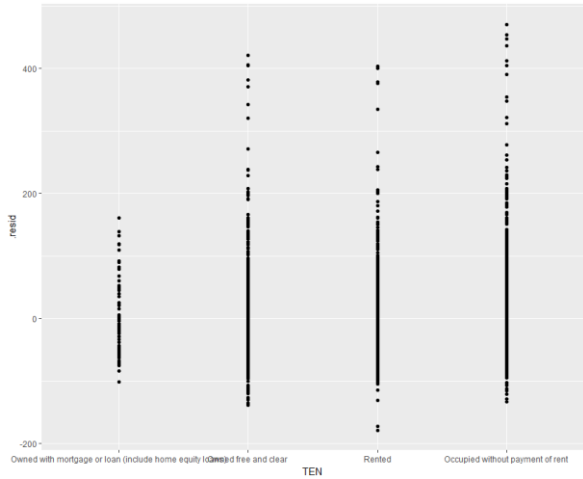


Figure 24. Residuals versus TEN

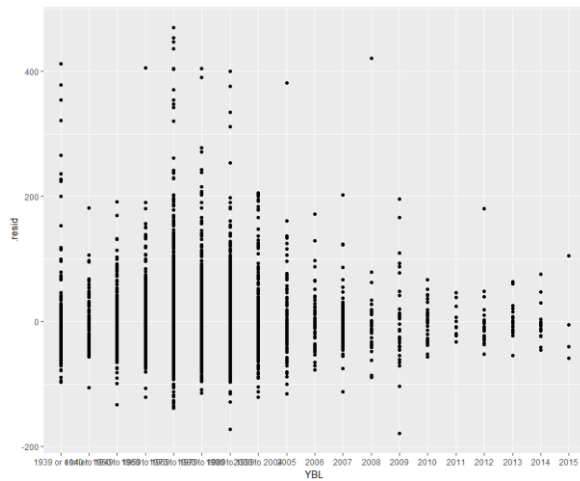


Figure 25. Residuals versus VALP

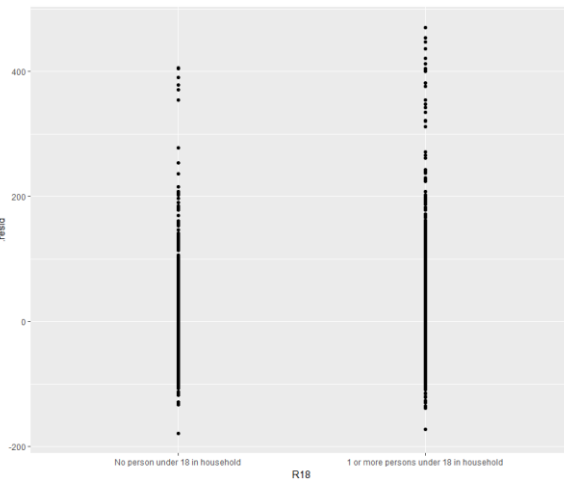


Figure 26. Residuals versus YBL

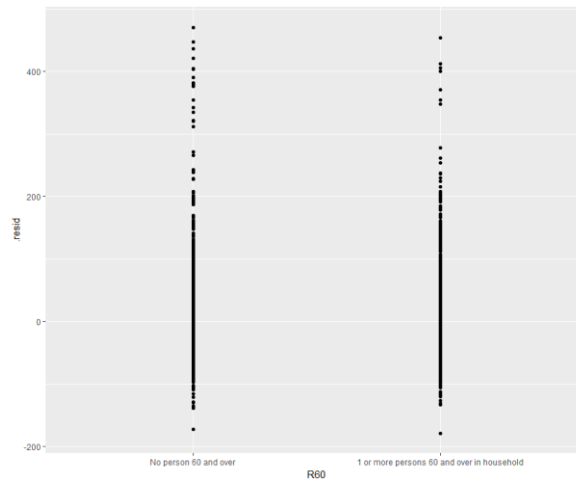


Figure 27. Residuals versus R18

Figure 28. Residuals versus R60

While each of the thirteen residuals versus explanatory variable plots showed some values of electricity being higher than most of the data points, there are too many points to be considered outliers and do not present any evidence that the assumptions for MLR are not met. This is consistent with the findings from earlier that suggest MLR is appropriate. Further, while the residuals versus predictor plots seem to suggest non-constant variance, the variance is not due only to single predictor and therefore the residuals versus fitted plot is a more appropriate

measure of model fit and does not indicate any MLR assumption violations that would halt inference from proceeding.

**Fitting a Model with Interactions.** A model including all thirteen of the predictors and their interactions of the following form was fit:

$$\begin{aligned}
 ELEP = & \beta_0 + (\beta_1 NP + \beta_2 ACR + \beta_3 BDSP + \beta_4 BLD + \beta_5 FULP + \beta_6 GASP + \beta_7 HFL \\
 & + \beta_8 RMSP + \beta_9 TEN + \beta_{10} VALP + \beta_{11} YBL + \beta_{12} R18 + \beta_{13} R60) * (\beta_1 NP \\
 & + \beta_2 ACR + \beta_3 BDSP + \beta_4 BLD + \beta_5 FULP + \beta_6 GASP + \beta_7 HFL + \beta_8 RMSP \\
 & + \beta_9 TEN + \beta_{10} VALP + \beta_{11} YBL + \beta_{12} R18 + \beta_{13} R60)
 \end{aligned}$$

To assess the fit of this model, a residuals versus fitted plot was generated. This plot is shown below:

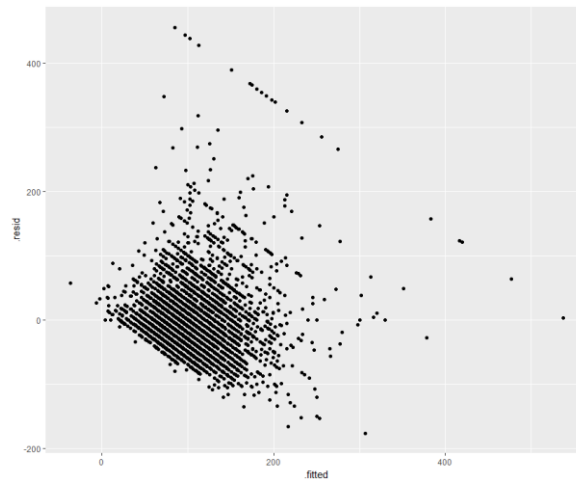


Figure 29. Residuals versus Fitted – Model with Interactions

This plot is very similar to the residuals versus fitted plot for the model without interactions and therefore does not present any evidence that the assumptions for MLR are not met, as were discussed previously.

**Choosing the Best Model.** After the two models were fit and the assumptions for MLR were proven to be satisfied, a decision needed to be made as to which of the models best fit the data. To make this determination, an ANOVA test was performed comparing the two models.

The results of this test found that the model with interactions was preferred (Extra Sum of Squares comparing model with interaction to model without interaction,  $p\text{-value} = 2.2e-16$ ). This analysis will thus proceed with this model.

**Summarizing the Preferred Model.** With the best model for the data selected, a summary was generated to determine whether there is a difference in the electricity expenses between Oregonians living in apartments and houses and how much that difference is. This showed that the estimated difference in the BLD term for people living in apartments versus houses is \$84.23, with a standard error of \$51.66, a t-statistic of 1.63, and a p-value of 0.10. The following table summarizes all the relevant coefficients for the question of interest from the fitted model:

<b>Term</b>	<b>Point Estimate</b>	<b>Standard Error</b>	<b>t-value</b>	<b>p-value</b>
BLDApartment	84.23	51.66	1.63	0.10
BLDApartment:GASP	-38.11	12.49	-3.05	0.00
BLDApartment:TENOwned free and clear	-68.84	29.68	-2.32	0.02
BLDApartment:TENOccupied without payment of rent	-57.77	26.27	-2.12	0.03

## **Conclusion**

Overall, with 95% confidence, there is no evidence that the mean price of electricity for people living in apartments is less than the mean price of electricity for people living in houses. It is estimated that the mean price of electricity for people living in houses is between \$17.05 below and \$185.51 above the mean price of electricity for people living in houses, with a point estimate

of \$84.23 (t-test,  $df=4008$ ,  $t=1.63$ ,  $p\text{-value}=0.10$ ). However, it is important to note that the preferred model included interactions, which means that the influence of BLD on the electricity expense is more complicated than simply assessing the value of this one coefficient. There are three significant interaction terms including the BLD variable in the model, which also can influence the electricity expense. These interactions appear to lower the price of electricity as well, which is an interesting observation. The interaction of an apartment household with gas price has a mean electricity price between \$50.60 and \$25.62 below the mean price of electricity for people without this interaction (t-test,  $df=4008$ ,  $t=-3.05$ ,  $p\text{-value}=0.00$ ). The interaction of an apartment household with tenure that is owned free and clear has a mean electricity price between \$98.52 and \$39.16 below the mean price of electricity for people without this interaction (t-test,  $df=4008$ ,  $t=-2.32$ ,  $p\text{-value}=0.02$ ). The interaction of an apartment household with tenure that is occupied without payment of rent has a mean electricity price between \$84.04 and \$31.50 below the mean price of electricity for people without this interaction (t-test,  $df=4008$ ,  $t=-2.32$ ,  $p\text{-value}=0.03$ ).

**Limitations.** Of note on the conclusions from this study is that the values calculated may be somewhat inaccurate because the data was imputed instead of resulting solely from actual observations. This may underestimate the standard error associated with the generated values. Also noteworthy is that the missing data in the dataset requires investigation to see if it was missing completely at random as this is the only type of missing data that would not affect the inferential conclusions.



## Prediction Problem

### Overview

The goal in this section is to create a model that could be used to predict electricity costs for a household in Oregon. To do this, best subset selection will be performed along with k-fold cross validation to select the best model. Afterward, a summary of the best model will be reported.

### Methods

**K-Fold Cross Validation Approach.** To determine the best model, thirteen models were fitted that included from one to thirteen predictors. Best subset selection was performed to determine which predictors to include in the model at each model size. Then, to test each model, the data was divided into k number of groups of even size. In this study, the value of k was equal to 10. Then, 10 experiments were performed where one of the 10 groups is used as a test set, and the remaining k-1 ( $10 - 1 = 9$ ) groups acted as the training set. The MSE was computed for each model and the lowest test MSE among each calculated for the different model sizes was selected as the preferred model. This indicated the optimal number of predictors to include in the model. The k-fold cross validation approach resulted in a model that included all thirteen of the original predictors and had a MSE of 3846.57. The  $R^2$  statistic for this model is 0.24.

### Conclusion

The best model as chosen by the k-fold cross validation includes the following terms:

Term	Coefficient	Standard Error	t-statistic	p-value
Intercept	-5.46e00	7.04e00	-0.78	0.44
NP	1.36e01	5.23e-01	26.07	<b>2e-16</b>
BLDOne-family house detached	2.16e01	4.40e00	4.92	<b>8.96e-07</b>
BLDOne-family house attached	2.67e00	4.53e00	0.59	0.56

BLD2 Apartments	4.35e00	4.02e00	1.08	0.28
BLD3-4 Apartments	-1.84e00	4.04e00	-0.46	0.65
BLD5-9 Apartments	-9.21e-01	4.29e00	-0.22	0.83
BLD10-19 Apartments	2.96e01	1.29e01	2.29	<b>0.02</b>
BLD20-40 Apartments	3.95e01	3.67e00	10.76	<b>2e-16</b>
BLD 50 or more apartments	2.32e01	3.98e00	5.84	<b>5.31e-09</b>
BLDBoat, Rv, van, etc.	3.15e01	3.39e00	9.28	<b>2e-16</b>
ACRHouse on one to less than ten acres	2.75e01	1.64e00	16.83	<b>2e-16</b>
ACRHouse on ten or more acres	2.62e01	2.44e00	10.73	<b>2e-16</b>
BDSP	6.39e00	7.79e-01	8.20	<b>2.58e-16</b>
FULP	8.58e-03	2.12e-03	4.04	<b>5.35e-05</b>
GASP	1.48e-01	1.24e-02	11.87	<b>2e-16</b>
HFLBottled, tank, or LP gas	-2.03e01	4.50e01	-0.45	0.65
HFLElectricity	4.08e01	3.78e00	10.79	<b>2e-16</b>
HFLFuel oil, kerosene, etc.	2.00e00	5.07e00	0.39	0.69
HFLCoal or coke	2.56e01	1.01e01	2.54	<b>0.01</b>
HFLWood	2.85e01	7.15e00	3.98	<b>6.81e-05</b>
HFLSolar energy	-1.47e01	2.27e01	-0.65	0.52
HFLOther fuel	-9.19e00	3.69e00	-2.49	<b>0.01</b>
HFLNo fuel used	9.46e00	4.08e00	2.32	<b>0.02</b>
RMSP	1.69e00	3.34e-01	5.06	<b>4.30e-07</b>
TENOwned free and clear	-8.97e00	4.30e00	-2.08	<b>0.04</b>
TENRented	-2.39e00	4.24e00	-0.56	0.57
TENOccupied without payment of rent	-3.54e00	4.34e00	-0.81	0.42
VALP	2.36e-05	2.42e-06	9.76	<b>2e-16</b>
YBL1940 to 1949	8.22e00	2.54e00	3.24	<b>0.00</b>
YBL1950 to 1959	5.42e00	2.20e00	2.47	<b>0.01</b>

YBL1960 to 1969	3.34e00	2.16e00	1.55	0.12
YBL1970 to 1979	8.73e00	1.86e00	4.70	<b>2.65e-06</b>
YBL1980 to 1989	4.73e00	2.14e00	2.21	<b>0.03</b>
YBL1990 to 1999	-7.32e-01	1.93e00	-0.38	0.70
YBL2000 to 2004	-2.64e00	2.35e00	-1.12	0.26
YBL2005	-4.44e00	3.92e00	-1.13	0.26
YBL2006	-6.25e00	4.37e00	-1.43	0.15
YBL2007	-4.92e00	4.28e00	-1.15	0.25
YBL2008	-7.02e00	5.74e00	-1.22	0.22
YBL2009	2.56e00	6.10e00	0.42	0.67
YBL2010	-7.63e00	6.75e00	-1.13	0.26
YBL2011	9.72e-01	8.69e00	0.11	0.91
YBL2012	-3.89e00	7.84e00	-0.50	0.62
YBL2013	-1.20e01	7.13e00	-1.68	0.09
YBL2014	-1.15e01	7.83e00	-1.46	0.14
YBL2015	-2.27e01	1.42e01	-1.60	0.11
R181 or more persons under 18 in household	4.47e00	1.75e00	2.55	<b>0.01</b>
R601 or more persons 60 and over in household	-3.70e00	1.20e00	-3.09	<b>0.00</b>

Of note is that while this model is found to be the preferred model when compared to others, in terms of predictive power, it may not be particularly strong. The  $R^2$  statistic is relatively low and the MSE is a bit higher than desired. This means that the predictions resulting from this model may not be as accurate as one would hope. The coefficients for each term presented in the table above indicate the influence on electricity expense that each term has based on the value of the term. One of the reasons the predictive power may be reduced is that earlier it was determined

that there were interactions between some of the terms and therefore this may not be addressed fully in this predictive model.

**Limitations.** While this model was selected as the best fit for the data, there are some limitations to consider. One such limitation to the model is that imputation was performed on the data meaning that several observations are affected by the imputed values, rather than actual observations. As a result, the standard error is likely underestimated. Also of note is that the missing data in the dataset requires investigation to see if it was missing completely at random as this is the only type of missing data that would not affect the inferential conclusions. Another limitation to this predictive model is that these households are only representative of Oregon households and should not be considered representative of households in other states. Also affecting the performance of this predictive model is how the sample size is large and we may run into issues such as small effects being considered statistically significant, raising the question of practical versus statistical significance.

### **Conclusion**

This analysis aimed to answer two questions of interest. The respective questions were “Is there a difference in electricity expenses for people living in houses versus apartments?” and “Can a model be created to predict electricity costs for a household in Oregon?” Ultimately, it was determined that there is no significant difference in electricity expense for people living in houses versus apartments among Oregon households. Additionally, a model was presented that could be used for predicting the electricity expense of an Oregon household, but the predictive power is less than ideal. Based on the results of this study, further investigation is recommended to identify additional characteristics of households that may influence the electricity expense and help make predictions more accurate.

## Appendix

**Sample Data:**

SERIALNO	NP	TYPE	ACR	BDSP	BLD	ELEP	FULP	GASP	HFL	RMSP	TEN	VALP	YBL	R18	R60
70	4	1	House on less than one acre	2	One-family house detached	70	2	3	Wood	4	Rented	NA	1939 or earlier	1 or more	none
106	0	1	NA	2	2 Apartments	NA	NA	NA	NA	3	NA	NA	1970 to 1979	NA	NA
163	2	1	House on less than one acre	2	One-family house detached	100	600	3	Fuel oil, kerosene, etc.	7	Owned with mortgage or loan	225000	1939 or earlier	none	none
178	1	1	House on less than one acre	3	One-family house detached	60	2	110	Utility gas	8	Owned free and clear	315000	1939 or earlier	none	1 or more
243	2	1	House on less than one acre	4	One-family house detached	80	2	20	Utility gas	8	Owned free and clear	200000	1950 to 1959	none	1 or more

**Description of Variables:****SERIALNO:**

**Definition:** Housing unit/GQ person serial number

**Possible Values:** Any Discrete Integer

**NP:**

**Definition:** Number of person records following this housing record

**Possible Values:** Integer between 0 and 20

## Appendix (Continued)

**TYPE:**

*Definition:* Type of unit

*Possible Values:* Categorical Integer

- (1) Housing unit
- (2) Institutional group quarters
- (3) Noninstitutional group quarters

**ACR:**

*Definition:* Lot size

*Possible Values:* Categorical Text

- N/A
- House on less than one acre
- House on one to less than ten acres
- House on ten or more acres

**BDSP:**

*Definition:* Number of bedrooms

*Possible Values:* Integer between 0 and 99

## Appendix (Continued)

**BLD:**

*Definition:* Units in structure

*Possible Values:* Categorical Text

- N/A (GQ)
- Mobile home or trailer
- One-family house detached
- One-family house attached
- 2 Apartments
- 3-4 Apartments
- 5-9 Apartments
- 10-19 Apartments
- 20-49 Apartments
- 50 or more apartments
- Boat, RV, van, etc.

**ELEP:**

*Definition:* Electricity (monthly cost)

*Possible Values:* Discrete Integer

- (bbb) N/A (GQ/vacant)
- (001) Included in rent or in condo fee
- (002) No charge or electricity not used
- (003 ... 999) \$3 to \$999 (Rounded and top-coded)

## Appendix (Continued)

**FULP:**

**Definition:** Fuel cost (yearly cost for fuels other than gas and electricity)

**Possible Values:** Categorical Text

- (bbbb) N/A (GQ/vacant)
- (0001) Included in rent or in condo fee
- (0002) No charge or these fuels not used
- (0003 ... 9999) \$3 to \$9999 (Rounded and top-coded)

**GASP:**

**Definition:** Gas (monthly cost)

**Possible Values:** Categorical Text

- (bbb) N/A (GQ/vacant)
- (001) Included in rent or in condo fee
- (002) Included in electricity payment
- (003) No charge or gas not used
- (004 ... 999) \$4 to \$999 (Rounded and top-coded)



## Appendix (Continued)

**HFL:**

*Definition:* House heating fuel

*Possible Values:* Categorical Text

- N/A (GQ/vacant)
- Utility gas
- Bottled, tank, or LP gas
- Electricity
- Fuel oil, kerosene, etc.
- Coal or coke
- Wood
- Solar energy
- Other fuel
- No fuel used

**RMSP:**

*Definition:* Number of Rooms

*Possible Values:* Discrete Integer between 0 and 99

## Appendix (Continued)

**TEN:**

*Definition:* Tenure

*Possible Values:* Categorical Text

- N/A (GQ/vacant)
- Owned with mortgage or loan (include home equity loans)
- Owned free and clear
- Rented
- Occupied without payment of rent

**VALP:**

*Definition:* Property value

*Possible Values:* Integer between 1 and 9,999,999

## Appendix (Continued)

**YBL:**

**Definition:** When structure first built

**Possible Values:** Categorical Text

- N/A (GQ)
- 1939 or earlier
- 1940 to 1949
- 1950 to 1959
- 1960 to 1969
- 1970 to 1979
- 1980 to 1989
- 1990 to 1999
- 2000 to 2004
- 2005

**R18:**

**Definition:** Presence of persons under 18 years in household (unweighted)

**Possible Values:** Categorical Text

- N/A (GQ/vacant)
- No person under 18 in household
- 1 or more persons under 18 in household

## Appendix (Continued)

**R60:**

**Definition:** Presence of persons 60 years and over in household (unweighted)

**Possible Values:** Categorical Text

- N/A (GQ/vacant)
- No person 60 and over
- 1 person 60 and over
- 2 or more persons 60 and over