Homework2

Min-kyeong (Min) Cha

Spring semester 2023

1 Python

1.1 Q1

	Control (s.d)	Treatment (s.d)	difference-in-means (p-value)
electricity	1181.33	1086.75	-3.40
	(454.31)	(423.96)	(0.00)
sqft	1633.05	1657.55	0.57
	(682.90)	(686.27)	(0.57)
temp	79.89	79.89	0.02
	(2.16)	(1.97)	(0.99)
Observations	501.0	499.0	

Table 1: Sample statistics by group and difference-in-mean test

Considering control variables, square feet and temperature, two groups with and without retrofit have similar means and standard deviation. On the other hand, the mean of electricity consumption of the two groups differs. Difference-in-mean test results show that the two groups have differences in electricity consumption, and the means of the other two variables are not significantly varying. As there is a difference in an explanatory variable while control variables remain the same, it seems randomization worked.

1.2 Q2

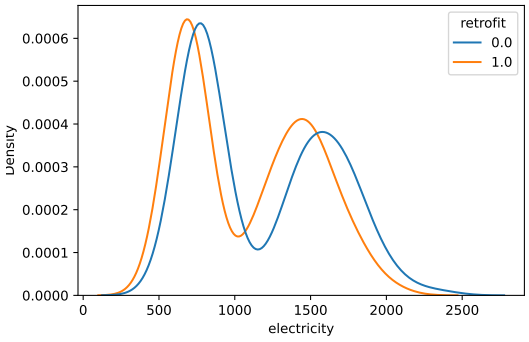

Figure 1: Kernel density plot of the electricity consumption by retrofit group.

The kernel density plot shows that retrofit did affect electricity consumption. In the figure, the density moved left for the treatment group(orange line) compared to the control group(blue line). That being said, retrofit made consumers with a wide range of electricity consumption reduce their usage.

1.3 Q3

	by hand (a)	simulated LS (b)	Package (c)
sqft	0.62	0.62	0.62
retrofit	-109.67	-109.67	-109.67
temp	3.26	3.25	3.26
Constant	-83.6	-83.59	-83.6
Observations	1000.0		

Table 2: Results of different ways of OLS

In Table 2, $\hat{\beta}_{OLS}$ is estimated in three different ways: by matrix multiplication (a), by simulated least squares (b), and using canned package (c). The coefficients estimated are the same regardless of which methods we choose.

2 Stata

2.1 Q1

	(1)	(2)	(3)
	Mean/Std. Dev.	Mean/Std. Dev.	difference/p-value
electricity	1181.33	1086.75	3.40
	(454.31)	(423.96)	(0.00)
sqft	1633.05	1657.55	-0.57
	(682.90)	(686.27)	(0.57)
temp	79.89	79.89	-0.02
	(2.16)	(1.97)	(0.99)
Observations	501	499	1000

Table 3: Sample statistics by group and difference-in-mean test

As in the previous question, while average electricity consumption varies among Control (2) and Treatment (1) group, the means of other variables are significantly similar.

2.2 Q2

The scatter plot shows positive linear relationship between electricity consumption and house size - the more square feet, the more electricity consumed.

2.3 Q3

In Table 4, all the variables statistically significantly influence electricity consumption (p¡0.1). Consistent with Figure 2, square feet have a positive coefficient; holding other variables fixed, one increase of square feet leads to a 0.62 increase in electricity consumption(kWh). Retrofit does reduce electricity consumption: on average, participating in a retrofit program reduces about 110kWh electricity consumption. Lastly, the higher the temperature, the more electricity is consumed. Also, Table 2 and Table 4 have the same regression coefficients.

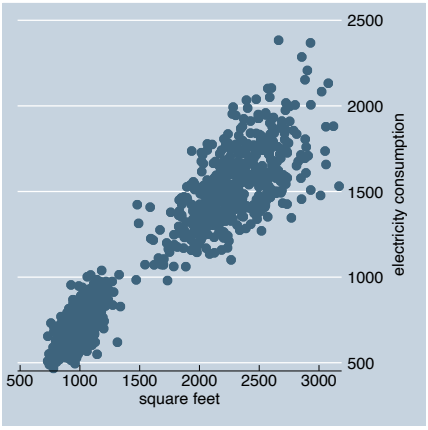

Figure 2: Scatter plot of the electricity consumption and square feet.

VARIABLES	(1) Ordinary least squares
sqft	0.615***
	(0.00678)
retrofit	-109.7***
	(7.943)
temp	3.255*
	(1.932)
Constant	-83.60
	(154.7)
Observations	1,000
R-squared	0.919

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Results of OLS