Inference Project

Angelicaqj

3/7/2021

The file OR_house.csv contains household level responses to the American Community Survey for households in Oregon. Technically this is a Public Use Microdata Sample (PUMS) from the 2015 1-year survey. The data were obtained from http://www2.census.gov/programs-surveys/acs/data/pums/2015/1-Year/.

A subset of variables is provided and only households that have at least one person, pay for their electricity, and are not group accommodation. You may assume this is a random sample of all such households in Oregon. The variable descriptions can be found in PUMS_Data_Dictionary_2011_2015.txt.

Explanatory Strategy

Explanatory Problem: Do people living in apartments pay less on electricity than those living in houses? And How much?

Since the primary goal is to answer if there is any difference in electricity cost between the people living in apartments and the people living in houses, one approach is adjusting the number of bedrooms and number of occupants in the household. Below is the inferential model to start with the analysis:

$$ELEP_i = \beta_0 + \beta_1 HousingType + \beta_2 BDSP_i + \beta_3 NP_i + \epsilon_i$$

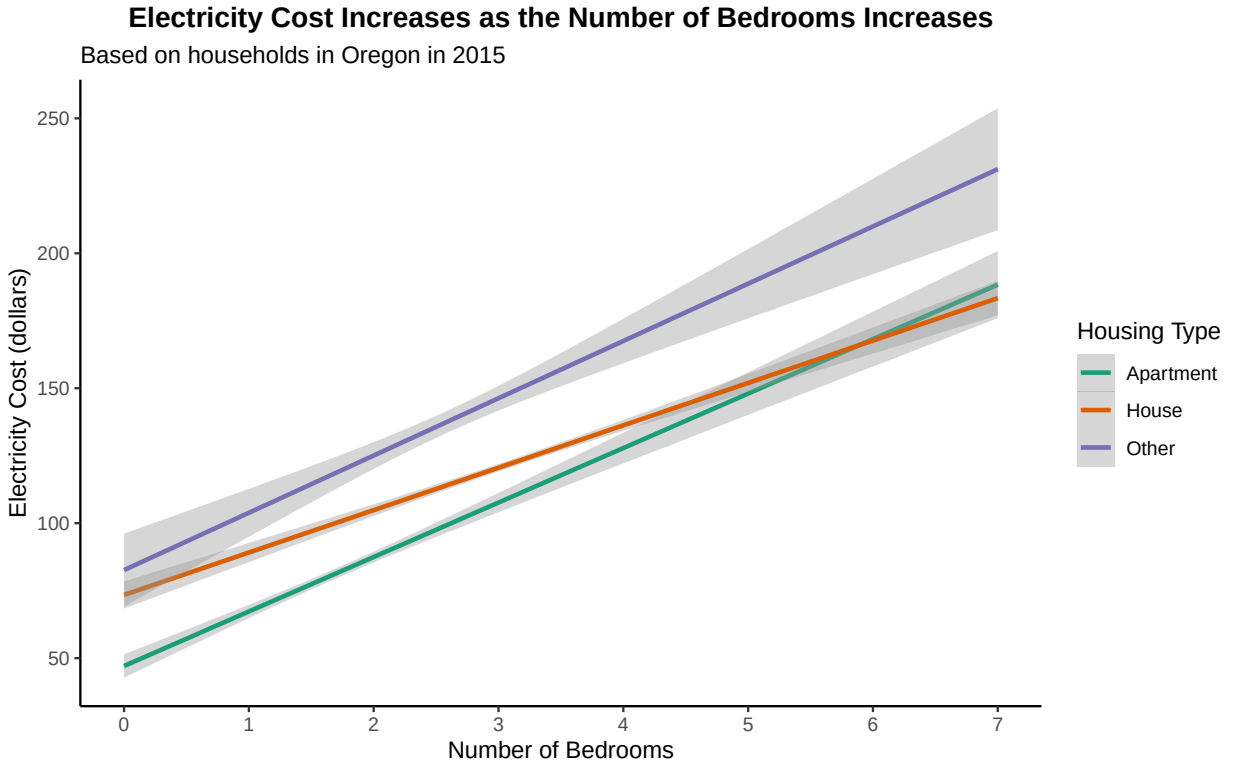
Description of the strategy

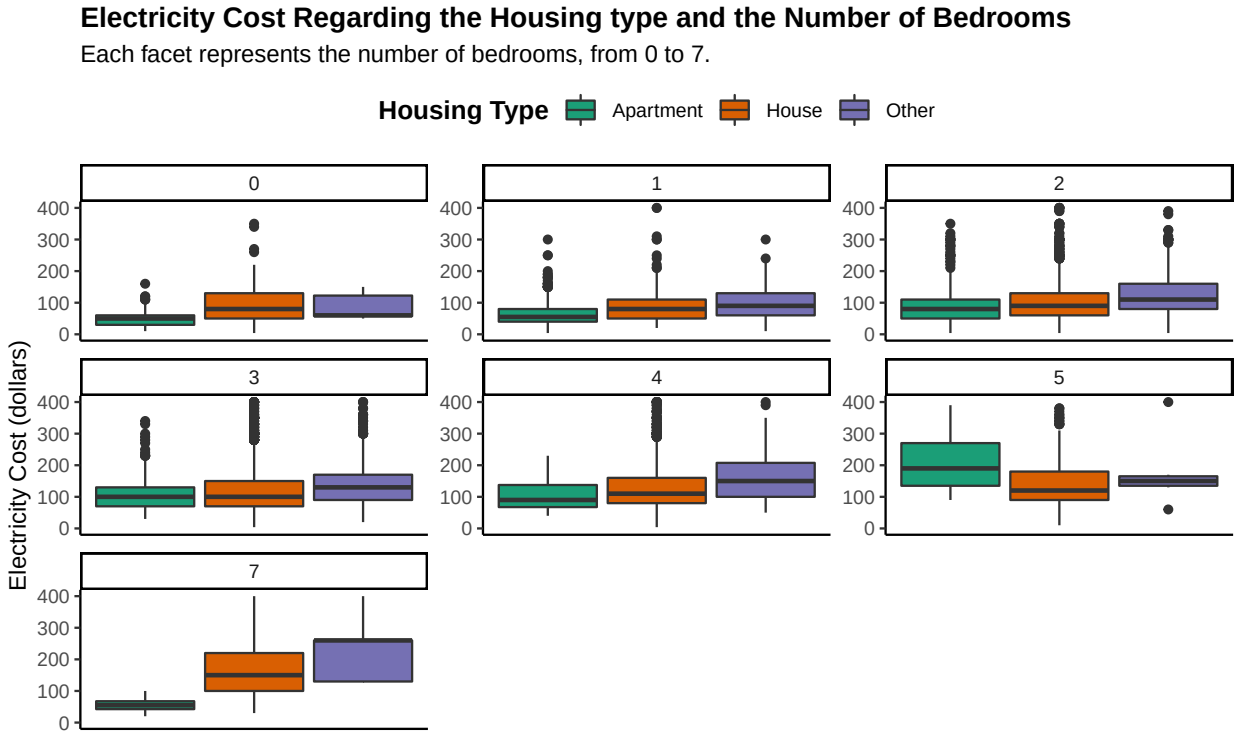
1. Exploratory Analysis by looking at the relationship between the response variable and one of the explanatory variables.
2. Fit 2 linear models where electrical cost is the response variable.

- Model 1 is the full model. It is including the number of bedrooms, number of occupants, and people living in apartments or houses as explanatory variables.
- Model 2 is the reduced model. It is including only people living in apartments or houses as the explanatory variable.

3. Diagnostic of the residuals with the most complicated model, in this case, model_1.
4. Comparing the 2 models performing analysis of variance to see if there is a significant difference between the 2 models.
5. Proceeding with inference to answer the question of interest if there is evidence that people living in apartments pay less on electricity than those living in houses.

1. Exploratory Analysis

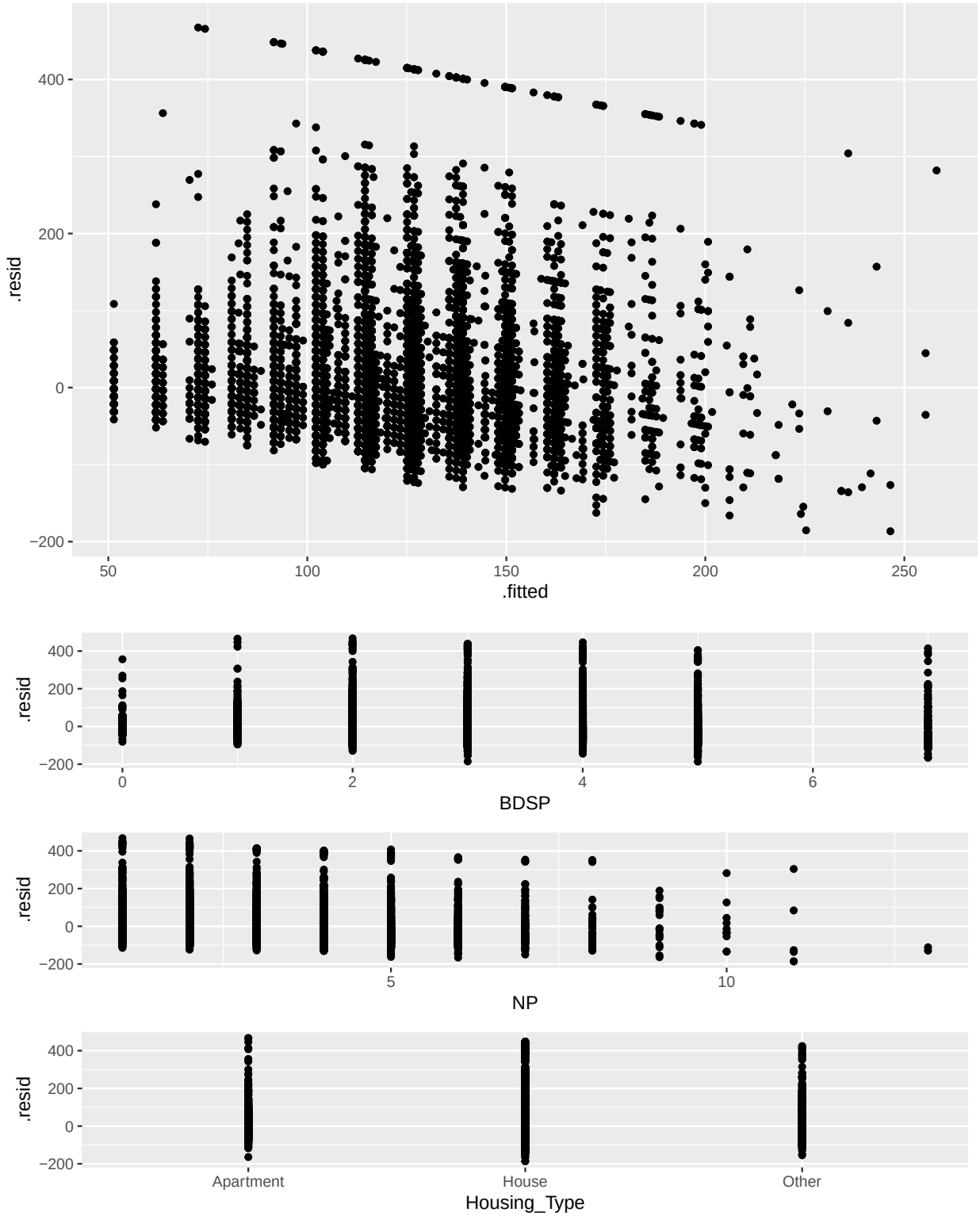




Both plots indicate a positive relationship between the response variable and the explanatory variable except for 5 bedrooms house type is decreasing the cost of electricity.

2. Fit 2 Models

3. Check fit



The first plot by comparing the residuals with the fitted values, it seems that there is a line on the top that can be a cluster related with one of the variables.

4. Compare the 2 models

```
## Analysis of Variance Table
##
## Model 1: ELEP ~ Housing_Type + BDSP + NP
## Model 2: ELEP ~ Housing_Type
##   Res.Df      RSS Df Sum of Sq      F    Pr(>F)
## 1  15161 74674828
## 2  15163 81576923 -2  -6902095 700.66 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##         df      BIC
## model_1  6 172035.9
## model_2  4 173357.4
```

Since the analysis of variance show us that the F-statistic is larger and the p-value is significantly small, there is enough evidence that model_1 is better to proceed with the inferential analysis.

5. Proceed with Inference

```
##
## Call:
## lm(formula = ELEP ~ Housing_Type + BDSP + NP, data = df_1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -186.49  -43.90  -14.48   25.12  467.43
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        39.0794     1.8467   21.16   <2e-16 ***
## Housing_TypeHouse  19.0163     1.7866   10.64   <2e-16 ***
## Housing_TypeOther  42.9409     2.4093   17.82   <2e-16 ***
## BDSP               10.5868     0.6969   15.19   <2e-16 ***
## NP                 12.3144     0.4502   27.35   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 70.18 on 15161 degrees of freedom
## Multiple R-squared:  0.1282, Adjusted R-squared:  0.128
## F-statistic: 557.5 on 4 and 15161 DF,  p-value: < 2.2e-16
```

```
##                       2.5 %   97.5 %
## (Intercept)       35.459695 42.69906
## Housing_TypeHouse 15.514329 22.51817
## Housing_TypeOther 38.218395 47.66349
## BDSP               9.220841 11.95272
## NP                11.431899 13.19681
```

There is strong evidence that people living in apartment pay less on electricity than those living in houses (p-value $<$ 2.2e-16). With 95% confidence, the mean of electricity cost for people living in houses is between 15.51 and 22.52 dollars greater than people living in apartments.