SDS 384: Causal Inference Methodology
Homework 2

March 24, 2020

Professor Zigler

Spencer Woody

Work with Qiaohui Lin
My code is available at `github.com/spencerwoody/sds384causal`

Variable name	Description
`Tx`	Whether the EGU has an SnCR installed in that year
`Outcome`	Annual emissions of NO_x in tons
`totOpTime`	Number of hours operated during the year
`HeatInput`	Measure of the amount of fuel burned
`pctCapacity`	Average percent of total operating capacity actually operated
`Phase2`	Indicator of participation on Phase II of the Acid Rain Program
`avgNOxControls`	Average number of other NOx emissions controls (besides SnCR)
`coal_no_scrubber`	Indicator of whether the EGU burns coal as primary fuel and does not have an SO2 scrubber installed
`coal_with_scrubber`	Indicator of whether the EGU burns coal as primary fuel and has an SO2 scrubber installed
`EPA.Region`	Which of 9 EPA defined regions in which the EGU is located

Table 1: Description of relevant variables in the `annualEGUs.csv` data.

This assignment centers around a data set very similar to the one used in the DAPSm paper by Papadogeorgou et al. (2018). The data contain information on power plants operating in the United States in 2002 and 2014, and are available on the Canvas site in the file `annualEGUs.csv`*. Specifically, the units in the data are Electricity Generating Units (EGUs) in 2002 and 2014, some of which were treated with a particular technology to reduce their emissions of NO_x, an important precursor to harmful air pollution. The technology is a Selective Catalytic Reduction or Selective Non Catalytic Reduction System, (SnCR). The outcome of interest is the level of NO_x emissions. Several other characteristics are measured on each power plant. Table 1 lists the variables that you will use for this analysis (you can ignore any other variables you see in the data). For all analyses of these data, log transform the* `Outcome` *variable.*

Exercise 1

Separately for 2002 and 2014, conduct an unadjusted "crude" analysis comparing the average NOx levels for treated and untreated units. Evaluate whether the observed covariates are balanced in this unadjusted analysis.

Here are throughout this assignment, we assume a linear model of the form

$$\log y_i = \beta_0 + \alpha z_i + x_i^{\intercal}\beta + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2) \tag{1}$$

where y_i is the annual emissions of NO_x for the ith power plant, z_i is the treatment indicator, and x_i is a vector of the other covariates listed in Table 1. The parameter α is the treatment effect.

In the crude analysis we use the entire dataset for years 2002 and 2014, respectively, to estimate the parameters in Eq. (1). Full regression estimates from this crude analysis are given in Tables 10 and 11 at the end of the document. The treatment effect estimates from both years are:

- 2002: -0.9525 (SE = 0.0669, $p < 0.0001$)
- 2014: -0.7069 (SE = 0.0450, $p < 0.0001$)

Here are throughout, we use the ordinary least squares estimate of the standard error, which is the outcome of the `lm` command in R.

We see that the causal effect of SnCR installations on annual NO_x emissions is significant and negative for both years, with the effect larger in magnitude for year 2002 than for year 2014.

Tables 2 and 3 present summaries of covariate balance between the treatment and control groups, using various (unpaired) tests of null hypothesis of no difference in mean of the covariate between the groups. We use the t-test, z-test, or chi-squared test depending on the data type. There is significant imbalance present in the dataset from both years. A large portion of the tests return very small p-values, a manifestation of significant differences in the distribution of covariates between treated and control units. This problem is especially bad for year 2014. Therefore, we should be cautious to characterize these figures as consistent estimates of the causal effect of interest since there could be confounding which has not been accounted for.

variable	variable.type	significance.test	test.statistic	p.value
totOpTime	continuous	t-test, difference in means	-4.009	< 0.0001
HeatInput	continuous	t-test, difference in means	-5.377	< 0.0001
pctCapacity	continuous	t-test, difference in means	-1.746	0.0818
Phase2	binary	z-test, difference in proportion	0.087	0.7684
avgNOxControls	continuous	t-test, difference in means	-6.589	< 0.0001
coal_no_scrubber	binary	z-test, difference in proportion	84.663	< 0.0001
coal_with_scrubber	binary	z-test, difference in proportion	7.951	0.0048
EPA.Region	categorical	chi-sq test of independence	283.146	< 0.0001

Table 2: Covariate balance check for crude analysis, year 2002

variable	variable.type	significance.test	test.statistic	p.value
totOpTime	continuous	t-test, difference in means	7.887	< 0.0001
HeatInput	continuous	t-test, difference in means	7.066	< 0.0001
pctCapacity	continuous	t-test, difference in means	9.653	< 0.0001
Phase2	binary	z-test, difference in proportion	72.950	< 0.0001
avgNOxControls	continuous	t-test, difference in means	-5.046	< 0.0001
coal_no_scrubber	binary	z-test, difference in proportion	53.994	< 0.0001
coal_with_scrubber	binary	z-test, difference in proportion	3.564	0.0591
EPA.Region	categorical	chi-sq test of independence	189.448	< 0.0001

Table 3: Covariate balance check for crude analysis, year 2014

Exercise 2

In this exercise you will use a variety of propensity score methods to estimate the causal effect of having an SnCR in a given year on NO_x emissions in that year, under the assumption that the covariates listed in Table 1 are sufficient to adjust for confounding (i.e., that having an SnCR installed is conditionally unconfounded with respect to NO_x emissions). For all parts of this exercise:

- *Use logistic regression with all of the variables in Table 1 (besides Tx and Outcome) included as covariates to estimate the propensity score.*
- *Be sure to check covariate balance for each analysis*
- *Conduct each analysis separately for 2002 and 2014, and comment (in ~3 sentences) on the differences between the analyses in the two years.*
- *I strongly suggest you read up on the following* R *packages to conduct these analyses:* `MatchIt`, `survey`, `ipw`, *and* `twang`.

(a) *When you arrive at a propensity score model, plot the histograms of the estimated propensity scores in treated and untreated units.*

Figure 1 shows histograms for the propensity score estimated using logistic regression, faceted by year and treatment / control groups.

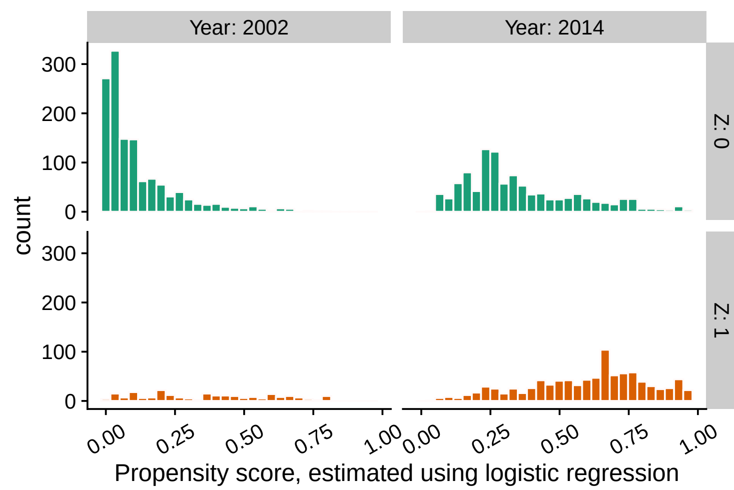

Figure 1: Propensity scores for treated and untreated units for both years in the study. These are estimated from logistic regression of Z (`Tx`) on all the other covariates in Table 1 (aside from `Outcome`).

(b) *Conduct a 1-1 nearest neighbor propensity score matching procedure without replacement*

As in Exercise 1, and throughout the rest of this problem, we use the linear model in Eq. (1) to estimate the treatment effect. We use matching without replacement, matching each treated unit to a control unit. Then these matched pairs are used to estimate the parameters of the linear model. Thus we estimate the average treatment effect on the treated (ATT),

- 2002: -1.1098 (SE = 0.0832, $p < 0.0001$) using 226 matched pairs

- 2014: -0.7262 (SE = 0.0443, p < 0.0001) using 921 matched pairs

As before in Exercise 1, there is a significant negative effect detected for both years, and the effect is larger in magnitude for year 2002. Also, the estimated effect for year 2002 is larger than it was for the crude analysis. Tables 12 and 13 provide the full results of the linear model.

Tables 4 and 5 below provide the balance assessment. For year 2002, there is a remarkable gain in covariate balance from matching. Whereas many covariates appeared imbalanced in the crude analysis of Exercise 1 (represented by small p-values for the null hypothesis of equivalent covariate means between treated and control units), now only the covariate totOpTime has a significant p-value at the 95% level. However, for year 2014 there is still quite substantial covariate imbalance; only the p-value for coal_no_scrubber is insignificant.

variable	variable.type	significance.test	test.statistic	p.value
totOpTime	continuous	t-test, difference in means	-2.005	0.0456
HeatInput	continuous	t-test, difference in means	-1.307	0.1919
pctCapacity	continuous	t-test, difference in means	-1.679	0.0939
Phase2	binary	z-test, difference in proportion	0.053	0.8174
avgNOxControls	continuous	t-test, difference in means	-0.581	0.5617
coal_no_scrubber	binary	z-test, difference in proportion	0.616	0.4326
coal_with_scrubber	binary	z-test, difference in proportion	0.554	0.4566
EPA.Region	categorical	chi-sq test of independence	5.318	0.8057

Table 4: Covariate balance check for one-to-one propensity score matching (Exercise 2(b)), year 2002

variable	variable.type	significance.test	test.statistic	p.value
totOpTime	continuous	t-test, difference in means	6.691	< 0.0001
HeatInput	continuous	t-test, difference in means	5.768	< 0.0001
pctCapacity	continuous	t-test, difference in means	8.138	< 0.0001
Phase2	binary	z-test, difference in proportion	32.167	< 0.0001
avgNOxControls	continuous	t-test, difference in means	-4.288	< 0.0001
coal_no_scrubber	binary	z-test, difference in proportion	34.534	< 0.0001
coal_with_scrubber	binary	z-test, difference in proportion	1.932	0.1645
EPA.Region	categorical	chi-sq test of independence	150.241	< 0.0001

Table 5: Covariate balance check for one-to-one propensity score matching (Exercise 2(b)), year 2014

(c) *Conduct a 1-1 nearest neighbor propensity score matching procedure without replacement and a caliper set to 0.1 standard deviations of the estimated propensity score distribution.*

After using a caliper to obtain matched pairs, the ATT estimates are now

- 2002: -0.9699 (SE = 0.0944, p < 0.0001) using 166 matched pairs (60 unmatched treatment units)
- 2014: -0.7576 (SE = 0.0554, p < 0.0001) using 496 matched pairs (425 unmatched treatment units)

As before, there is a significant negative effect detected for both years, and the effect is larger in magnitude for year 2002. Tables 14 and 15 provide the full results of the linear model.

Tables 6 and 7 show the covariate balance checks. In this case, every covariate appears to be balanced for each year because every p-value is insignificant, demonstrating the advantages of using a caliper. However, this comes at a price of discarding a sizable portion unmatched treatment units in the estimation of the treatment effect, leading to notably larger standard errors.

variable	variable.type	significance.test	test.statistic	p.value
totOpTime	continuous	t-test, difference in means	-0.119	0.9054
HeatInput	continuous	t-test, difference in means	0.253	0.8007
pctCapacity	continuous	t-test, difference in means	-0.039	0.9692
Phase2	binary	z-test, difference in proportion	0.016	0.8979
avgNOxControls	continuous	t-test, difference in means	0.616	0.5382
coal_no_scrubber	binary	z-test, difference in proportion	0.000	1.0000
coal_with_scrubber	binary	z-test, difference in proportion	0.161	0.6880
EPA.Region	categorical	chi-sq test of independence	5.633	0.7760

Table 6: Covariate balance check for one-to-one propensity score matching with a caliper (Exercise 2(c)), year 2002

variable	variable.type	significance.test	test.statistic	p.value
totOpTime	continuous	t-test, difference in means	-0.881	0.3784
HeatInput	continuous	t-test, difference in means	-0.527	0.5986
pctCapacity	continuous	t-test, difference in means	-1.071	0.2846
Phase2	binary	z-test, difference in proportion	0.243	0.6219
avgNOxControls	continuous	t-test, difference in means	-0.125	0.9002
coal_no_scrubber	binary	z-test, difference in proportion	0.294	0.5878
coal_with_scrubber	binary	z-test, difference in proportion	0.000	1.0000
EPA.Region	categorical	chi-sq test of independence	8.839	0.4523

Table 7: Covariate balance check for one-to-one propensity score matching with a caliper (Exercise 2(c)), year 2014

(d) *Conduct an analysis that subclassifies units based on the estimated propensity score*

Here we use subclassification based on the propensity score estimated from logistic regression, creating 4 subgroups. We again use a linear model for estimating the treatment effect, and use a weighted estimate using the sizes of the subgroups as weights (and also calculate a weighted estimate of the standard error in the same manner).

- 2002: -0.6381 (SE = 0.1285)
- 2014: -0.7314 (SE = 0.0890)

As before, there is a significant negative effect detected for both years. Now, however, the effect is slightly larger in magnitude for year 2014 than for year 2002.

Tables 8 and 9 contain covariate balance diagnostics, with each column containing the p-values for each subgroup. For the 2002 data, there are several instances of imbalance suggested by small p-values for the tests of null hypothesis of balance between control and treatment groups. This problem is even worse with the 2014 data. Using a greater number of subgroups might mitigate this problem.

variable	subgroup1	subgroup2	subgroup3	subgroup4
totOpTime	0.5071	0.1241	< 0.0001	0.3076
HeatInput	0.4012	0.6007	0.0029	0.1101
pctCapacity	0.9979	0.3034	0.0012	0.3677
Phase2	1.0000	1.0000	0.5589	0.2496
avgNOxControls	0.0337	0.0361	0.2993	0.9215
coal_no_scrubber	0.0194	1.0000	0.8819	1.0000
coal_with_scrubber	0.5845	0.8112	0.1682	1.0000
EPA.Region	0.0002	0.0104	0.2480	0.0530

Table 8: Covariate balance check for subclassification using propensity score (4 subclasses), year 2002

variable	subgroup1	subgroup2	subgroup3	subgroup4
totOpTime	< 0.0001	0.0013	0.0017	0.0329
HeatInput	< 0.0001	0.2439	0.7289	0.0812
pctCapacity	< 0.0001	0.0004	0.0059	0.1802
Phase2	0.4905	0.2273	0.8723	1.0000
avgNOxControls	0.9718	0.6884	0.0005	< 0.0001
coal_no_scrubber	0.7270	0.1304	0.5050	0.4342
coal_with_scrubber	< 0.0001	0.6324	0.9029	< 0.0001
EPA.Region	0.0006	< 0.0001	0.0036	0.0983

Table 9: Covariate balance check for subclassification using propensity score (4 subclasses), year 2014

(e) *Conduct an IPW analysis using weights $\frac{W_i}{\hat{e}(X_i)} + \frac{1-W_i}{1-\hat{e}(X_i)}$ and be sure to include a visual summary (e.g., histogram) of the estimated weights.*

We calculate the weights as described, using the propensity score estimated from logistic regression as $\hat{e}(X_i)$. We fit a weighted least squares estimate of the linear model in Eq. (1) using the `lm` command in R.

- 2002: -0.5951 (SE = 0.0404)
- 2014: -0.6447 (SE = 0.0404)

There is a significant negative effect detected for both years, and the effect is slightly larger in magnitude for year 2014 than for year 2002. Full results from the regression are presented presented in Tables 16 and 17 at the end of the document.

Figure 2 shows a visualization of the weights arranged by their magnitude and colored by treatment assignment. We notice that there are several observations with weights which are strikingly large, and therefore these weights may dominate the eventual estimated causal effects.

(f) *Conduct an IPW analysis using stabilized weights and be sure to include a visual summary (e.g., histogram) of the estimated weights.*

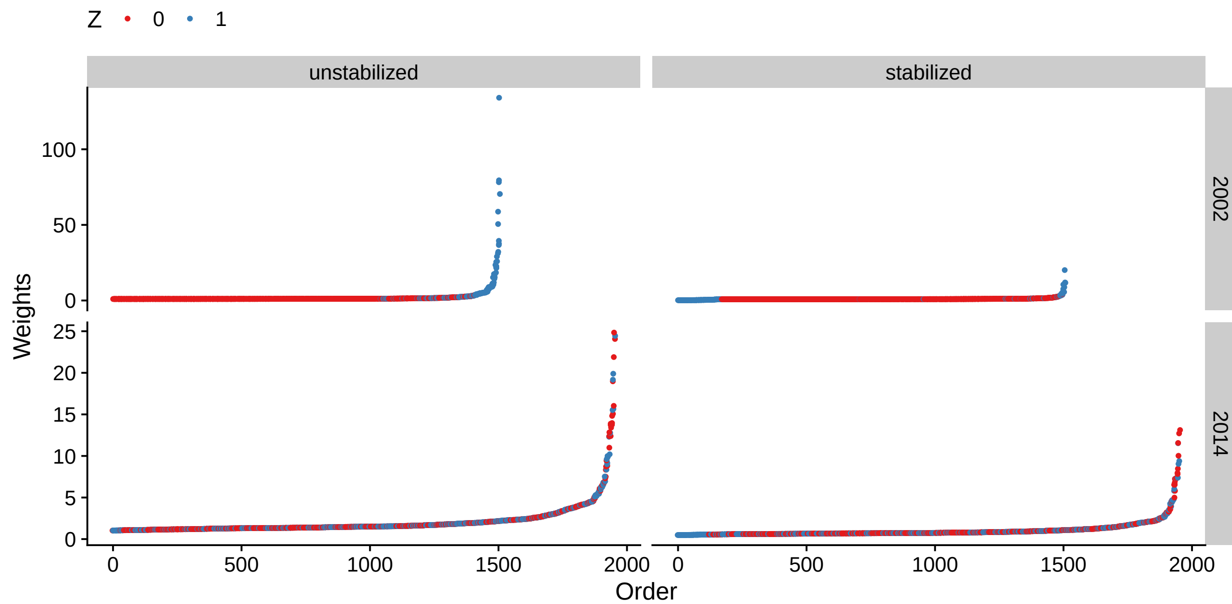

Figure 2: Unstabilized and stabilized inverse probability weights for Exercises (2e) and (2f).

Now the weights become $\frac{W_i \cdot \hat{\Pr}(W_i=1)}{\hat{e}(X_i)} + \frac{(1-W_i)\cdot \hat{\Pr}(W_i=0)}{1-\hat{e}(X_i)}$ where $\Pr(W_i = 1)$ is the marginal probability of assignment to treatment. That is, $\hat{\Pr}(W_i = 1)$ is the sample proportion of units which receive treatment. These stabilized weights are compared to the stabilized weights in Figure 3. We can see how the variance of the weights is reduced in this manner.

- 2002: -0.5349 (SE = 0.0572)
- 2014: -0.6432 (SE = 0.0407)

There is a significant negative effect detected for both years, and the effect is slightly larger in magnitude for year 2014 than for year 2002. However, even though the weights themselves have lower variance, the standard errors for the estimated causal effects are slightly larger than they were for the analysis using unstabilized weights. Full regression results are given in Tables 18 and 19 at the end of the document.

Exercise 3

Describe in ~5 sentences why the answers you obtained with the different propensity score methods in Exercise (2) were different from one another.

Every analysis in Exercise 2 used a linear model to estimate the causal effect of SnCR installation on log-emissions of NO_x. However, the estimates differed considerably. This is all due to how observations were either selected or weighted. That is, matching, both with and without a caliper, removed many observations from consideration when estimating the parameters of the linear model in Eq. (1) because there were not any units with adequately similar propensity scores to these units. Alternatively, the IPW analyses included every observation in the analysis, but weighted each one differently. Finally, the subclassification analysis also included every unit, but weighted each one according to the size of the subclassiciation (based on the estimated propensity score) to which it belonged.

Exercise 4

> *Repeat Exercise (2e), but use a more advanced prediction model (your choice) to estimate the propensity score. Describe (∼3 sentences) any differences.*

We performed a probit regression using Bayesian additive regression trees (BART; Chipman et al., 2010) with the `dbarts` package in R to estimate the propensity score. We estimated the treatment effect with weighted least squares using these weights for the linear model in Eq. (1), similarly to Problem 3(e).

- 2002: -0.6651 (SE = 0.0434)
- 2014: -0.8018 (SE = 0.0410)

Note that the estimated treatment effect is slightly lower for year 2002, and much lower for year 2014. These differences arise entirely from the discrepancy in the estimated propensity scores between logistic regression and from BART.

Figure 3 compares the propensity scores estimated from logistic regression (as used in all of Problem 2) to those estimated from BART. For the both years, BART tends to give a lower propensity score for control units ($Z = 0$), and a higher propensity score for treatment units ($Z = 1$). This pattern is especially notable for year 2014. This suggests that BART might be "overfitting" the estimated propensity score model.

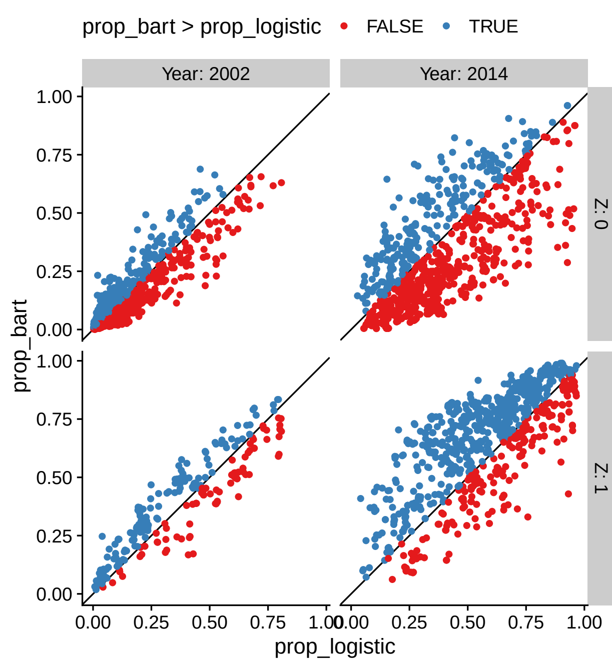

Figure 3: Comparison of estimated propensity scores from logistic regression and BART.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.1269	0.1261	24.80	0.0000
Tx	-0.9525	0.0669	-14.23	0.0000
totOpTime	0.0003	0.0000	20.27	0.0000
HeatInput	0.0000	0.0000	26.03	0.0000
pctCapacity	-0.5005	0.1727	-2.90	0.0038
Phase2	-0.0635	0.0581	-1.09	0.2747
avgNOxControls	-0.1118	0.0427	-2.62	0.0089
coal_no_scrubber	1.7313	0.0673	25.74	0.0000
coal_with_scrubber	1.6007	0.0937	17.09	0.0000
EPA.Region2	-0.1837	0.1136	-1.62	0.1060
EPA.Region3	0.0960	0.1131	0.85	0.3962
EPA.Region4	0.1514	0.1060	1.43	0.1534
EPA.Region5	-0.0944	0.1113	-0.85	0.3964
EPA.Region6	0.1331	0.1124	1.18	0.2362
EPA.Region7	-0.1584	0.1377	-1.15	0.2501
EPA.Region8	-0.0929	0.1377	-0.67	0.5001
EPA.Region9	-0.3204	0.1256	-2.55	0.0108
EPA.Region10	-0.1175	0.3222	-0.36	0.7153

Table 10: “Crude” linear regression model for estimating of treatment effect using 2002 data

Additional tables and figures

References

Hugh A. Chipman, Edward I. George, and Robert E. McCulloch. Bart: Bayesian additive regression trees. *Ann. Appl. Stat.*, 4(1):266–298, 03 2010. doi: 10.1214/09-AOAS285. URL `https://doi.org/10.1214/09-AOAS285`.

Georgia Papadogeorgou, Christine Choirat, and Corwin M Zigler. Adjusting for unmeasured spatial confounding with distance adjusted propensity score matching. *Biostatistics*, 20(2):256–272, 2018.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.3451	0.1263	18.57	0.0000
Tx	-0.7069	0.0450	-15.72	0.0000
totOpTime	0.0004	0.0000	18.85	0.0000
HeatInput	0.0000	0.0000	17.57	0.0000
pctCapacity	-0.6875	0.2424	-2.84	0.0046
Phase2	-0.1244	0.0556	-2.24	0.0255
avgNOxControls	-0.2243	0.0386	-5.81	0.0000
coal_no_scrubber	2.0186	0.0833	24.22	0.0000
coal_with_scrubber	1.9837	0.0809	24.51	0.0000
EPA.Region2	0.0422	0.1298	0.32	0.7453
EPA.Region3	0.0476	0.1272	0.37	0.7082
EPA.Region4	0.0133	0.1191	0.11	0.9111
EPA.Region5	-0.0205	0.1234	-0.17	0.8682
EPA.Region6	0.2671	0.1220	2.19	0.0287
EPA.Region7	0.0379	0.1465	0.26	0.7961
EPA.Region8	-0.1861	0.1491	-1.25	0.2120
EPA.Region9	-0.7299	0.1240	-5.89	0.0000
EPA.Region10	-0.1352	0.2280	-0.59	0.5531

Table 11: “Crude” linear regression model for estimating of treatment effect using 2014 data

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.3332	0.1859	17.93	0.0000
Tx	-1.1098	0.0832	-13.34	0.0000
totOpTime	0.0003	0.0000	9.67	0.0000
HeatInput	0.0000	0.0000	10.87	0.0000
pctCapacity	-0.9851	0.3778	-2.61	0.0094
Phase2	-0.0967	0.1190	-0.81	0.4171
avgNOxControls	-0.0534	0.0838	-0.64	0.5241
coal_no_scrubber	1.8828	0.1546	12.18	0.0000
coal_with_scrubber	1.8678	0.1952	9.57	0.0000
EPA.Region2	-0.0801	0.1383	-0.58	0.5628
EPA.Region3	0.2501	0.1679	1.49	0.1371
EPA.Region4	0.2099	0.1532	1.37	0.1712
EPA.Region5	0.1218	0.2445	0.50	0.6186
EPA.Region6	0.0095	0.1634	0.06	0.9538
EPA.Region7	-0.0014	0.4712	-0.00	0.9977
EPA.Region8	-0.1420	0.6301	-0.23	0.8217
EPA.Region9	-0.3774	0.1470	-2.57	0.0106
EPA.Region10	-0.1678	0.3545	-0.47	0.6362

Table 12: Linear regression model for estimating of treatment effect using 2002 data, after using one-to-one matching (Exercise 2(b))

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.4951	0.1272	19.61	0.0000
Tx	-0.7262	0.0443	-16.40	0.0000
totOpTime	0.0004	0.0000	18.86	0.0000
HeatInput	0.0000	0.0000	17.38	0.0000
pctCapacity	-0.8106	0.2432	-3.33	0.0009
Phase2	-0.2393	0.0608	-3.93	0.0001
avgNOxControls	-0.2479	0.0389	-6.37	0.0000
coal_no_scrubber	2.0182	0.0872	23.14	0.0000
coal_with_scrubber	2.0067	0.0817	24.57	0.0000
EPA.Region2	0.0125	0.1277	0.10	0.9223
EPA.Region3	0.0909	0.1257	0.72	0.4698
EPA.Region4	0.0552	0.1174	0.47	0.6384
EPA.Region5	-0.0282	0.1221	-0.23	0.8175
EPA.Region6	0.2705	0.1213	2.23	0.0259
EPA.Region7	0.1066	0.1546	0.69	0.4908
EPA.Region8	-0.1946	0.1493	-1.30	0.1926
EPA.Region9	-0.7155	0.1219	-5.87	0.0000
EPA.Region10	-0.1188	0.2241	-0.53	0.5959

Table 13: Linear regression model for estimating of treatment effect using 2014 data, after using one-to-one matching (Exercise 2(b))

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.1323	0.2012	15.57	0.0000
Tx	-0.9699	0.0944	-10.28	0.0000
totOpTime	0.0003	0.0000	7.99	0.0000
HeatInput	0.0000	0.0000	10.10	0.0000
pctCapacity	-0.5376	0.4134	-1.30	0.1943
Phase2	-0.0358	0.1311	-0.27	0.7849
avgNOxControls	-0.0580	0.0974	-0.60	0.5520
coal_no_scrubber	2.1735	0.1637	13.28	0.0000
coal_with_scrubber	1.8272	0.2172	8.41	0.0000
EPA.Region2	0.0156	0.1621	0.10	0.9236
EPA.Region3	0.0241	0.1795	0.13	0.8934
EPA.Region4	0.0796	0.1767	0.45	0.6525
EPA.Region5	0.2477	0.2577	0.96	0.3371
EPA.Region6	0.0560	0.1846	0.30	0.7620
EPA.Region7	-0.0550	0.3824	-0.14	0.8857
EPA.Region8	-0.7518	0.6312	-1.19	0.2346
EPA.Region9	-0.2674	0.1889	-1.42	0.1579
EPA.Region10	-1.0731	0.9009	-1.19	0.2345

Table 14: Linear regression model for estimating of treatment effect using 2002 data, after using one-to-one matching with a caliper (Exercise 2(c))

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.2247	0.2020	11.01	0.0000
Tx	-0.7576	0.0554	-13.67	0.0000
totOpTime	0.0004	0.0000	13.79	0.0000
HeatInput	0.0000	0.0000	13.71	0.0000
pctCapacity	-0.8967	0.3137	-2.86	0.0043
Phase2	-0.1476	0.0838	-1.76	0.0787
avgNOxControls	-0.2089	0.0531	-3.94	0.0001
coal_no_scrubber	2.1244	0.1200	17.71	0.0000
coal_with_scrubber	2.0743	0.1066	19.46	0.0000
EPA.Region2	0.3108	0.2032	1.53	0.1264
EPA.Region3	0.4681	0.2045	2.29	0.0223
EPA.Region4	0.3958	0.1947	2.03	0.0423
EPA.Region5	0.3375	0.1982	1.70	0.0890
EPA.Region6	0.4620	0.1950	2.37	0.0180
EPA.Region7	0.3952	0.2372	1.67	0.0959
EPA.Region8	-0.0833	0.2218	-0.38	0.7072
EPA.Region9	-0.4709	0.2047	-2.30	0.0216
EPA.Region10	-0.7215	0.3992	-1.81	0.0710

Table 15: Linear regression model for estimating of treatment effect using 2014 data, after using one-to-one matching with a caliper (Exercise 2(c))

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.1828	0.1129	28.20	0.0000
Tx	-0.5951	0.0404	-14.73	0.0000
totOpTime	0.0003	0.0000	18.70	0.0000
HeatInput	0.0000	0.0000	30.09	0.0000
pctCapacity	-0.7127	0.1780	-4.00	0.0001
Phase2	-0.1846	0.0525	-3.52	0.0005
avgNOxControls	-0.0141	0.0383	-0.37	0.7122
coal_no_scrubber	2.0877	0.0673	31.02	0.0000
coal_with_scrubber	1.8016	0.0861	20.92	0.0000
EPA.Region2	-0.1740	0.1062	-1.64	0.1016
EPA.Region3	0.2463	0.1024	2.40	0.0163
EPA.Region4	0.1218	0.0965	1.26	0.2073
EPA.Region5	-0.0723	0.1008	-0.72	0.4737
EPA.Region6	0.0904	0.1037	0.87	0.3833
EPA.Region7	-0.0247	0.1303	-0.19	0.8498
EPA.Region8	-0.2829	0.1374	-2.06	0.0396
EPA.Region9	-0.3766	0.1209	-3.12	0.0019
EPA.Region10	-0.2326	0.3523	-0.66	0.5091

Table 16: Linear regression model with inverse probability weights for estimating of treatment effect using 2002 data (Exercise 2(e))

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.3423	0.1311	17.87	0.0000
Tx	-0.6447	0.0404	-15.96	0.0000
totOpTime	0.0004	0.0000	19.95	0.0000
HeatInput	0.0000	0.0000	16.73	0.0000
pctCapacity	-0.5570	0.2084	-2.67	0.0076
Phase2	-0.1954	0.0584	-3.34	0.0008
avgNOxControls	-0.1584	0.0392	-4.04	0.0001
coal_no_scrubber	2.4075	0.0803	29.97	0.0000
coal_with_scrubber	2.3278	0.0768	30.32	0.0000
EPA.Region2	0.0331	0.1329	0.25	0.8032
EPA.Region3	0.1329	0.1289	1.03	0.3029
EPA.Region4	0.0253	0.1217	0.21	0.8355
EPA.Region5	-0.0472	0.1255	-0.38	0.7069
EPA.Region6	0.1828	0.1232	1.48	0.1380
EPA.Region7	0.3232	0.1465	2.21	0.0275
EPA.Region8	-0.4678	0.1449	-3.23	0.0013
EPA.Region9	-0.7337	0.1245	-5.89	0.0000
EPA.Region10	-0.1528	0.2527	-0.60	0.5455

Table 17: Linear regression model with inverse probability weights for estimating of treatment effect using 2014 data (Exercise 2(e))

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.2503	0.1151	28.23	0.0000
Tx	-0.5349	0.0572	-9.36	0.0000
totOpTime	0.0003	0.0000	20.27	0.0000
HeatInput	0.0000	0.0000	27.49	0.0000
pctCapacity	-0.5994	0.1718	-3.49	0.0005
Phase2	-0.0699	0.0559	-1.25	0.2112
avgNOxControls	-0.0864	0.0410	-2.11	0.0351
coal_no_scrubber	1.7745	0.0669	26.53	0.0000
coal_with_scrubber	1.5858	0.0904	17.53	0.0000
EPA.Region2	-0.2973	0.1107	-2.69	0.0073
EPA.Region3	-0.0205	0.1078	-0.19	0.8493
EPA.Region4	0.0122	0.1013	0.12	0.9041
EPA.Region5	-0.1942	0.1056	-1.84	0.0661
EPA.Region6	-0.0441	0.1085	-0.41	0.6841
EPA.Region7	-0.2497	0.1336	-1.87	0.0618
EPA.Region8	-0.2720	0.1353	-2.01	0.0445
EPA.Region9	-0.2359	0.1272	-1.85	0.0638
EPA.Region10	-0.3239	0.3999	-0.81	0.4181

Table 18: Linear regression model with stabilized inverse probability weights for estimating of treatment effect using 2002 data (Exercise 2(f))

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.3351	0.1314	17.77	0.0000
Tx	-0.6432	0.0407	-15.81	0.0000
totOpTime	0.0004	0.0000	19.94	0.0000
HeatInput	0.0000	0.0000	16.70	0.0000
pctCapacity	-0.5593	0.2079	-2.69	0.0072
Phase2	-0.1845	0.0584	-3.16	0.0016
avgNOxControls	-0.1534	0.0395	-3.89	0.0001
coal_no_scrubber	2.4006	0.0806	29.77	0.0000
coal_with_scrubber	2.3399	0.0769	30.41	0.0000
EPA.Region2	0.0487	0.1335	0.36	0.7153
EPA.Region3	0.1181	0.1294	0.91	0.3615
EPA.Region4	0.0249	0.1222	0.20	0.8385
EPA.Region5	-0.0609	0.1260	-0.48	0.6288
EPA.Region6	0.1914	0.1237	1.55	0.1221
EPA.Region7	0.2943	0.1473	2.00	0.0459
EPA.Region8	-0.4506	0.1458	-3.09	0.0020
EPA.Region9	-0.7405	0.1250	-5.92	0.0000
EPA.Region10	-0.1873	0.2546	-0.74	0.4619

Table 19: Linear regression model with stabilized inverse probability weights for estimating of treatment effect using 2014 data (Exercise 2(f))

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.1718	0.1165	27.23	0.0000
Tx	-0.6651	0.0434	-15.32	0.0000
totOpTime	0.0003	0.0000	19.24	0.0000
HeatInput	0.0000	0.0000	29.39	0.0000
pctCapacity	-0.7669	0.1769	-4.34	0.0000
Phase2	-0.1569	0.0557	-2.82	0.0049
avgNOxControls	-0.0828	0.0416	-1.99	0.0468
coal_no_scrubber	1.9802	0.0690	28.69	0.0000
coal_with_scrubber	1.6800	0.0916	18.34	0.0000
EPA.Region2	-0.1415	0.1060	-1.33	0.1822
EPA.Region3	0.2399	0.1045	2.30	0.0218
EPA.Region4	0.1718	0.0976	1.76	0.0787
EPA.Region5	-0.0032	0.1043	-0.03	0.9753
EPA.Region6	0.0617	0.1035	0.60	0.5514
EPA.Region7	-0.1163	0.1349	-0.86	0.3886
EPA.Region8	-0.1497	0.1441	-1.04	0.2992
EPA.Region9	-0.4095	0.1196	-3.42	0.0006
EPA.Region10	-0.2244	0.3406	-0.66	0.5102

Table 20: Linear regression model with inverse probability weights (estimated using BART) for estimating of treatment effect using 2002 data (Exercise 4)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.2510	0.1222	18.42	0.0000
Tx	-0.8018	0.0410	-19.57	0.0000
totOpTime	0.0004	0.0000	19.65	0.0000
HeatInput	0.0000	0.0000	16.74	0.0000
pctCapacity	-0.9289	0.2388	-3.89	0.0001
Phase2	-0.0768	0.0598	-1.29	0.1988
avgNOxControls	-0.2529	0.0388	-6.52	0.0000
coal_no_scrubber	2.1620	0.0828	26.12	0.0000
coal_with_scrubber	2.0603	0.0820	25.13	0.0000
EPA.Region2	0.2323	0.1240	1.87	0.0612
EPA.Region3	0.2306	0.1192	1.93	0.0532
EPA.Region4	0.1503	0.1108	1.36	0.1751
EPA.Region5	0.1180	0.1157	1.02	0.3077
EPA.Region6	0.2884	0.1125	2.56	0.0104
EPA.Region7	0.3200	0.1419	2.26	0.0242
EPA.Region8	-0.2502	0.1402	-1.79	0.0743
EPA.Region9	-0.7432	0.1176	-6.32	0.0000
EPA.Region10	0.0426	0.2432	0.18	0.8610

Table 21: Linear regression model with inverse probability weights (estimated using BART) for estimating of treatment effect using 2014 data (Exercise 4)