Homework 2 solutions

Economics 7103

1 Python

1. See table 1. If randomization worked, the simple difference-in-means is an unbiased estimate of the treatment effect.

	Control (s.d.)	Treatment (s.d.)	Difference (p val)
Electricity	1181.33	1086.75	94.58
	(454.31)	(423.96)	(0.00)
Sqft	1633.05	1657.55	-24.50
	(682.90)	(686.27)	(0.57)
Temp	79.89	79.89	-0.00
	(2.16)	(1.97)	(0.99)
Observations	501	499	1000

Table 1: Means by treatment and control group in the sample. The p value is from a two-way t-test for equivalence of means.

2. See figure 1, which is non-parametric evidence that the program worked:
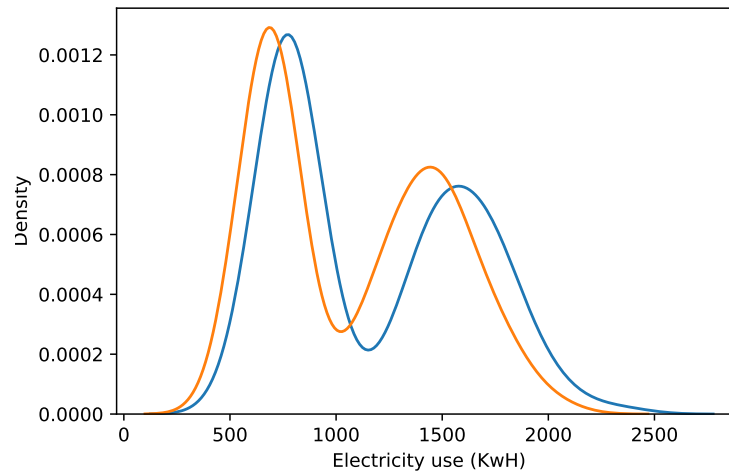

Figure 1: Histogram of treated and control electricity consumption.

3. Each method produces quite similar results that probably only differ in rounding error:

	(a)	(b)	(c)
Retrofit	-109.666	-109.666	-109.666
Sqft	0.615	0.615	0.615
Temperature	3.255	3.256	3.255
Constant	-83.603	-83.638	-83.603
Observations	1000	1000	1000

Table 2: Regression coefficients from OLS by hand (a), simulated OLS (b), and using the Statsmodels package (c).

2   Stata

1. See table 3

	(1) Treated Mean/SD	(2) Controls Mean/SD	(3) Difference Diff./p-value
electricity	1,086.75	1,181.33	94.58**
	(423.96)	(454.31)	(0.00)
sqft	1,657.55	1,633.05	-24.50
	(686.27)	(682.90)	(0.57)
temp	79.89	79.89	-0.00
	(1.97)	(2.16)	(0.99)
Observations obs	499	501	1,000

** p<0.01, * p<0.05

Table 3: Means by treatment group and control group in the sample. The p-value is from a two-way t-test for equivalence of means
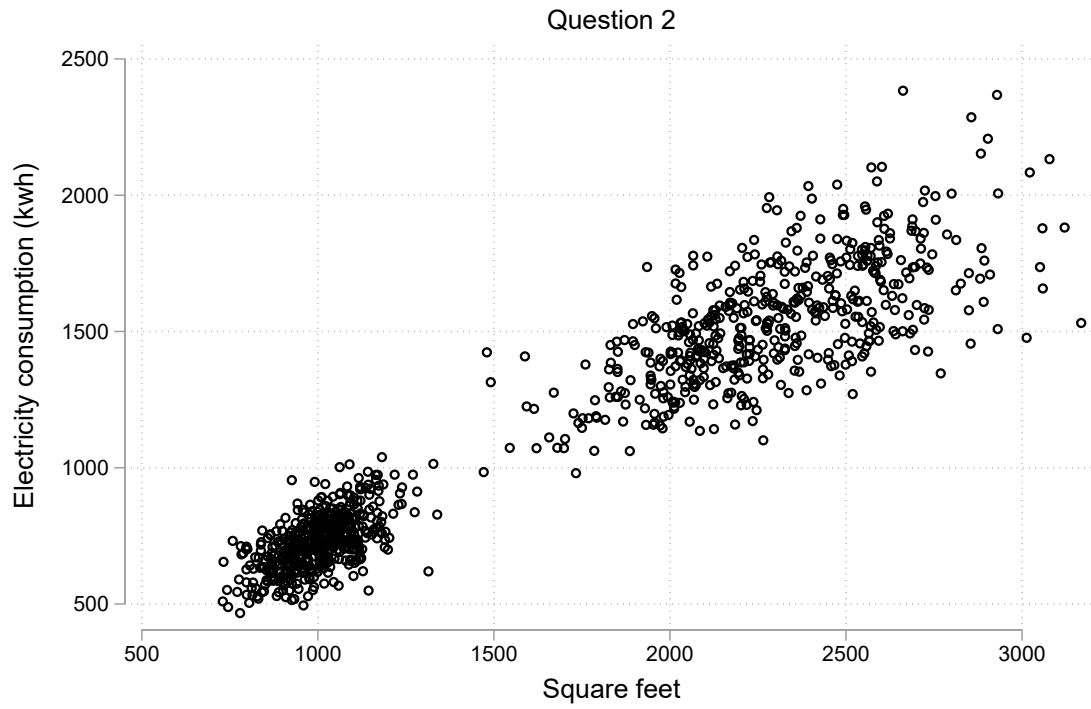
2. See figure 2
3. See table 4

Figure 2: Scatterplot of electricity consumption versus home size.

VARIABLES	(1) Electricity (kwh)
Square feet	0.615***
	(0.00678)
Treatment	-109.7***
	(7.943)
Temperature	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: Regression results with heteroskedasticity-robust standard errors.